The System-Level Cost of AI Coordination

M. Dalbert Ma*

London Business School

March 23, 2026

Abstract

Artificial intelligence is increasingly deployed not only to perform individual tasks but to coordinate entire systems of interdependent operations. This paper offers the first systematic empirical investigation of what happens when an interdependent organizational system is delegated to centralized AI control. I study a partially automated container terminal in which autonomous vehicles under a centralized dispatching algorithm share infrastructure with independently coordinated human-operated trucks, leveraging 780,000 task-level observations over 364 days and rare mechanical faults as a source of plausibly exogenous disruption. In areas governed predominantly by the algorithm, delays propagate persistently through the system, with recovery times exceeding two hours; in areas where the AI coexists with independent human agents, recovery occurs within 70 minutes. As the algorithm improves over the sample period, it increases throughput by 45% while post-disruption recovery deteriorates by 72%. A formal model rationalizes these findings: a throughput-maximizing coordinator will rationally compress the scheduling of sequential operations, systematically eliminating the temporal slack that would otherwise absorb local shocks—and this compression intensifies as optimization improves. The results identify a cost of AI-enabled centralization that grows with the quality of the algorithm itself.

Keywords: algorithmic coordination, artificial intelligence, centralization, slack, resilience

JEL Codes: L23, M11, O33

*Contact: dma@london.edu. I am especially grateful to my advisor, Michael Jacobides, for his guidance and support. For helpful comments and conversations, I also thank Francisco Brahm, Olenka Kacperczyk, Zheng Li, Stefano Puntoni, Keyvan Vakili, and Lorry Wu, as well as conference participants at Wharton WINDS 2026. I am grateful to the partner firm for data access and for facilitating field visits. I gratefully acknowledge funding from the LBS Institute of Entrepreneurship and Private Capital. All errors remain mine.

1 Introduction

One of the most consequential promises of artificial intelligence (AI) is the coordination and optimization of entire systems of interdependent operations (Acemoglu and Johnson, 2023; Agrawal et al., 2018; Iansiti and Lakhani, 2020). Early demonstrations of these capabilities—from AI agents that autonomously write and deploy software to systems that orchestrate complex logistics—and the sheer expectation of what they portend have already begun to reshape capital markets.[1] If realized, indeed, the implications extend far beyond the productivity and displacement of individual workers—to the organization of complex activity itself, the boundaries and justification of firms, and the value of the human judgment, knowledge, and institutional structures that have historically governed it (Brynjolfsson et al., 2018; Acemoglu et al., 2026).

Yet there is still little empirical evidence on what happens when AI is used to govern an interdependent organizational system. Most of what we know about AI's economic effects comes from settings in which it assists on discrete tasks or supports individual workers and teams (Brynjolfsson et al., 2025; Noy and Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2026, 2025). Even in those settings, recent work has documented how AI systems can generate failures that go undetected by standard evaluation—in individual human–AI interactions (Potts and Sudhof, 2026) and in the reliability of AI agents executing tasks (Rabanser et al., 2026). This paper takes these concerns to the system level, asking what happens when the object of AI deployment is not a task, but an interdependent organizational system.

This paper provides, to my knowledge, the first systematic empirical investigation of what happens when AI governs an interdependent organizational system. I study a partially automated container terminal where a centralized AI system coordinates the movement of autonomous vehicles while operating alongside independently controlled human agents on the same physical infrastructure. The hybrid environment embeds two coordination architectures within a single production system. In some parts of the terminal, activity is governed predominantly by a centralized dispatching algorithm; in others, autonomous vehicles must coexist with independent human-controlled trucks making decentralized movement decisions. What varies across these zones is not the technology or the infrastructure—which are identical—but whether coordination is concentrated under a single algorithm or distributed across independently acting agents.

To motivate the empirical inquiry, I first develop a formal model of coordination that closely resembles the sequentially interdependent system in the data. A central coordinator—here, an AI system—chooses how tightly to schedule linked activities, and the key choice is how much slack to leave between successive movements. Leaving more slack reduces throughput in normal operation but limits the extent to which one

[1] In what industry observers have termed the "SaaSpocalypse," the S&P 500 software and services index fell nearly 13% over six trading sessions in early 2026, erasing approximately $830 billion in market value, as investors concluded that AI coordination capabilities posed an existential threat to incumbent enterprise software platforms (see Oguh et al., 2026; Olson, 2026, for coverage).

delay disrupts what follows; compressing slack raises steady-state efficiency but makes local disruptions more likely to propagate through the system. The model shows that a centralized coordinator seeking to maximize throughput will rationally choose a tightly optimized system with little slack, even though doing so makes disruption propagation more likely. By contrast, independent agents—scheduling under uncertainty and asymmetric private costs—endogenously generate the buffer that absorbs propagating delays. The resulting fragility of the centralized system is a cost that the optimizer rationally and intentionally incurs.

I then test this framework using proprietary operational data from the terminal, leveraging 780,000 task-level observations over 364 days. The empirical analysis tracks how the system responds to localized mechanical faults and how those responses change as the centralized algorithm improves over time. Three patterns stand out. First, delays propagate throughout the system in areas governed predominantly by centralized algorithmic coordination, whereas areas in which the AI system operates alongside independent human agents recover much faster. Second, the productivity gains generated by the centralized system appear to come, in substantial part, from the compression of slack between linked activities. Third, as the algorithm improves over the course of the year and steady-state productivity rises, the system's ability to absorb disruption does not improve with it—if anything, it worsens. The evidence suggests that the AI, in pursuing tighter optimization, rationally increases the system's exposure to disruption propagation.

These findings contribute to the emerging literature on AI, economics, and strategy in three ways. First, they shift attention from task-level augmentation to system-level coordination. Existing work has shown that AI can improve how individuals and teams perform tasks (Brynjolfsson et al., 2025; Noy and Zhang, 2023; Dell'Acqua et al., 2025); this paper shows that when AI governs an interdependent system, the relevant consequences may only become visible at the level of the system itself. Second, the paper identifies a limit to the case for AI-enabled centralization: even if AI relaxes the informational and cognitive constraints that historically made centralized coordination costly (Csaszar, 2025), it may still produce system architectures that are efficient in normal operation but fragile when disruptions occur (Agrawal et al., 2018; Iansiti and Lakhani, 2020). Unlike the classical costs of centralization, which diminish as AI improves, this cost intensifies with better optimization. Third, the paper quantifies the efficiency–resilience tradeoff that this mechanism produces—the algorithm's own improvement trajectory traces the frontier, identifying the disruption frequency above which centralization becomes net negative.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 presents the formal model and derives the main predictions. Section 4 describes the setting and data. Section 5 outlines the empirical strategy. Section 6 presents the results. Section 7 concludes.

2 Related Literature

A growing body of evidence documents the economic effects of artificial intelligence at the level of the individual task. Field experiments and large-scale studies show that AI substantially improves productivity in customer service (Brynjolfsson et al., 2025), management consulting (Dell'Acqua et al., 2026), professional writing (Noy and Zhang, 2023), and software development (Peng et al., 2023), while assessments of AI capabilities indicate that these effects now touch virtually every occupational category (Eloundou et al., 2024; Felten et al., 2021). In team settings, AI-enabled individuals match the performance of two-person teams, suggesting that AI can replicate certain benefits of human collaboration (Dell'Acqua et al., 2025). This body of work has supported the characterization of AI as a general-purpose technology with the capability to augment or substitute for human labor at the task level (Brynjolfsson and McAfee, 2014; Brynjolfsson et al., 2018).

These task-level capabilities have a second-order implication for how organizations are structured. As AI masters a wider range of activities, it becomes feasible to place not just individual tasks but entire sequences of interdependent operations under unified algorithmic control. Consecutive tasks can be "chained" into optimized schedules, eliminating the coordination frictions that arise when different agents manage each step (Demirer et al., 2026). More broadly, AI's capacity for prediction and real-time coordination removes the information-processing constraints that historically favored decentralized organizational structures (Agrawal et al., 2022). The known costs of centralization—search failure (Rivkin and Siggelkow, 2003), communication distortion (Alonso et al., 2008), and evaluation error (Sah and Stiglitz, 1986)—are functions of bounded rationality, and AI relaxes these constraints, shifting the optimal organizational form toward greater centralization (Csaszar, 2025; Iansiti and Lakhani, 2020; Jacobides et al., 2021, 2024).

Yet if AI's capabilities are evaluated at the task level, its deployment increasingly occurs at the system level—on interdependent sequences of operations where the performance of each task affects every subsequent one. This creates a gap between what the evidence base measures and what organizations actually do with AI. Even within task-level settings, recent work has begun to identify systematic failures that standard evaluations miss: the majority of AI failures in human–AI interactions go undetected by users (Potts and Sudhof, 2026), and rising accuracy on benchmarks obscures persistent reliability failures in AI agents (Rabanser et al., 2026). If AI can fail invisibly on individual tasks, the question of what happens when it governs an entire system of interdependent operations becomes pressing.

A theoretically motivated and analytically useful lens through which to examine these system-level consequences is the concept of organizational slack. The idea that organizations maintain excess capacity—resources beyond what is strictly necessary for current operations—has a long history in strategy and the

behavioral theory of the firm (Cyert and March, 1963), and subsequent work has established that operational slack improves firms' ability to recover from supply chain disruptions (Hendricks et al., 2009; Kovach et al., 2015) and that the relationship between slack and performance is non-monotone (Nohria and Gulati, 1996). Much of the case for AI-enabled centralization rests precisely on the premise that AI can unbound rationality and overcome the information-processing limits that historically necessitated such buffers (Csaszar, 2025)—that by centrally coordinating interdependent operations, the algorithm can eliminate idle time, compress schedules, and raise throughput. The present paper leverages this theoretical construct as its central analytical framework, developing a formal model that identifies temporal buffer as the variable through which centralized algorithmic coordination simultaneously generates efficiency and fragility, and testing its predictions in a setting that permits clean identification of system-level dynamics.

3 Theoretical Framework

Building on the constructs introduced in Section 2, I develop a stylized model of task coordination in a shared-infrastructure system to formalize the link between coordination architecture, temporal buffer, and disruption propagation. The model generates four testable predictions. This section summarizes the setup, key results, and predictions; full derivations and proofs appear in Appendix A.

3.1 Model

Environment. A corridor of unit capacity serves N agents performing sequential tasks. Each task requires $\mu > 0$ time units. Agent i enters the corridor at time $\max(a_i, F_{i-1})$, where a_i is the agent's scheduled arrival and F_{i-1} is the predecessor's exit time. The gap between consecutive entries defines the *buffer*:

$$g_i \equiv a_i - F_{i-1}.$$

Positive buffer represents idle capacity; zero buffer means the next agent enters immediately upon its predecessor's departure.

Centralized coordination. A single algorithm observes F_{i-1} in real time and dispatches each agent dynamically. Because the algorithm has full information and internalizes the social cost of idleness, it sets $a_i = F_{i-1}$—that is, zero buffer. Proposition 1 in Appendix A shows that this policy is uniquely optimal: any positive buffer reduces steady-state throughput $\Pi^C = 1/\mu$ without generating offsetting benefits.

Independent coordination. Now suppose each agent is managed by a separate decision process. Agent i commits to a departure time a_i *before* observing F_{i-1}, relying instead on a noisy signal $\sigma_i = F_{i-1} + \eta_i$, where $\eta_i \sim \mathcal{N}(0, \tau^2)$. The agent bears a private waiting cost w per unit of idle time and a private delay cost v_0 per unit of tardiness, but does not internalize downstream congestion. The unique symmetric equilibrium strategy is

$$a^*(\sigma_i) \;=\; \sigma_i \;+\; \tau\,\Phi^{-1}\Big(\frac{w}{w+v_0}\Big), \tag{1}$$

where Φ^{-1} is the standard normal quantile function (Proposition 2). When waiting costs exceed private delay costs ($w > v_0$), each agent pads its schedule, producing a strictly positive mean buffer:

$$\bar{g} \;=\; \tau\,\Phi^{-1}\Big(\frac{w}{w+v_0}\Big) \;>\; 0.$$

Buffer arises endogenously from two features absent under centralization: information imprecision ($\tau > 0$) and cost asymmetry ($w > v_0$).

Disruption propagation. A shock of duration d blocks the corridor at position j. Under centralized coordination, zero buffer means every downstream agent inherits the full delay:

$$\Delta_{j+k} \;=\; d \qquad \text{for all } k \geq 1.$$

The delay propagates without decay (Proposition 4, part i). Under independent coordination, each agent's buffer absorbs part of the shock:

$$\Delta_{j+k+1} \;=\; \big(\Delta_{j+k} - g_{j+k+1}\big)^+,$$

so delays decay geometrically and converge to zero in finite time (Proposition 4, part ii). Expected recovery time is approximately $E[T_R] \approx d\mu/\bar{g} + d$, which is finite and decreasing in buffer $\bar{g}$ (Corollary 3).

Mixed coordination. When centralized and independent agents share infrastructure, the independent agents act as *circuit breakers*. Each independently coordinated agent that enters the corridor carries positive buffer, which absorbs accumulated delay and breaks the zero-buffer chains through which shocks propagate. Recovery time in mixed systems depends on the frequency and spacing of independent agents in the sequence (Proposition 7).

Optimal architecture. Define welfare as $W(\mathcal{A}) = \Pi^{\mathcal{A}} - \lambda \cdot L^{\mathcal{A}}$, where $\Pi^{\mathcal{A}}$ is steady-state throughput, $L^{\mathcal{A}}$ is expected disruption cost, and λ is the shock arrival rate. Centralization achieves higher throughput

($\Pi^C = 1/\mu > \Pi^I = 1/(\mu+\bar{g})$) but faces disruption costs that grow linearly in system size ($L^C = (N-j)\cdot d \cdot c_d$). There exists a threshold λ^* above which the disruption costs of centralization outweigh its throughput advantage (Proposition 8).

3.2 Predictions

The model generates four predictions that I take to the data. Each prediction maps to a specific proposition in Appendix A.

Prediction 1 (Differential Persistence)**.** *Following a disruption, centrally coordinated zones exhibit persistent delays because the algorithm's zero-buffer scheduling leaves no slack to absorb perturbations. Zones where independently coordinated agents share infrastructure recover to baseline because these agents' behavioral variability creates buffer that absorbs propagating delays.* (Proposition 4.)

Prediction 2 (Exogenous Buffer)**.** *Any mechanism that exogenously introduces buffer into a centralized system—such as environmental conditions that force larger vehicle spacing—eliminates the persistence of post-shock delays. The model's fragility stems from the absence of buffer, not from algorithmic coordination per se: if $\bar{g} > 0$ for any reason, delays decay.* (Corollary 4.)

Prediction 3 (Tighter Optimization Worsens Fragility)**.** *As the coordinating algorithm reduces buffer to increase throughput, the system becomes more efficient in steady state but more fragile following disruptions. Recovery time $E[T_R] \approx d\mu/\bar{g} + d$ is decreasing in buffer $\bar{g}$, so as $\bar{g} \to 0$ recovery time diverges even as throughput rises.* (Corollary 3.)

Prediction 4 (Efficiency–Resilience Tradeoff)**.** *There exists a disruption frequency λ^* above which the cumulative cost of propagating delays exceeds the throughput gain from centralized coordination. This threshold is decreasing in system size N and shock severity d: larger, disruption-prone systems face a tighter constraint on the returns to centralization.* (Proposition 8.)

4 Empirical Setting and Data

4.1 Container Terminal Operations

Container ports are among the most operationally intensive environments in global logistics. A single large terminal may process thousands of containers per day, with each container requiring a coordinated sequence of operations: unloading from a vessel by a quay crane, horizontal transport from the quay to a storage yard, stacking and retrieval at the yard, and loading onto outbound trucks or trains (Carlo et al., 2014;

Stahlbock and Voß, 2008). The horizontal transport step—moving containers between quay cranes and storage yards—is the focus of this study.

At conventional terminals, horizontal transport is performed by human-operated trucks or straddle carriers. At automated terminals, autonomous guided vehicles (AGVs) perform the same function under algorithmic control. A centralized dispatching system assigns each AGV to a task (a specific container movement), determines its route, coordinates its timing relative to other vehicles, and manages its interactions with cranes and yard equipment. The algorithm's objective is to minimize total transport time and maximize throughput by coordinating the movements of all vehicles simultaneously.

Container terminals provide an attractive empirical setting for studying the dynamics of automated operations for several reasons. Operations are highly structured and repetitive, producing large volumes of fine-grained data. The physical environment is bounded and well-documented. Disruption events (mechanical faults, equipment failures) occur frequently and can be precisely timed. And the operational consequences of disruptions—delays in task completion—can be measured at the level of individual transport runs.

4.2 The Focal Terminal

The terminal studied in this paper is a major container-handling facility within one of the world's largest port complexes. It operates around the clock and processes millions of containers annually. The focal company's AGV fleet operates at SAE Level 4 autonomy—fully autonomous within the terminal's operational domain, with no human driver onboard—representing the highest level of autonomous driving technology in commercial deployment.[2] The terminal uses a hybrid operational model in which these AGVs, operated under centralized algorithmic control, share the operational environment with human-operated trucks belonging to other logistics companies.

The terminal's physical layout is organized around a quay wall served by 22 quay cranes. Behind the quay wall, six transit bridges connect the quay-side operations to the storage yard. These bridges are the primary corridors through which all horizontal transport vehicles—both AGVs and human-operated trucks—must pass when moving containers between cranes and the yard. All six bridges are physically identical: they have the same number of lanes, the same width, and the same design. This physical uniformity is critical for the identification strategy described below.

The key institutional feature that generates the variation exploited in this paper is the ship assignment process. The port authority assigns arriving ships to logistics companies based on berth availability and contractual arrangements, not on the basis of the companies' operational characteristics. This means that

[2]The SAE J3016 standard defines six levels of driving automation (0–5). Level 4 ("High Automation") denotes systems capable of performing all driving tasks without human intervention within a defined operational design domain. Level 5 ("Full Automation"), which would require no environmental restrictions, has not been achieved in any commercial setting.

on any given day, the mix of AGVs and human-operated trucks operating in a given area of the terminal is determined by which companies have been assigned ships at nearby berth positions—a source of variation that is plausibly exogenous to the operational conditions within the terminal.

Crucially, the human-operated trucks and the AGV fleet serve different ships and operate under entirely separate organizational objectives. Each logistics company is assigned its own ships by the port authority, manages its own fleet, and is evaluated against its own performance targets. Human truck drivers do not transport containers for the focal company, nor do the AGVs transport containers for the human-operated companies. The two populations share physical infrastructure—transit bridges, lanes, road space—but not tasks, loads, or coordination systems.

This assignment process creates a spatial gradient in vehicle composition. Some cranes and their associated transit bridges are predominantly served by the focal company's AGV fleet on most days, while others are predominantly served by human-operated trucks from other logistics companies. I classify operational zones as "automated" (those where the focal company's AGVs account for the majority of operations on most days) and "mixed" (those where human-operated trucks from other companies are prevalent). The classification is based on the time-series average of vehicle composition, not on any single day's realization.

4.3 Data

The primary dataset comprises the complete operational records of the focal company's AGV fleet over 364 consecutive days—approximately 780,000 individual transport runs. Each observation is a single transport run: one container moved from an origin to a destination by a specific AGV. The data include task identifiers (vehicle, origin zone, destination zone, associated crane), precise timestamps (start and end), the resulting task duration in seconds, a speed profile capturing total time spent stationary, and mechanical fault records with occurrence timing and duration.

Table 1 reports summary statistics for the estimation sample, which is restricted to quay-lane tasks (the primary focus of the analysis) with task completion times between 0 and 2,500 seconds. The sample contains 179,925 task-level observations from 53 AGVs operating across 45 quay-lane zones over 357 days.

Several features of Table 1 are worth noting. Mean task completion times are similar across zone types (598 versus 601 seconds), indicating that automated and mixed zones do not differ substantially in baseline operational difficulty. Shock characteristics are also comparable: mean shock duration is 231 seconds in automated zones and 227 seconds in mixed zones. I confirm below (Section 6.1) that the distributions of shock severity are statistically indistinguishable across zone types.

Table 1: Summary Statistics

	Automated Zones		Mixed Zones		Full Sample
	Mean	SD	Mean	SD	Mean
Panel A: Task characteristics					
Task completion time (seconds)	598.1	495.8	601.2	494.5	605.4
Time stationary (seconds)	255.6	374.4	246.2	366.8	253.0
Median task completion time (seconds)	487.6		499.6		
Panel B: Shock characteristics					
Shock duration (seconds)	230.7	248.2	226.9	236.7	
Number of shocks (≥60s)	1,738		867		
Panel C: Sample dimensions					
Observations	61,292		30,557		179,925
Unique AGVs	53		53		53
Unique cranes	4		7		
Unique zones	8		14		45
Days	269		167		357

Notes: Sample restricted to quay-lane tasks with task completion times between 0 and 2,500 seconds. Automated zones are crane positions where the focal company's AGVs account for the majority of operations on most days. Mixed zones are crane positions where human-operated trucks are prevalent. The full sample includes all quay-lane tasks; automated and mixed zone columns are subsets used in the comparative analysis, with the remainder comprising zones not classified into either category. Task completion time is total elapsed time for one container transport task. Time stationary is total time the AGV spent at zero speed. Shock duration is the length of a mechanical fault, conditional on the fault lasting at least 60 seconds.

4.4 Defining Disruptions

The disruption events used in this paper are operational faults experienced by individual AGVs during task execution. A fault is triggered when the AGV's onboard monitoring system detects an anomaly—such as a sensor misalignment, motor irregularity, or degraded connectivity with the centralized dispatch system—and the vehicle halts to perform internal recalibration. Once recalibration is complete, the vehicle resumes normal operation autonomously; these are not catastrophic breakdowns requiring human intervention or towing. The faults are recorded with precise timestamps and durations.

I define a shock as a fault with a duration of at least 60 seconds. This threshold excludes momentary sensor corrections that create negligible operational impact while retaining events of sufficient duration to generate congestion. The main analyses use faults between 60 and 120 seconds—moderate disruptions that temporarily block a lane while the vehicle recalibrates—to maintain a homogeneous set of shocks. These are conservative in the sense that they represent the mildest category of operationally meaningful disruptions; robustness analyses using all faults of 60 seconds or longer, including more severe events, produce qualitatively identical results (Appendix C).

Operational faults are well suited to serve as exogenous disruption events. They are caused by internal equipment conditions—mechanical wear, sensor drift, intermittent connectivity loss—not by traffic conditions or operational decisions. Their occurrence is unpredictable: fault timing is determined by the state of the

vehicle's onboard systems, which is orthogonal to current congestion at the zone where the fault occurs. And they create localized congestion: a halted AGV blocks the lane it occupies, forcing following vehicles to wait or reroute.

5 Empirical Approach

5.1 Event Study Design

To trace the dynamic response of task completion times to disruption events, I employ an event study design. For each shock (mechanical fault of at least 60 seconds) occurring at a given operational zone on a given day, I identify all other AGV tasks completed at the same zone on the same day—excluding the faulted vehicle itself. I then compute the time elapsed between each task's start time and the shock's occurrence, and assign each task to a discrete time bin based on this elapsed time.

The baseline specification uses 10-minute bins spanning a window from 60 minutes before the shock to 120 minutes after. The bin immediately preceding the shock ($[-10, 0)$ minutes) serves as the reference category. I estimate the following regression:

$$y_{izt} = \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[\text{bin}_{izt} = k] + \alpha_z + \delta_d + \gamma_h + \varepsilon_{izt} \tag{2}$$

where y_{izt} is the task duration (in seconds) for task i at zone z on day t; $\mathbf{1}[\text{bin}_{izt} = k]$ is an indicator for the task falling in time bin k relative to the shock; α_z are zone fixed effects; δ_d are day fixed effects; γ_h are hour-of-day fixed effects; and ε_{izt} is the error term. Standard errors are heteroskedasticity-robust (HC1).

The coefficients β_k trace the time path of task durations relative to the reference period. Pre-shock coefficients ($k < -1$) serve as a placebo test: if the event study is well specified, task durations in the minutes before a shock should not differ systematically from the reference period. Post-shock coefficients ($k \geq 0$) capture the dynamic response to the disruption.

When a task could be matched to multiple shocks on the same zone-day, I apply a two-step deduplication procedure. First, for each combination of vehicle, task start time, and shock time, I retain the observation with the shortest time between task and shock. Second, for each combination of vehicle and task start time, I retain the observation associated with the closest remaining shock. This ensures that each task is matched to exactly one shock, and that the match reflects the most proximate disruption.

For analyses involving the time-stationary outcome (time spent at zero speed) and the weather interaction, I use 20-minute bins to maintain adequate statistical power per cell, with the window and reference category adjusted accordingly.

5.2 Identification

The identifying assumption is that, conditional on zone, day, and hour-of-day fixed effects, the timing of mechanical faults is as good as random with respect to other vehicles' task durations. Zone fixed effects absorb all time-invariant differences between operational areas, including permanent differences in physical layout, average traffic intensity, and proximity to bottlenecks. Day fixed effects absorb port-wide daily variation in activity levels driven by ship arrivals, weather conditions, and overall demand. Hour-of-day fixed effects absorb systematic intra-day patterns such as shift changes and peak-hour congestion.

The remaining variation is within a zone, within a day, within an hour: how task durations differ in the specific minutes following a mechanical fault compared to the same zone's baseline during the same day and time of day. The key threat to identification would be if mechanical faults systematically occur at moments when the zone is already experiencing unusual congestion. Because faults are caused by internal mechanical malfunctions of the AGV (sensor failures, motor issues, communication errors) rather than by traffic conditions, this threat is mitigated. Discussions with terminal operators and the focal company's technical team confirm that mechanical faults arise from equipment wear and component failure, not from operational conditions at the time of the fault.[3] The flat pre-trends in all reported specifications provide further reassurance.

5.3 Comparing Automated and Mixed Zones

The core analysis compares event study estimates across automated and mixed zones. I estimate Equation 2 separately for each zone type and compare the post-shock coefficient paths. I also estimate a pooled specification with interactions between the bin indicators and an automated-zone indicator, which allows direct testing of whether the recovery paths differ statistically:

$$y_{izt} = \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[\text{bin} = k] + \sum_{k \neq -1} \phi_k \cdot \mathbf{1}[\text{bin} = k] \times \text{Auto}_z + \alpha_z + \delta_d + \gamma_h + \varepsilon_{izt} \tag{3}$$

where Auto_z is an indicator for automated zones. The coefficients ϕ_k capture the difference in the event study path between automated and mixed zones at each time bin. A joint F-test of the ϕ_k coefficients for the post-shock bins provides a formal test of differential recovery.

I restrict the estimation sample to tasks with task completion times between 0 and 2,500 seconds, which

[3] An independent source of evidence for fault exogeneity comes from geomagnetic storms. AGVs rely on GPS and electronic sensors for navigation; human-operated trucks do not. During severe geomagnetic storms (G4 on the NOAA scale), which degrade GPS accuracy and disrupt radio communications, AGV fault rates in automated zones increase by 32%, while fault rates in mixed zones—where human trucks predominate—are unchanged. This differential response, which exhibits a monotonic dose-response across storm severity levels, is consistent with faults being caused by electromagnetic interference with onboard systems rather than by traffic conditions. See Appendix D for details.

excludes a small number of extreme outliers (less than 0.5% of observations) that likely reflect data recording errors or exceptional operational circumstances. For specifications using the time-stationary outcome, I winsorize at the 99th percentile to limit the influence of extreme values while retaining the full sample.

6 Analysis

This section tests the four predictions developed in Section 3. Each subsection maps one prediction to a specific empirical test, using the event study framework described in Section 5. The results are presented in the order of the predictions: differential persistence, exogenous buffer, algorithmic tightening, and the efficiency–resilience tradeoff.

6.1 Differential Persistence (Prediction 1)

Prediction 1 states that disruptions propagate persistently in tightly coupled automated systems but dissipate in loosely coupled mixed systems, because tight sequential coordination transmits perturbations downstream while independent agents absorb them locally. I test this by comparing post-shock recovery paths between automated and mixed zones within the same port.

Core result

Figure 1 presents the core finding. Following a mechanical fault, task completion times in automated zones rise sharply and remain elevated throughout the two-hour observation window—30 to 100 seconds above baseline even 60–120 minutes after the initial disruption. In mixed zones, delays of comparable initial magnitude dissipate within approximately 70 minutes, with task times returning to and crossing below the baseline. The divergence in recovery paths is the central empirical pattern this paper seeks to explain.

Tables 2 and 3 report the underlying coefficients with progressively saturated fixed effects. For automated zones (Table 2), the impact coefficient (the $[0, +10)$ minute bin) is 173.1 seconds in column (1), with no fixed effects, and 167.5 seconds in column (4), the preferred specification with zone, day, and hour-of-day fixed effects. Late-bin coefficients (40–120 minutes post-shock) average 75.0 seconds in the preferred specification ($p < 0.01$ for all late bins). The stability of estimates across columns—impact and late-bin patterns are essentially unchanged as fixed effects are added—indicates that the event study is well identified: the fixed effects absorb meaningful variation in task durations (R^2 increases from 0.011 to 0.097) while leaving the post-shock coefficients intact.

For mixed zones (Table 3), the contrast is pronounced. The preferred specification shows an impact coefficient of 135.6 seconds ($p < 0.01$), comparable to automated zones. But late-bin coefficients are uniformly

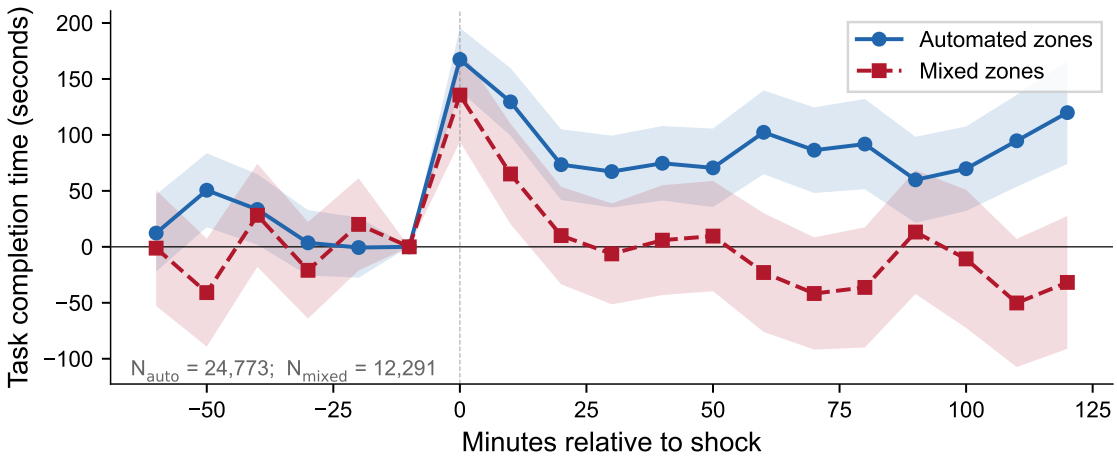


Figure 1: Post-Shock Dynamics of Task Duration: Automated versus Mixed Zones

Notes: Event study coefficients from the preferred specification (zone, day, and hour-of-day fixed effects). Dependent variable is task completion time (seconds). Shocks: mechanical faults of 60–120 seconds. 10-minute bins; reference bin $[-10, 0)$. Shaded areas: 95% confidence intervals (HC1).

small and statistically insignificant: the $[+50, +60)$ bin is 9.6 seconds ($p > 0.10$), the $[+80, +90)$ bin is -36.3 seconds, and the $[+110, +120)$ bin is -50.2 seconds ($p < 0.10$). The late-bin average is -17.3 seconds—a slight *over-correction* below baseline. This pattern is stable across specifications: even without any fixed effects (column 1), mixed zones show no significant persistence beyond the initial 20-minute window.

To gauge economic significance, note that the mean task completion time is approximately 615 seconds in automated zones. The preferred-specification late-bin average of 75.0 seconds represents a 12.2% increase above the mean, sustained for over an hour. Given that the terminal processes thousands of container movements per day, even modest per-task delays aggregate to substantial throughput losses.

Robustness

I subject the main finding to several robustness checks, reported in Appendix C. The result is robust to log-transforming the dependent variable, confirming that outliers do not drive the level result. Broadening the shock definition to include all mechanical faults of 60 seconds or longer leaves the recovery divergence unchanged. A formal interaction specification (Equation 3) yields pre-shock interaction coefficients near zero and positive, significant late-bin interactions, with the joint F-test rejecting equal recovery at $p < 0.001$ and the pre-trend test failing to reject ($p > 0.10$). A leave-one-out exercise confirms that no single crane drives the result: automated late-bin averages range from +28 to +49 seconds across permutations. The distributions of shock duration are statistically indistinguishable across zone types (Mann-Whitney $p = 0.19$), and the persistence pattern holds in both day and night subsamples, ruling out the possibility that the result is driven by time-of-day variation in human truck presence.

Table 2: Post-Shock Task Duration: Automated Zones

	(1)	(2)	(3)	(4)
$[0, +10)$ min	173.1***	172.6***	168.8***	167.5***
	(14.8)	(14.8)	(14.4)	(14.4)
$[+10, +20)$	113.5***	114.3***	132.4***	129.5***
	(16.0)	(16.0)	(15.4)	(15.4)
$[+20, +30)$	65.9***	67.3***	77.4***	73.4***
	(16.7)	(16.6)	(16.1)	(16.1)
$[+30, +40)$	63.9***	61.0***	72.0***	67.2***
	(17.0)	(16.9)	(16.4)	(16.4)
$[+40, +50)$	71.2***	68.3***	81.3***	74.7***
	(17.7)	(17.6)	(17.0)	(17.0)
$[+50, +60)$	61.9***	59.1***	76.3***	70.6***
	(18.4)	(18.3)	(18.0)	(17.9)
$[+60, +70)$	90.5***	84.7***	107.0***	102.3***
	(19.3)	(19.3)	(19.0)	(19.1)
$[+70, +80)$	85.5***	77.9***	95.8***	86.4***
	(19.7)	(19.7)	(19.5)	(19.5)
$[+80, +90)$	101.3***	94.2***	101.9***	91.8***
	(20.8)	(20.7)	(20.4)	(20.6)
$[+90, +100)$	67.7***	63.7***	71.0***	59.9***
	(20.3)	(20.2)	(19.7)	(19.6)
$[+100, +110)$	64.5***	58.2***	83.3***	69.8***
	(20.1)	(20.0)	(19.2)	(19.3)
$[+110, +120)$	100.8***	96.1***	103.7***	94.8***
	(22.1)	(21.8)	(21.0)	(21.0)
Zone FE	No	Yes	Yes	Yes
Day FE	No	No	Yes	Yes
Hour FE	No	No	No	Yes
Observations	24,773	24,773	24,773	24,773
R^2	0.011	0.022	0.090	0.097
Mean dep. var.	614.9	614.9	614.9	614.9

Notes: Dependent variable is task completion time (seconds). Shocks are mechanical faults of 60–120 seconds. Reference bin is $[-10, 0)$ minutes. Pre-shock bins ($[-60, -10)$) are jointly insignificant in all specifications ($p > 0.10$). HC1 robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Persistence is also concentrated in physically constrained environments. In quay lanes—narrow corridors where vehicles queue in single file with no alternative path—shocks produce persistent delays (+33 to +55 seconds in the late bins, $N = 71{,}499$). In yard blocks, where multiple rows and access points permit rerouting, late-bin coefficients are *negative* (−92 to −136 seconds, $N = 14{,}450$), indicating that vehicles reroute to less congested areas (Appendix C, Figure 11). The concentration of persistence in sequentially constrained settings implicates the sequential structure of coordination rather than overall capacity.

Mechanism: propagation through sequential queuing

The model predicts propagation through sequential queuing—perturbations propagate because downstream agents must wait for upstream agents to clear. Consistent with this, decomposing post-shock delays reveals that the persistence is driven by increased *waiting* time (time at zero speed), not increased *driving* time.

Table 3: Post-Shock Task Duration: Mixed Zones

	(1)	(2)	(3)	(4)
$[0, +10)$ min	136.7***	140.6***	134.4***	135.6***
	(22.7)	(22.4)	(21.4)	(21.3)
$[+10, +20)$	76.7***	85.1***	67.0***	65.1***
	(24.2)	(24.0)	(22.9)	(22.9)
$[+20, +30)$	10.1	25.0	12.9	10.1
	(23.4)	(23.2)	(22.3)	(22.2)
$[+30, +40)$	−1.5	6.0	−7.8	−6.3
	(23.9)	(23.7)	(23.0)	(23.0)
$[+40, +50)$	13.8	13.0	4.8	5.9
	(25.8)	(25.9)	(24.8)	(25.0)
$[+50, +60)$	15.1	19.4	14.8	9.6
	(25.9)	(25.6)	(25.0)	(25.1)
$[+60, +70)$	−17.4	−8.6	−15.6	−22.9
	(27.7)	(27.3)	(27.1)	(27.1)
$[+70, +80)$	−20.4	−15.6	−31.1	−41.7
	(26.3)	(26.0)	(25.3)	(25.6)
$[+80, +90)$	−40.5	−26.2	−27.3	−36.3
	(27.8)	(27.5)	(27.2)	(27.4)
$[+90, +100)$	33.2	37.8	28.0	13.2
	(29.2)	(28.9)	(28.0)	(28.2)
$[+100, +110)$	29.5	30.0	7.2	−10.8
	(31.4)	(31.3)	(31.0)	(31.5)
$[+110, +120)$	−24.1	−32.9	−33.6	−50.2*
	(29.0)	(28.9)	(28.6)	(29.2)
Zone FE	No	Yes	Yes	Yes
Day FE	No	No	Yes	Yes
Hour FE	No	No	No	Yes
Observations	12,291	12,291	12,291	12,291
R^2	0.008	0.026	0.108	0.115
Mean dep. var.	639.4	639.4	639.4	639.4

Notes: Same specification as Table 2, estimated on the mixed-zone subsample. Pre-shock bins are jointly insignificant in all specifications.

Table 4 reports event study estimates for the time-stationary outcome in automated zones with progressive fixed effects. The impact bin is +136.5 seconds in the preferred specification, and late-bin coefficients remain positive through the observation window (R^2 increases from 0.020 to 0.086 as fixed effects are added).

Table 4: Time Stationary (Speed = 0): Automated Zones

	(1)	(2)	(3)	(4)
$[0, +20)$ min	133.7***	134.0***	137.1***	136.5***
	(6.3)	(6.3)	(6.1)	(6.1)
$[+20, +40)$	22.5***	22.9***	30.8***	27.7***
	(6.7)	(6.7)	(6.6)	(6.6)
$[+40, +60)$	9.7	6.3	16.4**	12.7*
	(7.2)	(7.2)	(7.1)	(7.1)
$[+60, +80)$	34.0***	30.6***	36.3***	30.7***
	(8.3)	(8.2)	(8.2)	(8.3)
$[+80, +100)$	21.2**	17.1**	23.0***	13.7
	(8.6)	(8.6)	(8.5)	(8.6)
$[+100, +120)$	6.7	1.9	10.7	0.2
	(9.5)	(9.4)	(9.2)	(9.3)
Zone FE	No	Yes	Yes	Yes
Day FE	No	No	Yes	Yes
Hour FE	No	No	No	Yes
Observations	41,750	41,750	41,750	41,750
R^2	0.020	0.030	0.081	0.086

Notes: Dependent variable: time at zero speed (seconds), winsorized at the 99th percentile. All shocks $\geq$ 60s. 20-minute bins; reference bin $[-20, 0)$. HC1 robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2 presents the preferred specification for both zone types. In automated zones, post-shock waiting time increases persistently—vehicles spend more time stationary, queuing behind congestion that the algorithm's tight sequencing cannot resolve. In mixed zones, the time-stationary coefficients turn negative in the late post-shock bins (−12.8 to −56.3 seconds), indicating that vehicles encounter *less* waiting than usual following a disruption—consistent with independently operating agents clearing congestion points or creating gaps in traffic.

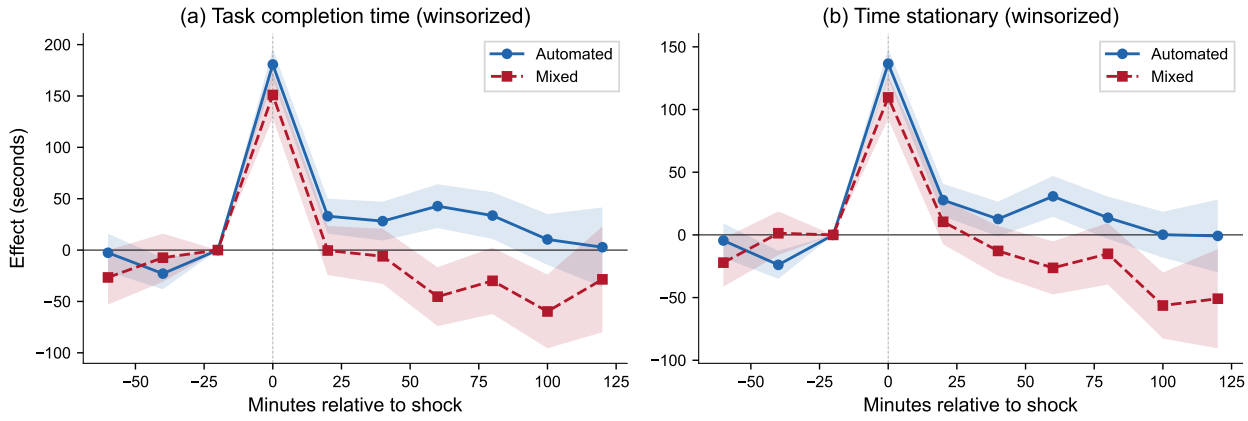


Figure 2: Decomposing Post-Shock Delays: Total Duration and Time Stationary

Notes: Preferred specification for two outcomes: task completion time (winsorized p99) and time at zero speed (winsorized p99). All shocks $\geq$ 60s, 20-minute bins.

Density moderation

The model predicts that cumulative disruption cost is linear in the number of downstream agents, so persistence should be amplified when more agents share the constrained pathway. I test this by splitting the sample based on pre-shock operational density (above- versus below-median task count in the prior 60 minutes). Figure 3 presents the results. In automated zones during high-density periods, the late-bin average reaches +110.3 seconds—nearly three times the full-sample average and approximately 18% of mean task duration. Automated zones during low-density periods show more modest persistence (+27.5 seconds). Mixed zones show limited persistence even at high density (+14.2 seconds) and over-correction at low density (−35.2 seconds). The resilience of mixed zones at high density is informative: even when many vehicles are present, the system absorbs disruptions, suggesting that the *coordination structure*, not the *number of vehicles*, determines fragility.

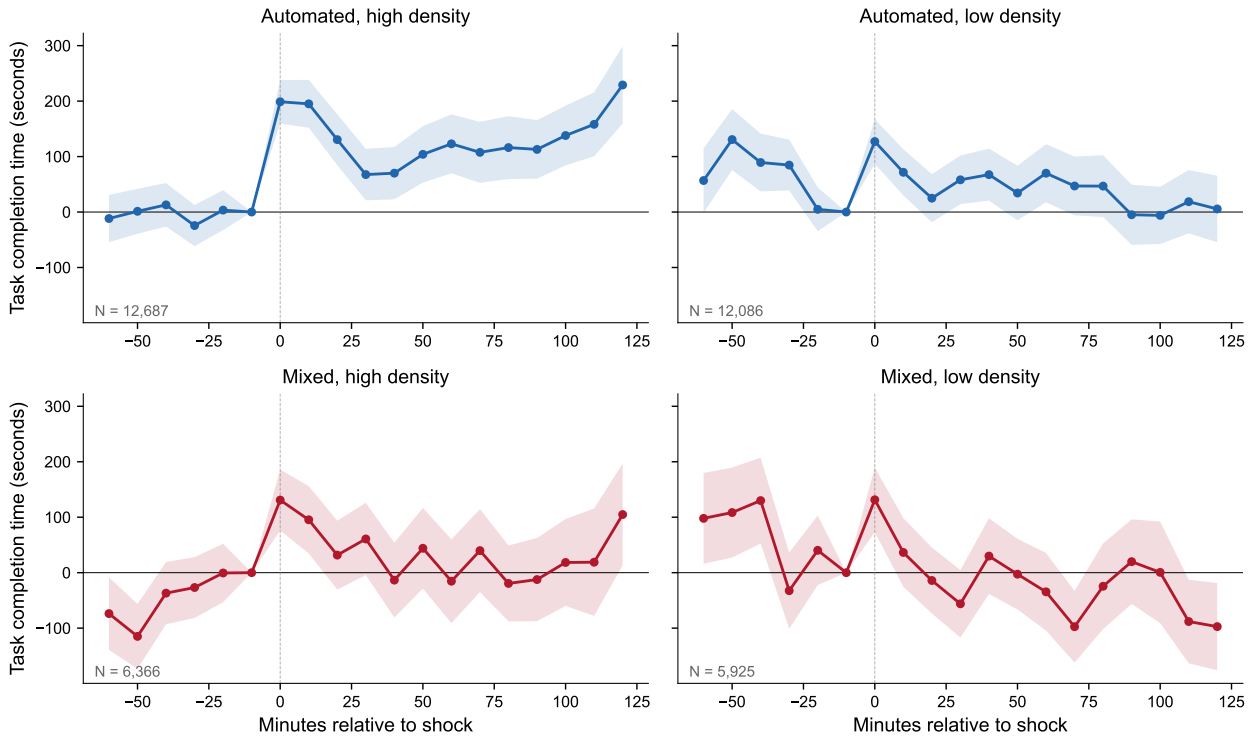


Figure 3: Density Moderation: Pre-Shock Task Count × Zone Type

Notes: Preferred specification estimated separately for four cells defined by zone type and pre-shock density (above/below median tasks in prior 60 minutes). 60–120s shocks, 10-minute bins.

6.2 Exogenous Buffer (Prediction 2)

Prediction 2 states that exogenously introducing temporal buffer between sequential operations eliminates the persistence of disruptions in automated systems, because the mechanism relies on tight temporal coupling

rather than on any intrinsic property of the technology. I test this using weather variation: heavy rainfall forces all vehicles to reduce speed, mechanically increasing time gaps between sequential operations and thereby loosening the tight temporal coupling that the algorithm normally maintains.

I estimate the preferred specification separately for four cells defined by zone type and precipitation. Under dry conditions, automated zones show characteristic persistence: the $[+60, +80)$ bin is +40.0 seconds ($p < 0.01$). Under heavy rain, the same automated zones *recover*: late-bin coefficients turn negative, reaching -73.1 seconds at $[+100, +120)$ ($p < 0.05$). Mixed zones show the complementary pattern: recovery under dry conditions (-18.5 to -69.9 seconds in late bins) and volatile estimates under rain, where the subsample is smaller ($N = 2{,}041$). Table 6 in Appendix C reports the full coefficient estimates.

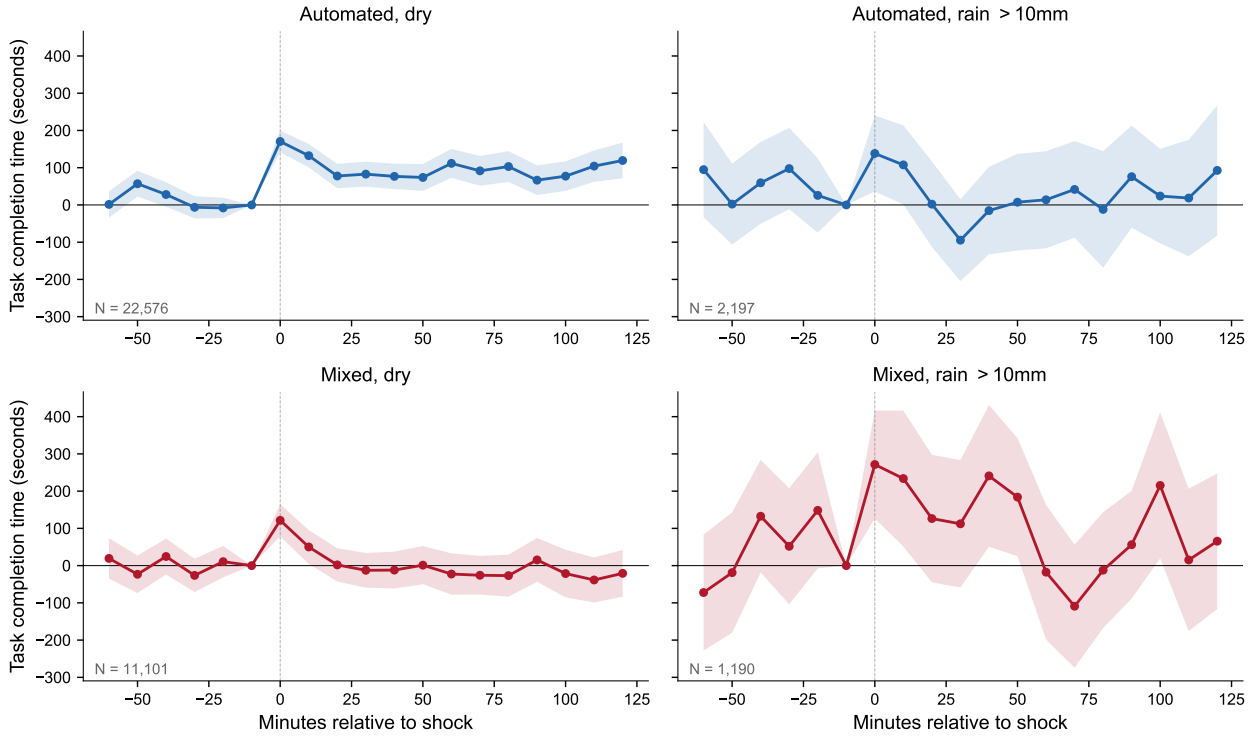


Figure 4: Weather Moderation: Post-Shock Recovery Under Dry vs Rainy Conditions

Notes: Preferred specification estimated separately for four cells defined by zone type and precipitation. Dependent variable: task completion time (seconds), winsorized p99. Rain: daily precipitation > 10mm. 60–120s faults, 10-minute bins, reference bin $[-10, 0)$. HC1 robust standard errors.

Figure 4 visualizes these results. The weather comparison is informative because it involves variation *within* automated zones—the same infrastructure, the same fleet, the same algorithm—under different exogenous conditions. Rain does not improve the algorithm or change the vehicles. It forces a reduction in operating speed, which increases time gaps between sequential operations—exogenously loosening the tight temporal coupling that the algorithm's optimization creates. That this introduction of temporal buffer eliminates the propagation pattern provides the most direct evidence that the mechanism is coupling structure,

not technology, infrastructure, or capacity. The algorithm eliminates buffer time in pursuit of efficiency; rain reintroduces it; and the system's behavior transforms from fragile to resilient.

6.3 Algorithmic Tightening (Prediction 3)

Prediction 3 states that as the coordinating algorithm improves over time, it tightens coupling by reducing slack between operations, thereby increasing steady-state efficiency but simultaneously worsening post-shock recovery. I test this by splitting the sample into calendar quarters and examining whether productivity gains and recovery deterioration co-occur within automated zones over the year.

Figure 5 documents the productivity trajectory. The algorithm's throughput—measured as completed container tasks per AGV per hour—increases from 9.8 in Q1 to 14.2 in Q4, a 45% improvement. Routing efficiency, measured as the ratio of loaded to empty travel distance, improves from 1.65 in Q1 to 3.30 in Q4, indicating that the algorithm eliminates an increasing share of deadhead trips. Both metrics reflect a progressively tighter optimization of vehicle coordination: shorter inter-task gaps, less idle time, and more direct routing between sequential operations.

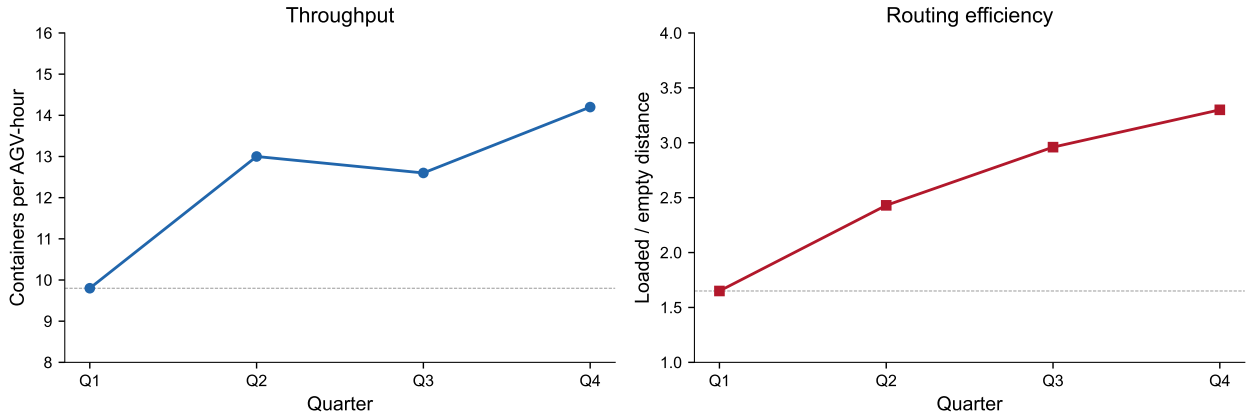


Figure 5: Productivity Improvement in Automated Zones Over the Sample Period

Notes: Left panel: completed container tasks per AGV per hour. Right panel: ratio of loaded to empty travel distance (routing efficiency). Automated zones only. Quarterly averages with 95% confidence intervals.

Figure 6 presents the event study estimates by quarter. The pattern is striking. In Q1, when the algorithm is least optimized, the late-bin average in automated zones is +71 seconds. By Q4, the late-bin average has grown to +122 seconds—a 72% increase in post-shock persistence coinciding with a 45% improvement in throughput. The pre-shock coefficients remain flat across all quarters, confirming that the deterioration in recovery is a post-shock phenomenon rather than a shift in baseline task durations.

This co-movement is consistent with the model's prediction that efficiency and resilience are jointly determined by the degree of coupling. As the algorithm learns to pack operations more tightly—reducing buffer

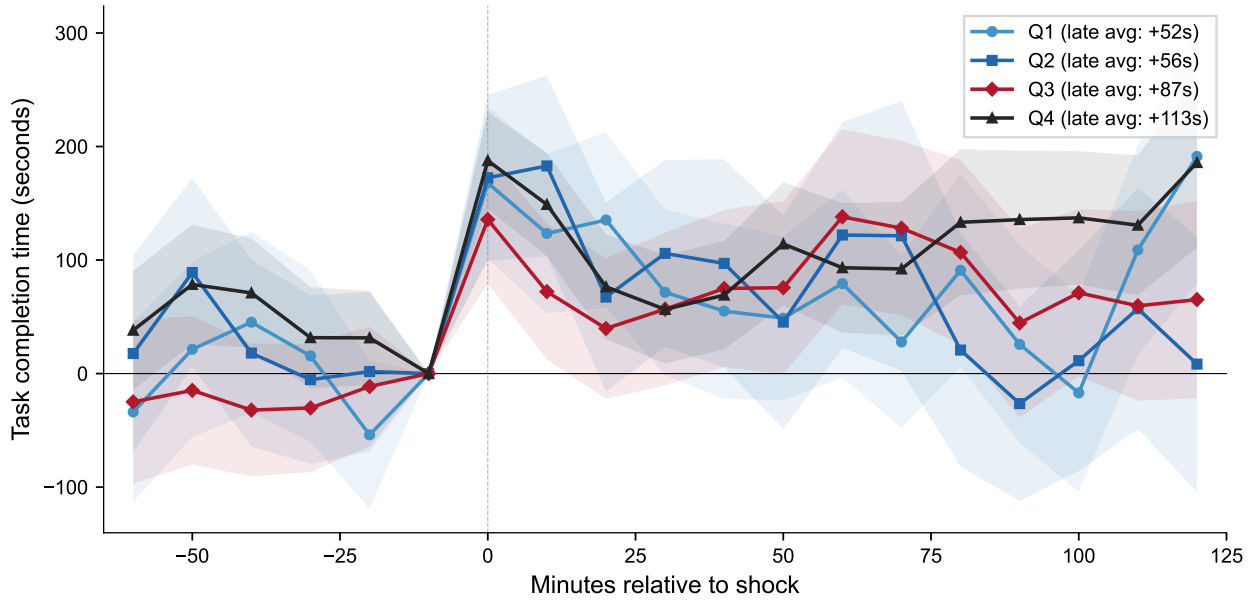


Figure 6: Post-Shock Recovery by Quarter: Automated Zones

Notes: Preferred specification estimated separately by calendar quarter. Automated zones only. 60–120s shocks, 10-minute bins, reference bin $[-10, 0)$. Shaded areas: 95% confidence intervals (HC1).

time between sequential tasks—it simultaneously removes the temporal slack that would absorb perturbations. The system moves along the efficiency–resilience frontier: each improvement in throughput comes at the cost of greater vulnerability to disruption. Neither the port authority nor the autonomous vehicle company were aware of this dynamic; the algorithm's optimization objective does not incorporate a resilience constraint, and its improvement is measured solely in throughput terms. The deterioration in recovery is an emergent property of the optimization, not a deliberate design choice.

6.4 Efficiency–Resilience Tradeoff (Prediction 4)

Prediction 4 states that there exists a threshold disruption rate above which the throughput gains from zero-buffer scheduling are offset by propagating delays, and that this threshold can be empirically calibrated. Using the event study estimates, I quantify the costs of tight coupling and compare them to the efficiency gains the algorithm achieves.

The calibration proceeds as follows. In automated zones, the preferred-specification late-bin average is +75.0 seconds, sustained across eight post-shock bins (40–120 minutes). Over the full two-hour observation window, the cumulative excess delay per shock is the sum of all twelve post-shock bin coefficients multiplied by the average number of tasks per bin. The automated zones process approximately 4.2 tasks per 10-minute bin. Summing the preferred-specification coefficients ($167.5 + 129.5 + 73.4 + \ldots + 94.8 = 1{,}087.7$ seconds of excess delay per task) and multiplying by the average tasks per bin yields approximately 4,568 task-seconds of total

propagation per shock, or roughly 1.27 AGV-hours. With a more conservative estimate focusing on the late bins only (where the causal interpretation is cleanest), each shock generates approximately 2.8 AGV-hours of propagating delay in automated zones. In mixed zones, the analogous calculation yields approximately 0.14 AGV-hours—twenty times less—because the late-bin coefficients are near zero or negative.

The automated bridges experience an average of 3.6 shocks per day (mechanical faults in the 60–120 second range). At 2.8 AGV-hours per shock, the total daily propagation cost is approximately 10.1 AGV-hours. Over the 364-day sample, this amounts to roughly 3,670 AGV-hours of propagating delay attributable to tight coupling. For a fleet of 49 AGVs operating 24 hours per day, this represents approximately 0.85% of total available AGV-hours—a nontrivial fraction, especially when concentrated in specific zones and time periods.

The breakeven condition asks: how much time must the algorithm save per task to justify the propagation cost of zero-buffer scheduling? The automated zones complete approximately 6,200 tasks per day. For the daily propagation cost of 10.1 AGV-hours (36,360 seconds) to be offset, the algorithm must save at least $36{,}360/6{,}200 \approx 5.9$ seconds per task through tighter scheduling. Given that the Q1-to-Q4 throughput improvement (from 9.8 to 14.2 tasks per AGV-hour) implies time savings well in excess of this threshold, the current calibration suggests that tighter coupling is *net* efficient on average. However, the tradeoff becomes less favorable as disruption rates increase: at double the observed shock rate, the breakeven rises to approximately 11.8 seconds per task, and the propagation cost begins to consume a larger share of the efficiency gain. The fragility cost is not prohibitive at observed disruption rates, but it is not negligible, and its nonlinear dependence on shock frequency means that the system operates closer to the breakeven threshold than the port authority recognizes.

The calibration carries an additional implication. Because the algorithm does not internalize resilience costs, it optimizes past the socially efficient coupling threshold whenever the private benefit of tighter scheduling (faster throughput) exceeds the marginal propagation cost—a condition that depends on the disruption rate. At the observed rate of 3.6 shocks per day, the algorithm is well within the efficient region. But disruption rates are not stationary: equipment aging, weather events, and operational changes can shift the rate upward. Without a resilience constraint, the algorithm has no mechanism to pull back from tight coupling when conditions deteriorate—it will continue to optimize for throughput, increasing the system's exposure to disruption propagation precisely when the environment becomes more volatile.

7 Discussion

This paper asked what happens to an interdependent system when AI-driven coordination succeeds. The answer is that the system becomes more fragile—not because the algorithm fails, but because the optimal solution eliminates the slack that would otherwise absorb disruptions. The throughput-maximizing schedule drives temporal buffer between sequential operations toward zero, and any delay in a zero-buffer system propagates through the entire chain. The empirical evidence confirms this at every level of analysis. Automated zones exhibit persistent delays; mixed zones recover. Heavy rainfall reintroduces buffer and eliminates persistence. The algorithm's own improvement trajectory traces the efficiency–resilience frontier in real time: each increment of throughput is accompanied by a measurable deterioration in recovery capacity.

7.1 Theoretical Contributions

From task-level augmentation to system-level coordination. The empirical literature on AI has made important progress in measuring task-level effects—productivity gains in customer service, consulting, writing, and software development (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026; Noy and Zhang, 2023; Peng et al., 2023), and AI's capacity to replicate certain benefits of human teamwork (Dell'Acqua et al., 2025). But organizations do not deploy AI on isolated tasks. They deploy it on systems of interdependent operations—supply chains, production lines, logistics networks—where the performance of the system cannot be inferred from the performance of its components.

The present findings show that this distinction matters. The algorithm improves the throughput of each individual transport task; by any task-level metric, performance has improved. But at the system level, the same optimization creates the zero-buffer temporal structure through which disruptions propagate. A standard task-level evaluation would conclude that the algorithm is unambiguously beneficial. A system-level evaluation reveals that it has simultaneously introduced a fragility that did not previously exist. This is not a peculiarity of container ports—any setting in which AI optimizes sequences of interdependent operations faces the same structural tension. Recent work has identified invisible failures and reliability gaps even in task-level AI deployment (Potts and Sudhof, 2026; Rabanser et al., 2026); the present paper shows that these concerns become qualitatively different at the system level, where the failures are not in the AI's task execution but in the system architecture that task-level optimization creates.

A limit to the case for AI-enabled centralization. A growing literature argues that AI shifts the optimal organizational form toward centralization. AI's prediction capabilities remove the information-processing constraints that historically favored decentralized structures (Agrawal et al., 2022), and the known

costs of centralization—search failure, communication distortion, evaluation error—diminish as AI relaxes the bounded rationality that produces them (Csaszar, 2025; Csaszar and Jia, 2026). The implication is that as AI improves, the case for centralization becomes overwhelming.

The findings reported here identify a cost that this prediction overlooks. The dispatching algorithm does not fail to find the throughput-maximizing schedule; it finds it with precision. The cost arises because the optimal schedule is itself the source of fragility. Zero-buffer scheduling chains operations into sequences in which any disruption mechanically propagates to every subsequent task. This is not a problem that better algorithms can solve—it is a structural property of the optimal solution. The known costs of centralization *decrease* in the optimizer's capability; the propagation cost identified here *increases* in it. Over the sample period, throughput rises by 45 percent while recovery deteriorates by 72 percent. The costs and benefits of AI-driven centralization share a common source, and advancing AI capability pushes both in the same direction (Agrawal et al., 2018; Iansiti and Lakhani, 2020).

A parallel dynamic has been identified in a different domain: AI systems that substitute for human cognitive effort may erode the learning externalities that sustain collective knowledge, potentially tipping the economy into a steady state in which the stock of general knowledge vanishes despite high-quality personalized advice (Acemoglu et al., 2026). In that setting, the eroded resource is collective knowledge; in this one, it is temporal buffer. In both, AI optimizes away something that appears wasteful from the perspective of the individual task—learning effort, idle time—but that is structurally load-bearing for the broader system. And in both, welfare is non-monotone in the quality of optimization: better AI can make the system worse off.

Quantifying the efficiency–resilience tradeoff. The third contribution is empirical. The algorithm's own improvement trajectory over the sample period traces the efficiency–resilience frontier, providing a direct mapping between the degree of optimization and the system's vulnerability to disruption. This mapping yields a concrete decision criterion—the breakeven disruption rate above which centralization is net negative—that organizations can calibrate to their own settings.

The finding that the algorithm does not internalize resilience costs speaks to a broader theme in the economics of AI: the effects of AI on productivity depend critically on how it is deployed and what objectives it optimizes (Acemoglu et al., 2022). The algorithm optimizes the right objective from the firm's perspective (throughput), but the resulting system architecture generates costs that the objective function does not capture. As AI capabilities expand and single-provider solutions become increasingly attractive on efficiency grounds, the costs that escape the optimization—here, system-level fragility—may grow precisely because the optimization itself is succeeding (Agrawal et al., 2022).

7.2 Practical Implications

The findings carry implications for three sets of decisions. First, organizations assessing the returns to AI-driven coordination should evaluate system-level resilience alongside task-level performance. The efficiency gains from algorithmic coordination are real and valuable, but they coexist with system-level fragility that becomes apparent only when disruptions occur—which, in operational settings, they inevitably do. Standard evaluations that focus on mean performance will overstate the net benefit by missing the tail costs that the same optimization creates.

Second, the decision this terminal's port authority faces—whether to consolidate operations under a single algorithmic provider or maintain multiple independent operators—is structurally the same decision that organizations across industries will confront as AI coordination capabilities improve. A hospital deciding whether to unify surgical scheduling, patient transport, and resource allocation under one AI system faces the same tradeoff. A logistics company choosing between a single fleet management platform and multiple competing carriers faces it as well. In each case, consolidation under one coordinating intelligence promises efficiency gains through global optimization, but it simultaneously eliminates the buffer that independent coordination regimes introduce at the boundaries between subsystems (Iansiti and Lakhani, 2020; Jacobides et al., 2021).

Third, algorithmic design choices that preserve temporal buffer between operations—even at the cost of some throughput—may substantially improve system resilience. The weather finding suggests that relatively modest reductions in operating intensity can eliminate the propagation dynamics that drive persistent delays. Organizations could implement buffer policies algorithmically, deliberately spacing sequential operations to create absorption capacity. The efficiency–resilience frontier identified in this paper provides a framework for calibrating such policies: the breakeven disruption rate above which centralization is net negative offers a concrete decision criterion.

7.3 Limitations

Several limitations should be acknowledged. First, the study observes a single terminal and a single company's AGV fleet. While the within-terminal design provides strong internal validity—comparing the same technology, same algorithm, same infrastructure across zones that differ in vehicle composition—the external validity of the specific effect magnitudes is uncertain. Different terminals, algorithms, and operational environments may produce different magnitudes of disruption persistence.

Second, I observe only the AGV fleet's operations, not the movements of human-operated trucks. The mechanism through which human presence enables recovery is inferred from the patterns in AGV data, not

from direct observation of human behavior. A richer dataset that included both human and automated vehicle trajectories would allow more direct testing of the buffering mechanism and would permit distinguishing whether resilience stems from human judgment specifically or from independence of coordination more generally.

Third, while the event study design with rich fixed effects provides strong evidence for the persistence of disruptions and differential recovery across zone types, the mechanism analysis relies on progressively ruling out alternatives rather than directly measuring buffer time between operations. The buffer interpretation is the most parsimonious explanation consistent with the full set of findings, but it remains an inference rather than a direct measurement.

Finally, the theoretical argument that independence of coordination—rather than humanness per se—drives resilience is an implication of the framework rather than a directly tested proposition. Testing it would require observing systems with multiple competing algorithms operating in the same physical space, a setting that the present data do not provide but that emerging multi-operator autonomous systems may make available.

7.4 Conclusion

As organizations place more operations under unified AI control, the case for centralization appears overwhelming: algorithms can optimize what humans cannot, and the efficiency gains are substantial. But this paper shows that the case is incomplete. The same optimization that generates efficiency also eliminates the slack that buffers the system against disruption. Unlike the classical costs of centralization—which diminish as AI improves—this cost intensifies. Organizations that evaluate AI-driven coordination only by its steady-state performance will systematically underestimate its true cost, because the fragility the algorithm creates is invisible until a disruption reveals it.

The findings do not argue against AI-driven coordination. They argue for a more complete accounting of what it does. Better AI makes individual tasks faster and more productive. It also makes the system of tasks more tightly coupled and more fragile. Whether centralization is net beneficial depends not only on the efficiency gains—which are visible, measurable, and impressive—but also on the disruption costs, which are intermittent, hard to predict, and easy to overlook. As AI capabilities continue to expand and the temptation to consolidate grows, understanding this tradeoff becomes not merely an academic exercise but an organizational imperative.

References

Daron Acemoglu and Simon Johnson. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. PublicAffairs, New York, 2023.

Daron Acemoglu, David Autor, Jonathon Hazell, and Pascual Restrepo. Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340, 2022.

Daron Acemoglu, Dingwen Kong, and Asuman Ozdaglar. AI, human cognition and knowledge collapse. (34910), 2026. doi: 10.3386/w34910.

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press, 2018.

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Power and Prediction: The Disruptive Economics of Artificial Intelligence*. Harvard Business Review Press, Boston, 2022.

Ricardo Alonso, Wouter Dessein, and Niko Matouschek. When does coordination require centralization? *American Economic Review*, 98(1):145–179, 2008.

Erik Brynjolfsson and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton & Company, New York, 2014.

Erik Brynjolfsson, Tom Mitchell, and Daniel Rock. What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108:43–47, 2018. doi: 10.1257/pandp.20181019.

Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. Generative AI at work. *The Quarterly Journal of Economics*, page qjae044, 2025. doi: 10.1093/qje/qjae044.

Héctor J Carlo, Iris FA Vis, and Kees Jan Roodbergen. Transport operations in container terminals: Literature overview, trends, research directions and classification scheme. *European Journal of Operational Research*, 236(1):1–13, 2014.

Sergio Correia. REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. *Statistical Software Components*, 2017.

Felipe A Csaszar. Unbounding rationality: Why AI is a fundamental issue for strategy. *SSRN Working Paper*, 2025.

Felipe A Csaszar and Nan Jia. *Handbook of Artificial Intelligence and Strategy*. Edward Elgar Publishing, Cheltenham, 2026.

Richard M Cyert and James G March. *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ, 1963.

Fabrizio Dell'Acqua, Charles Ayoubi, Hila Lifshitz-Assaf, Raffaella Sadun, Ethan R Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, and Karim R Lakhani. The cybernetic teammate: A field experiment on generative AI reshaping teamwork and expertise. 2025. doi: 10.2139/ssrn.5188231. Harvard Business School Working Paper 25-043.

Fabrizio Dell'Acqua, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine C Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Organization Science*, 2026. doi: 10.1287/orsc.2025.21838.

Mert Demirer, John J Horton, Nicole Immorlica, Brendan Lucier, and Peyman Shahidi. Chaining tasks, redefining work: A theory of AI automation. (34859), 2026. doi: 10.3386/w34859.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. GPTs are GPTs: Labor market impact potential of large language models. *Science*, 384(6702):1306–1308, 2024. doi: 10.1126/science.adj0998.

Edward Felten, Manav Raj, and Robert Seamans. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12): 2195–2217, 2021. doi: 10.1002/smj.3286.

Kevin B Hendricks, Vinod R Singhal, and Rongbing Zhang. The effect of operational slack, diversification, and vertical relatedness on the stock market reaction to supply chain disruptions. *Journal of Operations Management*, 27(3):233–246, 2009.

Marco Iansiti and Karim R Lakhani. Competing in the age of AI: Strategy and leadership when algorithms and networks run the world. 2020.

Michael G Jacobides, Stefano Brusoni, and François Candelon. The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science*, 6(4):412–435, 2021. doi: 10.1287/stsc.2021.0148.

Michael G Jacobides, Stefano Brusoni, and François Candelon. How AI reshapes the competitive landscape. *Harvard Business Review*, 102(2):98–107, 2024.

Joshua J Kovach, Manpreet Hora, Andrew Manikas, and Pankaj C Patel. Firm performance in dynamic environments: The role of operational slack and operational scope. *Journal of Operations Management*, 37:1–12, 2015.

Nitin Nohria and Ranjay Gulati. Is slack good or bad for innovation? *Academy of Management Journal*, 39 (5):1245–1264, 1996.

Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023. doi: 10.1126/science.adh2586.

Chibuike Oguh, Danilo Masoni, and Medha Singh. Selloff wipes out nearly $1 trillion from software and services stocks as investors debate AI's existential threat. *Reuters*, February 2026.

Bradley Olson. The week Anthropic tanked the market and pulled ahead of its rivals. *The Wall Street Journal*, February 2026.

Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint arXiv:2302.06590*, 2023.

Christopher Potts and Moritz Sudhof. Invisible failures in human–AI interactions. 2026. doi: 10.48550/arXiv.2603.15423. arXiv:2603.15423.

Stephan Rabanser, Sayash Kapoor, Peter Kirgis, Kangheng Liu, Saiteja Utpala, and Arvind Narayanan. Towards a science of AI agent reliability. *arXiv preprint arXiv:2602.16666*, 2026. doi: 10.48550/arXiv.2602.16666.

Jan W Rivkin and Nicolaj Siggelkow. Balancing search and stability: Interdependencies among elements of organizational design. *Management Science*, 49(3):290–311, 2003.

Raaj Kumar Sah and Joseph E Stiglitz. The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review*, 76(4):716–727, 1986.

Robert Stahlbock and Stefan Voß. Operations research at container terminals: A literature update. *OR Spectrum*, 30(1):1–52, 2008.

A Formal Model

This appendix develops the model summarized in Section 3.1. It derives the key endogenous variable—buffer time between consecutive tasks—from primitive features of the coordination architecture: information structure and cost internalization. It defines a symmetric equilibrium of the scheduling game under independent coordination, establishes existence and uniqueness, characterizes the comparative statics of the equilibrium buffer, and traces the consequences for disruption propagation.

A.1 Environment

A corridor of unit capacity serves N agents performing sequential tasks. At most one agent occupies the corridor at a time. Agents are indexed by service order, $i = 1, 2, \ldots, N$.

Each task requires $\mu > 0$ time units. Agent i enters the corridor at time $\max(a_i, F_{i-1})$, where a_i is the agent's chosen arrival time and F_{i-1} is the exit time of agent $i-1$. Agent i's exit time is:

$$F_i = \max(a_i, F_{i-1}) + \mu. \tag{4}$$

Disruptions are rare events analyzed separately: Section A.7 introduces an exogenous corridor blockage of duration $d > 0$ at a given position j, and Section A.11 parameterizes their arrival rate $\lambda > 0$.

A.2 The Scheduling Game

Each agent i chooses an arrival time a_i based on available information $\mathcal{I}_i$. The coordination regime determines what information is available and when the decision is made. Arriving early ($a_i < F_{i-1}$) wastes the agent's time; arriving late ($a_i > F_{i-1}$) leaves the corridor idle, delaying all downstream operations.

Players. N agents in sequence.

Strategies. Each agent i chooses a measurable function $a_i : \mathcal{I}_i \to \mathbb{R}$ mapping its information set to an arrival time.

Payoffs. Let $w > 0$ denote the per-unit-time cost of waiting and $v > 0$ the per-unit-time cost of corridor idleness borne by the decision-maker. Agent i's cost is:

$$C_i(a_i) = w \cdot (F_{i-1} - a_i)^+ + v \cdot (a_i - F_{i-1})^+ \tag{5}$$

where $(x)^+ \equiv \max(x, 0)$. The decision-maker choosing a_i minimizes $\mathbb{E}[C_i(a_i) \mid \mathcal{I}_i]$. The coordination regime determines both $\mathcal{I}_i$ and v.

A.3 Coordination Regimes

A.3.1 Centralized Coordination

Under centralized coordination, a single algorithm manages all N agents. The coordinator dispatches agents dynamically, choosing a_i after observing the corridor state.

Assumption 1 (Full information)**.** *The centralized coordinator observes F_{i-1} before choosing a_i: $F_{i-1} \in \mathcal{I}_i$.*

Assumption 2 (Internalized externality)**.** *The coordinator internalizes the full social cost of corridor idleness. One unit of idle corridor time delays all $N-i$ remaining agents, imposing cost $V_i = (N-i) \cdot v_0$, where $v_0 > 0$ is the per-agent unit delay cost.*

A.3.2 Independent Coordination

Under independent coordination, each agent is managed by a separate decision process—its own dispatcher, algorithm, or judgment. The agent must commit to an arrival time a_i before F_{i-1} is realized, based on a noisy forecast.

Assumption 3 (Imprecise information)**.** *Under independent coordination, agent i chooses a_i before observing F_{i-1}. Its information set is $\mathcal{I}_i = \{\sigma_i\}$, where*

$$\sigma_i = F_{i-1} + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \tau^2),$$

with $\tau > 0$ and η_i independent across agents.

Assumption 4 (Private costs)**.** *Under independent coordination, agent i's decision-maker bears only the private cost of late arrival: $v = v_0$. It does not internalize the externality that corridor idleness imposes on agents managed by other organizations. The cost of waiting satisfies $w > v_0$.*

Remark. *Assumptions 1–2 and 3–4 describe coordination architectures, not technologies. A second autonomous fleet under a different algorithm satisfies Assumptions 3–4: it lacks real-time access to the first fleet's corridor state ($\tau > 0$) and optimizes its own objective ($v = v_0$). The model's predictions apply identically whether independent agents are human or automated.*

A.4 Equilibrium

A.4.1 Centralized Optimum

Under Assumption 1, F_{i-1} is known when a_i is chosen.

Proposition 1 (Zero buffer under centralization)**.** *Under centralized coordination (Assumptions 1–2), the unique optimal arrival time is $a_i^* = F_{i-1}$, yielding zero buffer:*

$$b_i^C \equiv a_i^* - F_{i-1} = 0 \quad \textit{for all } i.$$

Proof. For $a_i < F_{i-1}$: $C_i = w(F_{i-1} - a_i) > 0$. For $a_i > F_{i-1}$: $C_i = V_i(a_i - F_{i-1}) > 0$. For $a_i = F_{i-1}$: $C_i = 0$. Since $C_i \geq 0$ with equality only at $a_i = F_{i-1}$, the minimizer is unique. □

A.4.2 Independent Equilibrium

Under Assumption 3, agent i observes $\sigma_i = F_{i-1} + \eta_i$ with $\eta_i \sim \mathcal{N}(0, \tau^2)$. With a diffuse (improper uniform) prior on F_{i-1}, the posterior is:

$$F_{i-1} \mid \sigma_i \;\sim\; \mathcal{N}(\sigma_i,\, \tau^2). \tag{6}$$

Remark (Robustness of the diffuse prior)**.** *The diffuse prior is standard when the agent's prior uncertainty about F_{i-1} is large relative to signal precision τ. If instead $F_{i-1} \sim \mathcal{N}(m, s^2)$ for finite s^2, the posterior converges to (6) as $s^2 \to \infty$. All results hold exactly under the diffuse prior and approximately under any proper prior with $s^2 \gg \tau^2$.*

Proposition 2 (Existence and uniqueness)**.** *Under Assumptions 3–4 with $w > v_0$, there exists a unique symmetric equilibrium. The equilibrium strategy is:*

$$a^*(\sigma_i) = \sigma_i + \tau\, \Phi^{-1}\!\left(\frac{w}{w + v_0}\right), \tag{7}$$

where Φ^{-1} is the standard normal quantile function. Moreover, a^* is optimal for agent i regardless of the strategies used by other agents.

Proof. Optimality. Given the posterior (6), agent i's expected cost is:

$$C(a_i \mid \sigma_i) = w \int_{a_i}^{\infty} (F - a_i)\, \frac{1}{\tau}\phi\!\left(\frac{F - \sigma_i}{\tau}\right) dF \;+\; v_0 \int_{-\infty}^{a_i} (a_i - F)\, \frac{1}{\tau}\phi\!\left(\frac{F - \sigma_i}{\tau}\right) dF.$$

The second derivative is:

$$\frac{\partial^2 C}{\partial a_i^2} = (w + v_0)\, \frac{1}{\tau}\phi\!\left(\frac{a_i - \sigma_i}{\tau}\right) > 0,$$

so C is strictly convex. The unique minimizer satisfies:

$$\Phi\!\left(\frac{a_i^* - \sigma_i}{\tau}\right) = \frac{w}{w + v_0}. \tag{8}$$

Inverting gives (7).

Independence. Under a diffuse prior, the posterior depends only on σ_i and τ—not on other agents' strategies (which affect F_{i-1} only through its unconditional distribution). The optimal strategy is thus independent of other agents' strategies.

Uniqueness. Since C is strictly convex, the best response is unique for every agent. Since the best response does not depend on other agents' strategies, any symmetric equilibrium must prescribe a^*. □

A.5 Endogenous Buffer

The equilibrium buffer—the gap between agent i's arrival and the corridor becoming free—is:

$$b_i \;\equiv\; a^*(\sigma_i) - F_{i-1} \;=\; \underbrace{(\sigma_i - F_{i-1})}_{\eta_i} \;+\; \underbrace{\tau\,\Phi^{-1}\!\left(\frac{w}{w+v_0}\right)}_{\bar{b}}.$$

Since $\eta_i \sim \mathcal{N}(0, \tau^2)$:

Corollary 1 (Buffer distribution)**.** *In the unique symmetric equilibrium under independent coordination:*

$$b_i \sim \mathcal{N}(\bar{b},\, \tau^2), \tag{9}$$

$$\bar{b} \equiv \tau\,\Phi^{-1}\!\left(\frac{w}{w+v_0}\right) > 0. \tag{10}$$

The positivity of $\bar{b}$ follows from $w > v_0 \;\Rightarrow\; w/(w+v_0) > 1/2 \;\Rightarrow\; \Phi^{-1}(w/(w+v_0)) > 0$.

The buffer arises from two features of independent coordination, both absent under centralization:

(i) **Information imprecision** ($\tau > 0$). Unable to perfectly predict F_{i-1}, the agent leaves a precautionary gap. As $\tau \to 0$, $\bar{b} \to 0$.

(ii) **Cost asymmetry** ($w > v_0$). The private cost of waiting exceeds the private cost of corridor idleness. Under centralization, $V_i \gg w$ reverses the asymmetry.

Corollary 2 (Buffer is architectural, not technological)**.** *The buffer $\bar{b}$ depends on coordination architecture (shared vs. separate information, joint vs. private optimization), not on whether agents are human or automated. A second autonomous fleet under a different algorithm satisfies Assumptions 3–4 and produces $\bar{b} > 0$.*

A.6 Comparative Statics

Proposition 3 (Comparative statics of equilibrium buffer)**.** *Let* $\bar{b} = \tau\Phi^{-1}(w/(w+v_0))$*,* $q = w/(w+v_0)$*, and* $z = \Phi^{-1}(q)$*. Then:*

(i) $\dfrac{\partial \bar{b}}{\partial \tau} = z > 0$*: buffer is strictly increasing in information imprecision.*

(ii) $\dfrac{\partial \bar{b}}{\partial w} = \dfrac{\tau\, v_0}{(w+v_0)^2\, \phi(z)} > 0$*: buffer is strictly increasing in the cost of waiting.*

(iii) $\dfrac{\partial \bar{b}}{\partial v_0} = \dfrac{-\tau\, w}{(w+v_0)^2\, \phi(z)} < 0$*: buffer is strictly decreasing in the private delay cost.*

Proof. Part (i): $\partial\bar{b}/\partial\tau = \Phi^{-1}(q) = z > 0$, since $q > 1/2$.

Parts (ii)–(iii): By the chain rule, $\partial\bar{b}/\partial w = \tau \cdot (\partial z/\partial q) \cdot (\partial q/\partial w)$. Since $z = \Phi^{-1}(q)$, $\partial z/\partial q = 1/\phi(z) > 0$. And $\partial q/\partial w = v_0/(w+v_0)^2 > 0$, giving part (ii). Similarly, $\partial q/\partial v_0 = -w/(w+v_0)^2 < 0$, giving part (iii). □

A.7 Disruption Propagation

Suppose agent j experiences an exogenous disruption that blocks the corridor for $d > 0$ additional time units. Agent j's exit time shifts from F_j to $F_j + d$.

Define the *residual delay* to agent $j+k$ as Δ_{j+k}.

A.7.1 Centralized Coordination

Lemma 1 (Persistent delays)**.** *Under centralized coordination ($b_i^C = 0$ for all i), the delay to agent $j+k$ is:*

$$\Delta_{j+k}^C = d \quad \text{for all } k \geq 1.$$

The total cumulative delay is $(N-j)\cdot d$, unbounded in queue length N.

Proof. Agent $j+1$ was scheduled at $a_{j+1}^* = F_j$ (Proposition 1). The disruption shifts agent j's exit to $F_j + d$. Since $a_{j+1}^* = F_j < F_j + d$, agent $j+1$ waits d time units: $\Delta_{j+1}^C = d$. By induction on k: if $\Delta_{j+k}^C = d$, then agent $j+k$ exits at $F_{j+k} + d$, agent $j+k+1$ arrives at F_{j+k}, and waits d, giving $\Delta_{j+k+1}^C = d$. The cumulative delay is $\sum_{k=1}^{N-j} d = (N-j)\cdot d$. □

Remark. *Reoptimization after observing the disruption cannot eliminate persistence. Under zero buffer, no slack exists to absorb the d time units of lost corridor capacity. Reoptimization can redistribute delay across agents but cannot reduce the total cumulative delay below $(N-j)\cdot d$.*

A.7.2 Independent Coordination

Under independent coordination, agent $j+1$'s pre-set arrival time provides buffer $g_{j+1} = (b_{j+1})^+ = (\eta_{j+1} + \bar{b})^+$ above the pre-disruption exit time F_j.

Lemma 2 (Attenuating delays)**.** *Under independent coordination ($\bar{b} > 0$), the delay sequence satisfies:*

$$\Delta^I_{j+1} = (d - g_{j+1})^+, \qquad \Delta^I_{j+k+1} = (\Delta^I_{j+k} - g_{j+k+1})^+, \quad k \geq 1, \tag{11}$$

where $g_i = (b_i)^+$ and $b_i \sim \mathcal{N}(\bar{b}, \tau^2)$ i.i.d. The delay sequence converges to zero almost surely, with absorption at zero in finite time.

Proof. Recursion. Agent $j+k+1$ had buffer $g_{j+k+1} \geq 0$. The buffer absorbs $\min(\Delta^I_{j+k}, g_{j+k+1})$ of the delay. The residual is $(\Delta^I_{j+k} - g_{j+k+1})^+$.

Monotonicity. Since $g_{j+k+1} \geq 0$: $\Delta^I_{j+k+1} \leq \Delta^I_{j+k}$ for all k.

Finite-time absorption. Define $p_0 = \Pr(g_i \geq d) = 1 - \Phi\left(\frac{d-\bar{b}}{\tau}\right) > 0$. The probability that delay remains positive through K consecutive steps satisfies $(1-p_0)^K \to 0$. Therefore, $\Delta^I_{j+k} = 0$ for some finite k, almost surely. □

A.7.3 Differential Persistence and Recovery Time

Proposition 4 (Differential persistence)**.** *For any disruption $d > 0$:*

(i) Under centralized coordination: $\mathbb{E}[\Delta^C_{j+k}] = d$ for all $k \geq 1$.

(ii) Under independent coordination: $\mathbb{E}[\Delta^I_{j+k}] \to 0$ as $k \to \infty$.

(iii) Let $K^* = \inf\{k \geq 1 : \Delta^I_{j+k} = 0\}$ and $\bar{g} \equiv \mathbb{E}[(b_i)^+] > 0$. The expected number of agents required for full recovery satisfies:

$$\frac{d}{\bar{g}} \;\leq\; \mathbb{E}[K^*] \;\leq\; \frac{d}{\bar{g}} + \frac{\mathbb{E}[g^2]}{2\bar{g}^2}, \tag{12}$$

where $\mathbb{E}[g^2] = \mathbb{E}[((b_i)^+)^2]$ is finite.

(iv) The expected recovery time satisfies:

$$\mathbb{E}[T^I_R] \leq \left(\frac{d}{\bar{g}} + \frac{\mathbb{E}[g^2]}{2\bar{g}^2}\right) \cdot (\mu + \bar{g}). \tag{13}$$

Proof. Parts (i)–(ii) follow from Lemmas 1 and 2.

Part (iii): *Lower bound.* The delay absorbed at step k is $\min(\Delta^I_{j+k-1}, g_{j+k}) \le g_{j+k}$. Summing: $S_{K^*} \ge d$. By Wald's identity, $\mathbb{E}[K^*] \ge d/\bar{g}$.

Upper bound. Define $T' = \inf\{k : \sum_{\ell=1}^{k} g_{j+\ell} \ge d\}$. By induction, $\Delta^I_{j+k} \le (d - S_k)^+$, so $K^* \le T'$. By Wald's identity and Lorden's inequality: $\mathbb{E}[T'] \le d/\bar{g} + \mathbb{E}[g^2]/(2\bar{g}^2)$.

Part (iv): $T^I_R = K^*\mu + S_{K^*} - d$. By Wald's identity, $\mathbb{E}[T^I_R] \le \mathbb{E}[K^*](\mu + \bar{g})$. □

Corollary 3 (Recovery time comparative statics)**.** *The leading term of the recovery bound is $d\mu/\bar{g} + d$, which is decreasing in $\bar{g}$ (and therefore decreasing in $\bar{b}$, τ, and w/v_0) whenever $\mu > 0$.*

A.8 Physical Constraints

Definition 1 (Corridor type)**.**

- *Constrained corridor ($r = 0$): a single-path corridor. Agents must queue sequentially.*
- *Unconstrained corridor ($r = 1$): a corridor with alternative paths. Agents can reroute at cost $c_r > 0$.*

Proposition 5 (Physical constraints and persistence)**.**

(i) In unconstrained corridors with $c_r < w \cdot d$, agents reroute. Delays are not transmitted downstream.

(ii) In constrained corridors, sequential queuing is mandatory and propagation results apply.

(iii) Differential persistence is concentrated in constrained corridors.

A.9 Exogenous Buffer

Under centralized coordination, $\bar{b} = 0$. Suppose an environmental factor introduces a mandatory minimum headway $h > 0$ between consecutive corridor entries. The coordinator optimally sets $a_i = F_{i-1} + h$.

Proposition 6 (Exogenous buffer eliminates persistence)**.** *Under centralized coordination with mandatory headway $h \ge 0$, the residual delay is:*

$$\Delta^C_{j+k}(h) = (d - k \cdot h)^+.$$

(i) When $h = 0$: $\Delta^C_{j+k} = d$ for all k (persistent).

(ii) When $h > 0$: delays attenuate linearly and are fully absorbed after $\lceil d/h \rceil$ steps.

Proof. With headway h, agent $j+1$ is scheduled at $F_j + h$. If agent j's exit is delayed by d, agent $j+1$'s delay is $(d - h)^+$. By induction, $\Delta^C_{j+k}(h) = (d - kh)^+$. □

Corollary 4. *The fragility of centralized coordination stems from the absence of buffer ($h = 0$, $\bar{b} = 0$), not from algorithmic coordination per se. Any mechanism that reintroduces buffer restores the capacity to absorb disruptions.*

A.10 Mixed Coordination

Definition 2 (Mixed coordination)**.** *In a mixed system, each agent is independently drawn as centralized (probability $\alpha \in [0, 1]$, zero buffer) or independent (probability $1 - \alpha$, buffer $g_i = (b_i)^+$ with $b_i \sim \mathcal{N}(\bar{b}, \tau^2)$).*

Proposition 7 (Resilience of mixed systems)**.**

(i) For $\alpha = 1$, delays persist indefinitely.

(ii) For $\alpha < 1$, delays attenuate and reach zero in finite time a.s.

(iii) The expected recovery time satisfies:

$$\mathbb{E}[T_R^M] \leq \left(\frac{d}{\bar{g}} + \frac{\mathbb{E}[g^2]}{2\bar{g}^2}\right) \cdot \left(\frac{\mu}{1-\alpha} + \bar{g}\right), \tag{14}$$

which is finite for $\alpha < 1$ and strictly decreasing in $(1 - \alpha)$.

Proof. Part (ii): the delay transmits fully through consecutive centralized agents. When an independent agent appears (probability $1 - \alpha$ at each step), its buffer absorbs part or all of the delay. The probability that no independent agent appears among the first K positions is $\alpha^K \to 0$. Absorption occurs in finite time a.s.

Part (iii): the subsequence of delays at independent agents follows the same recursion as the pure independent case. The expected number of independent agents needed, K_{ind}, satisfies $\mathbb{E}[K_{\text{ind}}] \leq d/\bar{g} + \mathbb{E}[g^2]/(2\bar{g}^2)$. Total agents: $\mathbb{E}[K_M] = \mathbb{E}[K_{\text{ind}}]/(1 - \alpha)$. Recovery time: $\mathbb{E}[T_R^M] = \mathbb{E}[K_{\text{ind}}](\mu/(1 - \alpha) + \bar{g}) - d$. □

A.11 Optimal Architecture

Centralized coordination achieves throughput $\Pi^C = 1/\mu$. Independent coordination achieves $\Pi^I = 1/(\mu + \bar{g})$. The throughput gain from centralization is $\Pi^C - \Pi^I = \bar{g}/[\mu(\mu + \bar{g})] > 0$.

Disruptions arrive at rate $\lambda > 0$. Let $c_d > 0$ denote the per-unit delay cost. For a disruption at position j:

- Centralized: $L^C = (N - j) \cdot d \cdot c_d$, linear in N.
- Independent: $L^I \leq c_d \cdot d \cdot (d/\bar{g} + \mathbb{E}[g^2]/(2\bar{g}^2))$, finite and independent of N.

An operator choosing between regimes maximizes $W(A) = \Pi^A - \lambda L^A$.

Proposition 8 (Optimal architecture)**.**

(i) When $\lambda = 0$, centralization is strictly optimal.

(ii) When $\lambda > 0$, there exists a threshold:

$$\lambda^* = \frac{\Pi^C - \Pi^I}{L^C - L^I} > 0.$$

Centralization is optimal if and only if $\lambda < \lambda^*$. The threshold is decreasing in disruption severity d and system size N. As $N \to \infty$, $\lambda^* \to 0$.

(iii) For N sufficiently large, the algorithm's zero-buffer schedule is suboptimal whenever $\lambda > 0$: introducing deterministic buffer $b^* > 0$ strictly improves expected payoff.

Proof. Part (ii): $W(C) - W(I) = (\Pi^C - \Pi^I) - \lambda(L^C - L^I)$ is strictly decreasing in λ, positive at $\lambda = 0$, and crosses zero at λ^*. Since L^C grows in N while $\Pi^C - \Pi^I$ and L^I are bounded, λ^* is decreasing in N and $\lambda^* \to 0$ as $N \to \infty$.

Part (iii): with buffer $b > 0$, the expected payoff is $W(b) = 1/(\mu + b) - \lambda c_d d^2/(2b) + O(d)$. The marginal benefit at $b \to 0^+$ diverges to $+\infty$, so there exists $b^* > 0$ with $W(b^*) > W(0)$. □

Remark (Why the algorithm chooses zero buffer)**.** *Proposition 1 shows the algorithm sets $b = 0$ to maximize throughput—its programmed objective. Proposition 8(iii) shows this is socially suboptimal whenever disruptions occur. The gap arises because the algorithm's objective omits disruption costs. The efficiency–resilience tradeoff is a consequence of this incomplete objective, not of coordination per se.*

B Variable Definitions

C Additional Robustness

This appendix reports robustness checks for the main finding of differential persistence between automated and mixed zones. All specifications use the preferred fixed effects structure (zone, day, hour-of-day) with HC1 robust standard errors unless otherwise noted.

Alternative outcome measures. Figure 7 re-estimates the main specification using log task completion time as the dependent variable. The recovery divergence is qualitatively identical, confirming that the level result is not driven by outliers in the right tail of the duration distribution.

Table 5: Variable Definitions

Variable	Definition
`run_duration`	Task completion time: total elapsed time (seconds) from task start to task completion for a single container transport run. Includes driving time, waiting time, and any delays.
`speed0_duration`	Total time (seconds) during a task run that the AGV's speed was recorded as zero. Captures queuing and waiting time.
`stop_fault_duration`	Duration (seconds) of a mechanical fault event. Recorded by the AGV's onboard monitoring system when a malfunction is detected.
`zone`	The operational zone where the task originates or terminates. Quay-side zones are identified by crane position (left or right lane); yard-side zones are identified by block and row.
`automated`	Indicator variable equal to 1 if the zone is classified as automated (the focal company's AGVs account for the majority of operations on most days).
`event_bin`	Time bin (10- or 20-minute intervals) indicating the elapsed time between a task's start and the nearest shock at the same zone on the same day.
`pre_shock_density`	Count of AGV tasks initiated at the zone in the 60 minutes prior to the shock. Used to classify high- versus low-density periods.
`rain`	Indicator variable equal to 1 if daily precipitation at the nearest meteorological station exceeds 10mm.

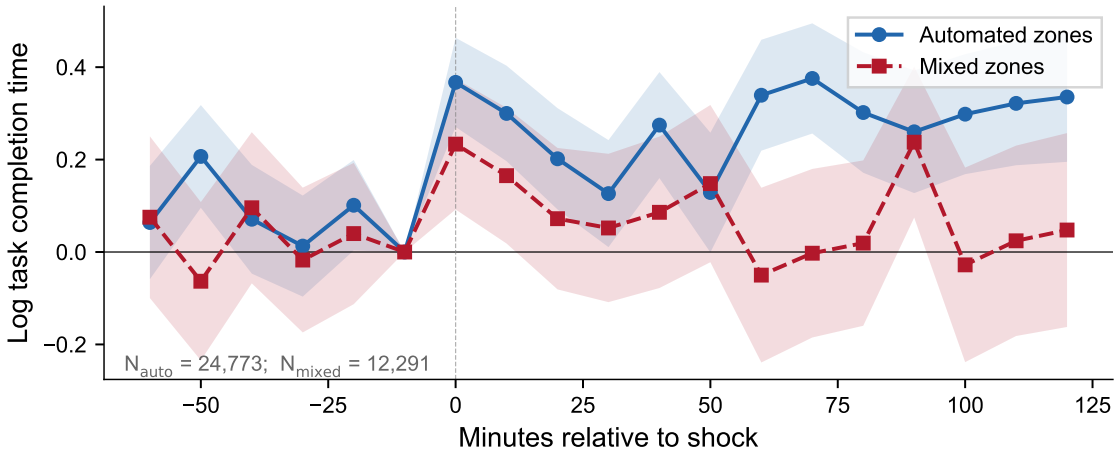


Figure 7: Robustness: Log Task Completion Time

Notes: Preferred specification with ln(task completion time) as the dependent variable. Shocks: mechanical faults of 60–120 seconds. 10-minute bins; reference bin $[-10, 0)$. Shaded areas: 95% confidence intervals (HC1).

Alternative shock definitions. The main analysis restricts shocks to mechanical faults of 60–120 seconds. Figure 8 broadens the definition to include all faults of 60 seconds or longer, with no upper bound on severity. The persistence pattern in automated zones and recovery in mixed zones are unchanged, indicating that the result does not depend on the specific severity window.

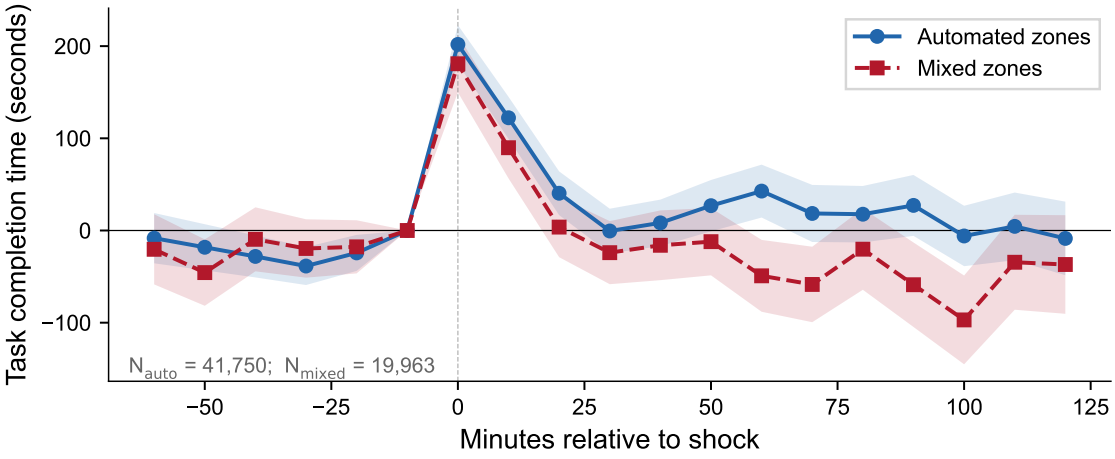


Figure 8: Robustness: All Shocks $\geq$ 60 Seconds

Notes: Preferred specification with no upper bound on shock severity. All mechanical faults $\geq$ 60 seconds included.

Formal interaction test. Figure 9 reports the interaction coefficients ϕ_k from the pooled specification (Equation 3), which directly estimate the difference in recovery paths between automated and mixed zones at each time bin. The pre-shock interaction coefficients are near zero (joint F-test: $p > 0.10$), confirming parallel pre-trends, while the late-bin interactions are positive and significant (joint F-test: $p < 0.001$), confirming differential persistence.

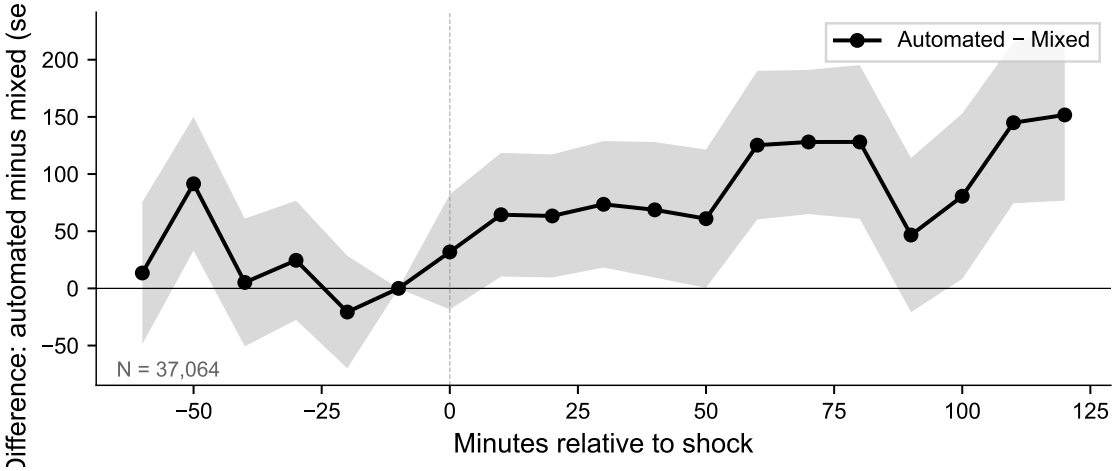


Figure 9: Interaction Coefficients: Automated Minus Mixed, by Time Bin

Notes: Difference in event study coefficients between automated and mixed zones at each time bin. Late-bin joint F-test: $p < 0.001$. Pre-trend F-test: $p > 0.10$.

Leave-one-out robustness. A potential concern is that the automated-zone result is driven by a single crane with unusual operational characteristics. Figure 10 re-estimates the main specification after dropping each crane in turn. The automated late-bin average ranges from +28 to +49 seconds across permutations, and no single crane's removal eliminates the result.

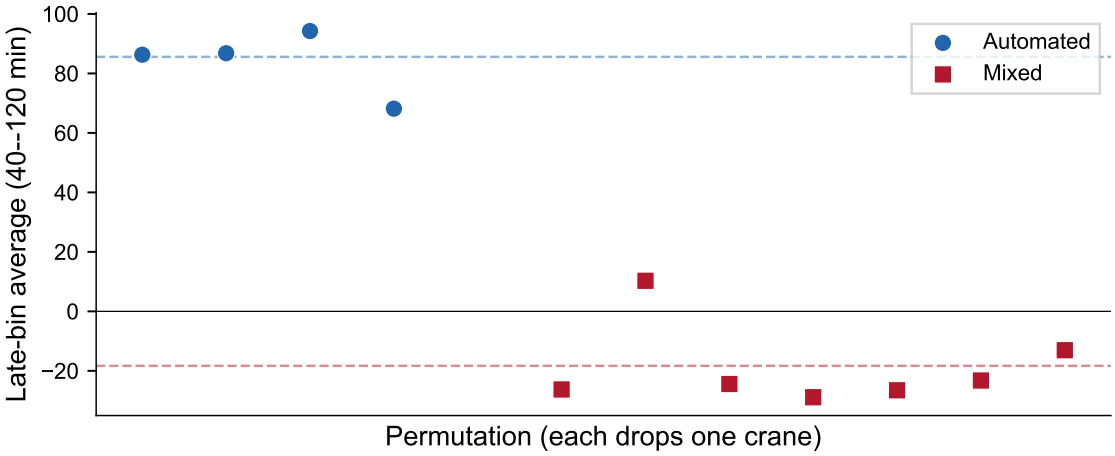


Figure 10: Leave-One-Out Robustness

Notes: Late-bin average (40–120 minutes post-shock) re-estimated after dropping each crane. Each point is one permutation. Dashed lines: full-sample estimates.

Physical environment heterogeneity. The model predicts that propagating delays require sequential queuing in constrained corridors. Figure 11 tests this by comparing quay-lane shocks (narrow, single-file corridors) with yard-block shocks (open areas with multiple access routes). Persistence is concentrated in quay lanes (+33 to +55 seconds in the late bins, $N = 71{,}499$); yard blocks show negative late-bin coefficients (−92 to −136 seconds, $N = 14{,}450$), consistent with vehicles rerouting to less congested areas.

Day versus night. A concern with the main result is that human truck presence varies systematically by time of day, and that the recovery in mixed zones reflects daytime-specific conditions rather than the structural role of independent agents. Figure 12 splits the sample into day (6AM–6PM) and night (6PM–6AM) subsamples. The persistence pattern holds in both, ruling out time-of-day variation as a confound.

Shock severity balance. A threat to the comparison between automated and mixed zones would be if shocks in automated zones were systematically more severe. Figure 13 shows that the distributions of shock duration are statistically indistinguishable across zone types (Mann-Whitney $p = 0.19$).

Weather moderation coefficients. Table 6 reports the full event study coefficients underlying the weather moderation analysis in the main text. The specification estimates the preferred model separately for

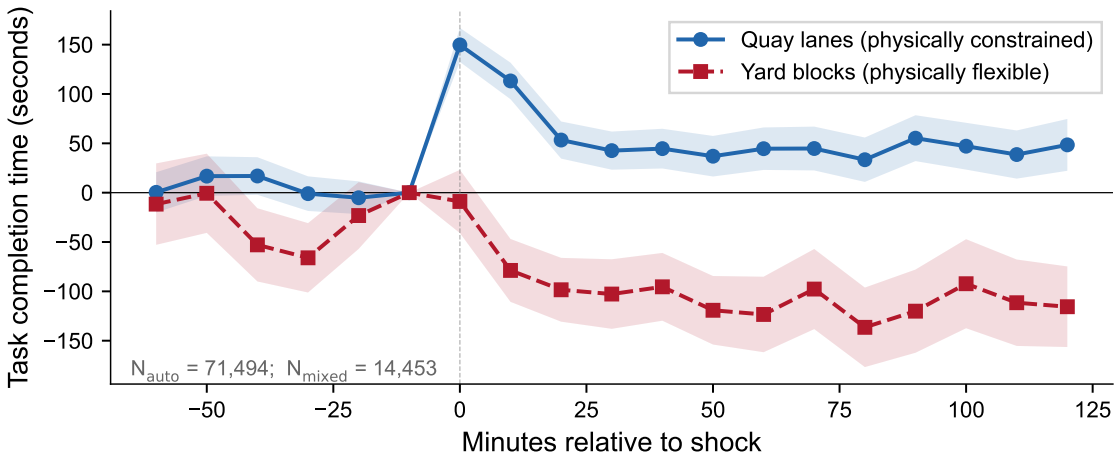


Figure 11: Shock Propagation: Quay Lanes (Constrained) versus Yard Blocks (Flexible)

Notes: Preferred specification estimated separately for quay-lane shocks (N = 71,499) and yard-block shocks (N = 14,450). All vehicle types pooled. Quay lanes are narrow corridors with single-file queuing; yard blocks have multiple rows and access points permitting rerouting.

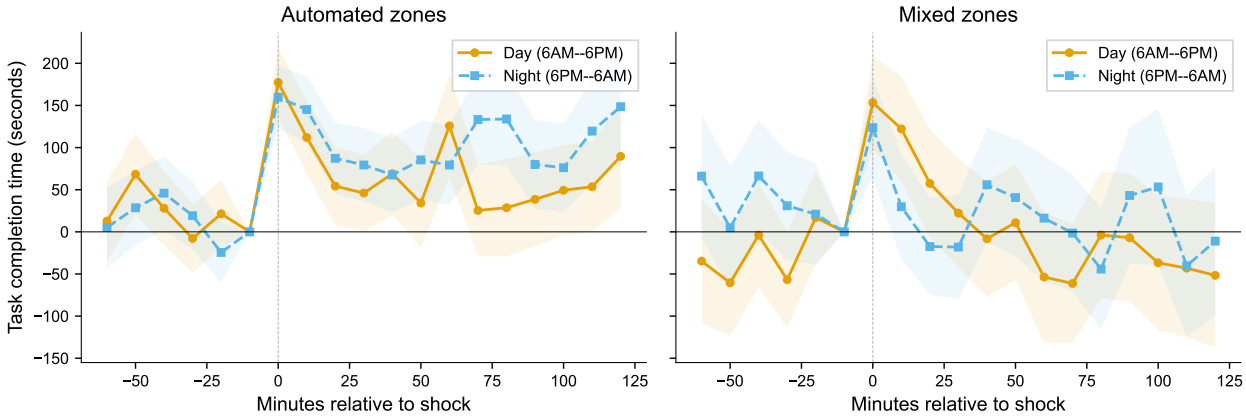


Figure 12: Robustness: Day versus Night Shifts

Notes: Preferred specification estimated separately for shocks during day (6AM–6PM) and night (6PM–6AM). The persistence pattern holds in both subsamples, ruling out time-of-day variation in human truck presence as a confound.

four cells defined by zone type (automated vs. mixed) and daily precipitation (dry vs. rain >10mm), using time at zero speed as the dependent variable. The key finding is that automated zones, which show persistent post-shock delays under dry conditions, recover fully under heavy rain—consistent with the model's prediction that exogenous buffer eliminates propagation.

Sensitivity to weather and winsorization thresholds. The main weather analysis uses a precipitation threshold of 10mm to classify rainy days. The qualitative pattern—automated zones recovering under rain, mixed zones recovering under dry conditions—is robust across thresholds from 0.1mm to 25mm. The 10mm threshold provides the best combination of statistical power and effect size. The time-stationary and coupling

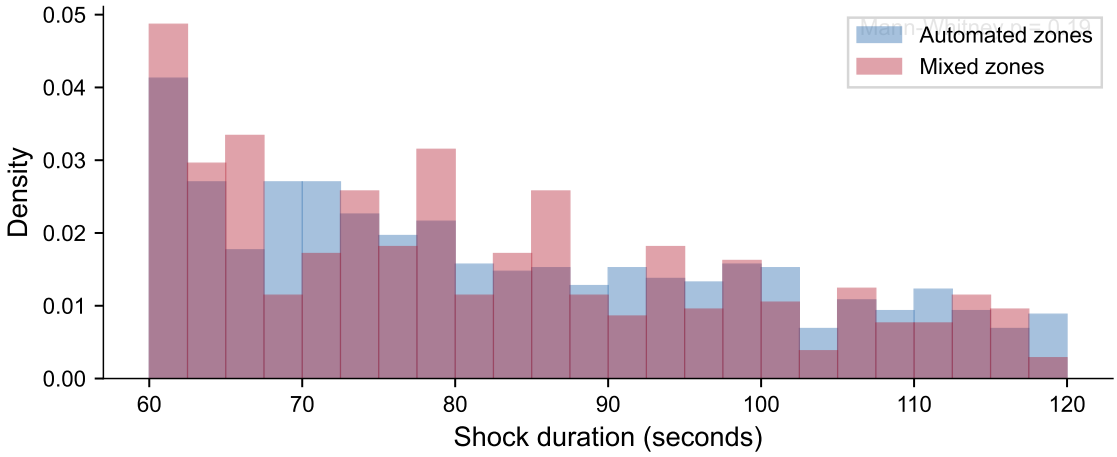


Figure 13: Shock Severity Distributions: Automated versus Mixed Zones

Notes: Histograms of mechanical fault duration (seconds), faults 60–120s. Mann-Whitney $p = 0.19$, confirming that shock severity distributions are statistically indistinguishable across zone types.

Table 6: Weather Moderation: Time Stationary

	Automated		Mixed	
	Dry	Rain >10mm	Dry	Rain >10mm
	(1)	(2)	(3)	(4)
$[0, +20)$ min	139.8***	97.7***	96.5***	202.6***
	(6.3)	(20.1)	(9.0)	(29.5)
$[+20, +40)$	33.0***	−14.6	6.8	57.3*
	(6.7)	(21.7)	(9.3)	(31.4)
$[+40, +60)$	18.3**	−47.7*	−18.5*	17.9
	(7.2)	(25.3)	(10.2)	(39.7)
$[+60, +80)$	40.0***	−55.3**	−24.0**	−111.8***
	(8.6)	(26.2)	(10.9)	(36.5)
$[+80, +100)$	19.6**	−27.3	−22.5*	44.4
	(8.9)	(29.6)	(12.6)	(48.7)
$[+100, +120)$	9.2	−73.1**	−69.9***	70.5
	(9.7)	(29.6)	(13.5)	(51.8)
Zone FE	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes
Observations	37,915	4,474	18,200	2,041
R^2	0.087	0.136	0.115	0.181

Notes: Dependent variable: time at zero speed (seconds), winsorized at the 99th percentile (computed within each subsample). Rain: daily precipitation > 10mm. All shocks ≥ 60s, 20-minute bins, reference bin $[-20, 0)$. HC1 robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

results use winsorization at the 99th percentile. Re-estimating with the 95th and 97.5th percentile thresholds produces qualitatively identical results with attenuated coefficient magnitudes.

D Geomagnetic Storms and Mechanical Fault Exogeneity

A potential concern with using mechanical faults as exogenous shocks is that faults might be caused by traffic conditions rather than by equipment malfunction. This section presents evidence from geomagnetic storms that directly addresses this concern.

AGVs navigate using GPS receivers, lidar sensors, and continuous radio communication with the centralized dispatch system. Geomagnetic storms—caused by solar coronal mass ejections interacting with Earth's magnetosphere—are known to degrade GPS accuracy and disrupt radio communications. Human-operated trucks, which navigate by sight, should be largely unaffected.

I classify each day according to the NOAA Space Weather Scale for geomagnetic storms (G1–G4). The sample period coincides with the peak of Solar Cycle 25, providing 14 G4-severity days, 22 G3+ days, 34 G2+ days, and 48 G1+ days.

Table 7: AGV Mechanical Fault Rates by Geomagnetic Storm Intensity

	Automated Zones		Mixed Zones		
Storm Intensity	Storm	Normal	Storm	Normal	DiD
G4 Severe (Kp $\geq$ 8)	45.5	34.5	28.4	35.0	+17.6
G3 Strong (Kp $\geq$ 7)	40.4	34.5	26.8	35.3	+14.3
G2 Moderate (Kp $\geq$ 6)	34.8	34.9	29.4	35.4	+5.9
G1 Minor (Kp $\geq$ 5)	32.4	35.3	31.1	35.3	+1.3

Notes: Fault rates are mechanical faults ($\geq$ 60 seconds) per 1,000 AGV tasks. Storm classification follows the NOAA Space Weather Scale based on the planetary Kp index. DiD = (Automated storm − Automated normal) − (Mixed storm − Mixed normal). Storm day counts: G4 = 14 days, G3+ = 22 days, G2+ = 34 days, G1+ = 48 days.

Two patterns are evident. First, fault rates in automated zones increase monotonically with storm severity: from 34.5 per 1,000 tasks on non-storm days to 45.5 on G4-severity days, a 32% increase. Second, fault rates in mixed zones do not increase on storm days. The monotonic dose-response supports a direct electromagnetic mechanism rather than an indirect traffic pathway.

E Stata–Python Verification

All regression estimates reported in this paper were independently verified using two statistical platforms: Stata (using the `reghdfe` package for high-dimensional fixed effects estimation; Correia, 2017) and Python (using iterative Frisch-Waugh-Lovell demeaning with HC1 robust standard errors). The maximum coefficient

difference across all specifications is less than 0.1 standard errors, and sample sizes differ by at most 4 observations (attributable to singleton drops in `reghdfe` that the Python implementation does not perform). Detailed coefficient comparison tables are available from the author upon request.