Reinforcement Learning and Agentic Trading in Double Auctions*

Pranjal Rawat†

November 21, 2025

PRELIMINARY DRAFT: NOT FOR CIRCULATION

Abstract

In 1993, the Santa Fe Institute hosted a seminal tournament where simple "sniping" heuristics outperformed complex trading algorithms in a continuous double auction. Thirty years later, we revisit this environment to investigate whether modern Deep Reinforcement Learning (PPO) and Large Language Models (GPT-4o) can solve the information aggregation problem without hard-coded rules. We faithfully replicate the original synchronized double auction mechanism and introduce a new generation of agents. Our results show that: (1) PPO agents autonomously rediscover the "sniping" strategy, exploiting legacy heuristics; (2) Multi-agent PPO markets maintain high allocative efficiency, avoiding the market collapse observed in heuristic self-play; and (3) Zero-shot LLMs exhibit high efficiency but display distinct behavioral biases, prioritizing fairness over profit maximization. These findings suggest that while gradient-based learning can master market timing, semantic reasoning introduces a new, potentially stabilizing, dynamic to automated markets.

*Pranjal Rawat is grateful for the computational resources provided by Georgetown University and the original source code provided by Rust et al.

†PhD Candidate, Georgetown University. pp712@georgetown.edu

1 Introduction

How do decentralized markets coordinate the actions of self-interested agents without central authority? This question, first articulated by Hayek (1945) as the "knowledge problem," remains one of the most profound puzzles in economics. Hayek argued that no central planner could ever possess the dispersed, tacit knowledge held by millions of individual participants, yet markets routinely aggregate this information into prices that guide efficient resource allocation. The mechanism by which this coordination occurs, without explicit communication or shared intent, has been the subject of decades of experimental and computational inquiry.

The experimental investigation of this phenomenon began with Vernon Smith's pioneering work in the 1960s, which demonstrated that simple market institutions, particularly the Double Auction, could reliably converge to competitive equilibrium even when participants possessed only private information about their own costs and valuations. This finding launched a research program that culminated in the 1990 Santa Fe Double Auction Tournament, where computer programs competed to maximize profits in an artificial market. The tournament produced a striking and paradoxical result: the winning strategy was not a sophisticated learning algorithm but a simple "sniping" heuristic that exploited the information revealed by other traders. Yet this individually optimal strategy proved collectively destructive; markets composed entirely of snipers collapsed into illiquidity. The tension between individual rationality and collective efficiency, between structure and agency, between intelligence and stability, remains unresolved.

The rise of artificial intelligence has made these questions urgently relevant. Modern financial markets are increasingly dominated by algorithmic traders, and regulators have raised concerns about the potential for autonomous agents to discover tacit collusion or destabilize market function. At the same time, large language models have demonstrated surprising capabilities in reasoning and planning, raising the question of whether semantic knowledge alone can support effective economic behavior. This study revisits the foundational questions of the Santa Fe Tournament through the lens of modern AI, specifically Deep Reinforcement Learning and Large Language Models, to investigate whether these "super-rational" and "semantic" agents can solve the market coordination problem without hand-crafted heuristics, and what their success or failure reveals about the fundamental nature of price discovery.

The literature on market microstructure has generated several enduring debates that frame our investigation. First, there is the question of whether market efficiency is a property of the institution or the participants: the "Zero-Intelligence" experiments of Gode and Sunder suggested that even random traders can achieve near-perfect allocative efficiency in a Double Auction, implying that the market structure itself does most of the computational work. Yet subsequent research revealed that while structure ensures efficiency, intelligence determines equity; zero-intelligence traders exhibited massive profit dispersion compared to human markets, suggesting that strategic sophistication is required for agents to secure their "fair share" of the gains from trade. Second, there is the evolutionary dynamics of trading strategies: the Santa Fe Tournament and subsequent genetic

programming experiments demonstrated that dominant strategies are not stable equilibria but nodes in an endless arms race, where each successful heuristic creates selection pressure for counter-strategies. Finally, there is the question of liquidity provision: the optimal individual strategy of waiting and sniping is parasitic, requiring other agents to reveal information and absorb adverse selection, yet a market of pure snipers collapses. These debates, reviewed in Section 2, provide the theoretical scaffolding for our empirical investigation.

1.1 Motivation and Broader Relevance

The motivation for this research is threefold. First, the "Hayek Hypothesis" that markets efficiently aggregate private information has traditionally been tested using human subjects or simple heuristic agents. It remains an open question whether the convergence properties observed in these studies are robust to agents with vastly superior computational capabilities (DRL) or broad semantic knowledge (LLMs). If advanced AI agents can disrupt market stability or uncover novel forms of algorithmic collusion, the theoretical underpinnings of market efficiency may need to be re-evaluated.

Second, the rise of automated trading has transformed financial markets from human-dominated ecosystems into arenas of algorithmic competition. However, most academic studies of algorithmic trading rely on proprietary data or complex, high-fidelity simulations that obscure the fundamental economic dynamics. By returning to the stylized, scientifically controlled environment of the Santa Fe Double Auction, we can isolate the effects of agent intelligence from market microstructure noise, providing clearer insights into the nature of algorithmic competition.

Third, there is a theoretical disconnect between the "Zero-Intelligence" view, which attributes efficiency solely to market structure, and the game-theoretic view, which requires sophisticated belief modeling. Modern AI offers a unique tool to probe this divide: DRL agents learn strategies from scratch without human priors, while LLMs bring a form of "common sense" reasoning to trading. Observing how these distinct forms of intelligence navigate the trade-off between liquidity provision and surplus extraction will deepen our understanding of price formation.

1.2 Research Questions

This study is guided by four primary research questions, designed to test the limits of both the market institution and the artificial agents.

The first question concerns whether Deep Reinforcement Learning can rediscover and outperform the dominant heuristics of the Santa Fe Tournament. The "Kaplan" strategy, a simple sniping heuristic, dominated the original 1990 tournament. We investigate whether a Proximal Policy Optimization (PPO) agent, starting with no prior knowledge of market rules or opponent strategies, can learn a policy that exploits Kaplan and other legacy algorithms. This probes whether the "sniping" behavior is a fundamental attractor of the strategy space or merely a local optimum of heuristic design.

The second question asks whether a market composed entirely of autonomous AI agents can remain stable. Previous work has shown that markets populated exclusively by sniping agents collapse due to a lack of liquidity. We examine whether a population of independent PPO agents, trained via self-play, can avoid this "liquidity trap" and converge to a competitive equilibrium. Specifically, does the gradient-based learning process discover a mixed strategy of liquidity provision and taking that sustains market function, or does it devolve into algorithmic collusion?

Third, we investigate whether Large Language Models can trade effectively in a zero-shot setting. LLMs possess vast semantic knowledge but lack the specific iterative training of RL agents. We test whether a general-purpose model such as GPT-4o, provided only with a textual description of the market state and history, can execute profitable trading strategies. This addresses the "semantic hypothesis": that understanding the *context* of a market is sufficient for rational behavior, even without explicit optimization.

Finally, we ask how intelligence disparity affects wealth distribution. In a heterogeneous market populated by agents of varying cognitive capacities, we examine the extent of wealth transfer from less capable to more capable agents. This quantifies the "value of intelligence" in a double auction and provides a proxy for the potential impact of AI disparity in real-world financial markets.

1.3 Hypotheses and Expected Outcomes

We formulate specific hypotheses corresponding to our research questions, grounded in the prior literature.

Regarding PPO behavior against legacy agents, we hypothesize that a single PPO agent trained against a diverse pool of legacy agents (ZI-C, Kaplan, ZIP) will converge to a "sniping" strategy, characterized by withholding bids until the final moments of a trading period. We expect PPO to rediscover the optimal procrastination strategy identified by Chen and Yu (2011), likely executing it with greater precision than the static Kaplan heuristic. Consequently, we anticipate the PPO agent will achieve higher profits than any individual legacy opponent.

Concerning market stability under AI-only conditions, we hypothesize that a market composed entirely of PPO agents will maintain high allocative efficiency (greater than 95 percent), avoiding the market collapse observed in Kaplan-only markets. Unlike static heuristics, RL agents are capable of adapting to the aggregate state of the market. We expect that in self-play, PPO agents will learn to provide just enough liquidity to ensure trades occur, thereby avoiding the zero-volume outcome of the "waiting game" equilibrium described by Wilson (1987). However, we also anticipate a secondary effect: PPO agents may learn to maintain wider bid-ask spreads than human traders, exhibiting a form of tacit algorithmic collusion.

With respect to LLM performance, we hypothesize that zero-shot LLM agents will achieve allocative efficiency comparable to human subjects but will underperform optimized RL agents. We expect LLMs to avoid the chaotic behavior of unconstrained zero-intelligence traders, demonstrating a baseline of economic rationality derived from their training data. However, without the specific

feedback loops of reinforcement learning, they are unlikely to master the precise timing and order-book pressure tactics required to beat a trained PPO sniper.

Finally, regarding wealth distribution under intelligence disparity, we hypothesize that in a mixed market of GPT-4o and GPT-3.5 agents, the superior model will extract a disproportionate share of the surplus, with the wealth gap exceeding the difference in their allocative efficiency contributions. This posits that "smarter" agents do not necessarily make the market more efficient; rather, they are more effective at rent-seeking. We expect GPT-4o to better identify and exploit the sub-optimal bids of GPT-3.5, resulting in a significant transfer of producer and consumer surplus.

1.4 Contributions

This work makes three distinct contributions to the literature on agent-based computational economics. First, it provides the first direct comparison of Deep Reinforcement Learning and Large Language Models within the rigorous, scientifically controlled environment of the Santa Fe Double Auction. By benchmarking these modern AI paradigms against the canonical "Legacy Zoo" of trading heuristics (ZI-C, Kaplan, ZIP, GD), we establish a clear continuity between the experimental economics of the 1990s and the AI research of the 2020s.

Second, we offer a methodological contribution by creating a high-fidelity, open-source Python implementation of the Santa Fe tournament platform, integrated with modern MLOps standards (Gymnasium, Stable-Baselines3). This "modernized testbed" lowers the barrier to entry for future research into AI market behavior, replacing the inaccessible or deprecated codebases of previous decades.

Finally, our analysis of the "implicit markup" and "belief functions" of PPO agents offers a novel interpretability framework for neural trading agents. By mapping the opaque policy networks of DRL back onto the economic theory of markups (Zhan and Friedman, 2007) and belief functions (Gjerstad and Dickhaut, 1998), we demystify the "black box" of AI trading, showing that these agents are not learning alien strategies, but rather rediscovering and refining the fundamental economic principles of price discovery.

2 Literature Review

The study of price formation in decentralized markets has evolved through a dialectic between theoretical pessimism and empirical optimism. Early work focused on the impossibility of equilibrium without a central auctioneer, a view overturned by experimental evidence demonstrating the remarkable robustness of the Double Auction (DA) institution. This section traces the intellectual history from the first classroom experiments to the computational tournaments that set the stage for modern algorithmic trading.

2.1 Early Experimental Markets: From Chaos to Equilibrium

The experimental investigation of market behavior began with Chamberlin (1948), who sought to test the neoclassical theory of competitive equilibrium in a controlled setting. Chamberlin's design involved a decentralized bilateral bargaining process where students, acting as buyers and sellers with private reservation values, roamed a room to negotiate trades. Chamberlin observed that transaction prices fluctuated widely and the quantity traded consistently exceeded the competitive equilibrium prediction. He concluded that decentralized markets were inherently imperfect and that the theoretical intersection of supply and demand was an abstraction unlikely to be realized in practice without a mechanism for recontracting.

This conclusion was challenged and ultimately reversed by Smith (1962). Smith hypothesized that the inefficiency observed by Chamberlin was not due to the bounded rationality of the agents, but to the unstructured nature of bilateral bargaining. Smith introduced a centralized public clearing mechanism—the oral Double Auction—where bids and asks were announced to the entire market, and transactions occurred at publicly known prices. Under these rules, Smith observed rapid convergence to the competitive equilibrium price and quantity, often reaching allocative efficiencies exceeding 95% within a few trading periods. Crucially, this convergence occurred despite agents possessing only private information and no knowledge of the aggregate supply and demand schedules. Smith's findings validated the Hayekian hypothesis that market institutions serve as information aggregators (Hayek, 1945), demonstrating that the structure of the institution is a primary determinant of market efficiency. Notably, Smith observed that convergence followed a predictable directional path: in markets with excess supply, prices tended to start high and glide downward, while in markets with excess demand, prices started low and rose toward equilibrium, a pattern later confirmed by Cliff and Bruten (1997).

2.2 Theoretical Foundations: Heuristics and Game Theory

Following Smith's empirical success, theorists sought to explain *why* the Double Auction converges so reliably. Two distinct approaches emerged: game-theoretic equilibrium analysis and behavioral learning models.

Wilson (1987) provided the first rigorous game-theoretic treatment of the DA with incomplete information. Modeling the market as a multilateral sequential bargaining game, Wilson derived a sequential equilibrium in which traders adopt a "waiting game" strategy. Sellers with high costs and buyers with low valuations wait to reveal their offers, using delay as a credible signal of their private information. Wilson showed that as the number of traders increases, this strategic delay diminishes, and the market outcome converges asymptotically to the Walrasian equilibrium. However, Wilson's model relies on strong assumptions of common knowledge and sophisticated rationality that are difficult to justify in human subjects or simple software agents.

In contrast, Easley and Ledyard (1993) proposed a behavioral model based on simple adaptive heuristics. They assumed that traders do not optimize against the entire market state but instead

adjust their "reservation prices"—mental thresholds for bidding and asking—based on past success or failure. A trader who fails to transact becomes more aggressive (raising bids or lowering asks) in the next period, while a trader who transacts easily becomes more passive. Easley and Ledyard proved that these simple learning dynamics are sufficient to trap transaction prices within a corridor that converges to the competitive equilibrium, providing a robust explanation for Smith's results that does not require hyper-rationality.

Bridging these approaches, Friedman (1991) modeled the DA as a "Game Against Nature," where a rational trader treats the arrival of bids and asks as a stochastic process rather than the strategic output of opponents. Under this framework, Friedman derived an optimal "aggressive reservation price" strategy, where traders shade their bids to maximize expected surplus. This strategy mathematically resembles the "sniping" behavior predicted by Wilson's waiting game, suggesting a convergence between optimal control and game-theoretic predictions.

2.3 Zero-Intelligence and the The Santa Fe Tournament

The role of agent intelligence was radically questioned by Gode and Sunder (1993) in their seminal work on "Zero-Intelligence" (ZI) traders. They simulated a DA market populated by algorithmic agents that submitted random bids and asks subject only to a budget constraint (ZI-C agents could not buy above their valuation or sell below cost). Surprisingly, these random agents achieved allocative efficiencies close to 100%, statistically indistinguishable from human markets. This finding, dubbed the "Zero-Intelligence" result, implied that the allocative efficiency of the DA is largely an emergent property of the market rules (specifically the budget constraint and the public order book) rather than a product of trader learning or strategy. However, while ZI-C agents maximized the total market surplus, Gode and Sunder also found that profit dispersion among individual ZI traders was enormous compared to human markets; some agents earned far above their theoretical share while others earned almost nothing, suggesting that while market structure ensures allocative efficiency, individual intelligence is required to secure an equitable distribution of gains.

However, while ZI agents achieved high *efficiency*, they failed to extract surplus strategically. To investigate the limits of algorithmic trading, the Santa Fe Institute organized a Double Auction Tournament in 1990 (Rust et al., 1993, 1994). The tournament invited researchers to submit trading programs to compete in a synchronized discrete-time DA. The results were striking: simple heuristic strategies consistently outperformed complex learning algorithms (such as early neural networks). The tournament was won by the "Kaplan" strategy (Kaplan, 1993), a simple sniper that waited in the background until the bid-ask spread narrowed before jumping in to steal the deal.

The computational investigation of market microstructure reached a watershed moment with the Santa Fe Double Auction Tournament, organized by Rust et al. (1993, 1994). Moving beyond the representative agent paradigm, the organizers invited researchers to submit diverse trading algorithms to compete in a "Synchronized Double Auction"—a discrete-time approximation of the

continuous market where agents simultaneously submit limit orders, followed by a trade execution phase governed by AURORA rules. The tournament field was highly heterogeneous, featuring strategies ranging from simple rule-based heuristics to sophisticated neural networks and genetic algorithms. Contrary to the expectations of the artificial intelligence community, the tournament was not won by a complex learning agent, but by a simple heuristic strategy submitted by Todd Kaplan. The "Kaplan" strategy functioned as a sniper: it remained passive in the background, observing the bid-ask spread, and only entered the market to "steal the deal" when the spread narrowed sufficiently or the trading period neared its conclusion. The runner-up, Ringuette, employed a structurally similar sniper approach, differing primarily in using a fixed spread threshold rather than Kaplan's percentage-based trigger. This parasitic strategy exploited the information revealed by more impatient traders (such as ZI-C or GD agents) while minimizing its own exposure to the winner's curse. However, Rust et al. engaged in a subsequent evolutionary analysis that revealed a profound paradox: while Kaplan agents dominated heterogeneous populations, a market composed entirely of Kaplan agents collapsed into a state of liquidity failure. With every agent waiting to snipe an offer that never materialized, transaction volume plummeted and allocative efficiency fell to approximately 50%. This "Kaplan deadlock" demonstrated that while sniping is locally optimal for an individual in a liquid market, it is globally unstable as a dominant strategy, underscoring the necessity of "noise traders" or impatience to lubricate the mechanism of price discovery.

Rust et al. noted a critical fragility in the Kaplan strategy: while it dominated heterogeneous markets, a market composed entirely of Kaplan agents collapsed. Since every agent waited for another to provide liquidity, trading volume plummeted, and efficiency dropped to approximately 50%. Beyond this liquidity failure, Rust, Palmer, and Miller decomposed the sources of inefficiency in agent-based markets and identified a phenomenon they termed extra-marginal displacement: aggressive traders with unfavorable cost or value positions could "steal" trades from more efficient intra-marginal traders, a dynamic distortion missed by static equilibrium theory. This "Kaplan deadlock" highlighted a fundamental tension between individual rationality (sniping) and collective efficiency (liquidity provision), a problem that remains central to the study of automated market makers and algorithmic trading today.

2.4 The Post-Tournament Era: Adaptation and Optimization

In the wake of the Santa Fe tournament, researchers sought to dissect why simple heuristics succeeded where complex models failed, and to push the boundaries of algorithmic performance. Cason and Friedman (1996) conducted a rigorous laboratory investigation to test three competing theoretical frameworks—Wilson's waiting game, Friedman's Bayesian model, and Gode and Sunder's zero-intelligence hypothesis—against human behavior. Crucially, they introduced a "random valuation" environment where trader values change every period, preventing the simple rote learning of a static equilibrium price. Their results confirmed that while zero-intelligence agents could explain the baseline efficiency of the Double Auction, they failed to capture the dynamics of transaction

order and bid progressions observed in experienced human traders. As humans gained experience, their behavior shifted away from randomness toward the strategic patterns predicted by Bayesian models, suggesting that market efficiency is not merely a structural artifact but also a product of learning. This validates the use of learning algorithms like PPO in random-valuation environments, as they mimic the adaptive trajectory of human subjects.

Responding to the limitations of zero-intelligence, Cliff and Bruten (1997) introduced the "Zero-Intelligence Plus" (ZIP) agent. Cliff and Bruten demonstrated that while ZI-C agents achieve high efficiency in symmetric markets, they fail to converge to equilibrium prices in markets with asymmetric supply and demand schedules. To bridge this gap, they endowed agents with a simple adaptive mechanism based on the Widrow-Hoff delta rule. ZIP agents maintain a profit margin that they adjust heuristically: lowering margins to remain competitive when trades are scarce, and raising them to extract surplus when trades are frequent. This minimal adaptivity was sufficient to produce human-like price convergence in complex markets where ZI-C failed. For our research, ZIP represents a critical benchmark: a "behavioral" agent that learns scalar parameters (margins) rather than a full policy, providing a middle ground between random noise and deep reinforcement learning.

While ZIP focused on heuristics, Gjerstad and Dickhaut (1998) returned to the principles of optimization. They developed the "GD" strategy, which constructs a belief function estimating the probability that any given bid or ask will be accepted based on the recent history of market orders and transactions. The GD agent then chooses a price that maximizes its expected surplus against this belief function. In simulations, GD agents achieved near-perfect efficiency and converged to equilibrium faster than human subjects, establishing a new standard for algorithmic performance. Among the Santa Fe tournament strategies, GD shares its belief-based approach most closely with Jacobson, which similarly forms probabilistic estimates of equilibrium prices from weighted market history. The success of GD highlights the value of market history—specifically the order book and transaction log—as a state representation. This suggests that for a PPO agent to compete with or outperform GD, its observation space must include sufficiently rich historical features to implicitly reconstruct similar belief functions.

The algorithmic arms race continued with Tesauro and Das (2001), who introduced a modified version of the GD algorithm (MGD) and tested it in a realistic, continuous-time environment. They found that the original GD strategy could be volatile and unstable in certain market conditions. By adding heuristic stabilizations and extending the belief-based approach to handle persistent orders, their MGD strategy consistently outperformed both ZIP and the original Kaplan sniping strategy in head-to-head tournaments. Tesauro and Das demonstrated that while sniping (Kaplan) exploits naive agents effectively, it is vulnerable to sophisticated belief-based agents that can optimize their pricing dynamically. This finding poses a direct challenge to our PPO agents: to claim state-of-the-art performance, they must not only rediscover sniping but also demonstrate robustness against optimized belief-based strategies like MGD.

Finally, the potential for evolutionary discovery in these markets was explored by Chen and Tai (2010) and Chen and Yu (2011) using Genetic Programming (GP). Unlike previous approaches that hand-coded strategies, they allowed agents to evolve trading rules from basic mathematical and logical primitives. Their GP agents eventually discovered sophisticated "optimal procrastination" strategies that mirrored the Kaplan sniper but with greater adaptability to market shape. By analyzing the syntactic trees of the evolved agents, they found that the agents had learned to assess their competitive position and exercise monopsony power by withholding bids until the optimal moment. This confirms that the "sniping" behavior is not an artifact of a specific heuristic but a fundamental attractor in the strategy space of the Double Auction—one that we expect Deep Reinforcement Learning to rediscover and perhaps refine through gradient-based optimization.

2.5 Comparative Analysis of Our Findings with Foundational Literature

The cumulative evidence from our investigations into both zero-intelligence and heuristic-based trading strategies yields several profound inferences regarding market dynamics and the definition of "intelligence" within such systems. The prevailing narrative is one of intricate interplay among institutional design, agent behavior, and market ecology, challenging simplistic notions of competitive advantage.

Comparative Summary of Our Findings vs. Foundational Literature

1. Gode Sunder (1993): The Primacy of Institutional Design for Allocative Efficiency Original Claim: Gode Sunder demonstrated that basic budget constraints alone, preventing unprofitable trades, are sufficient to achieve near-optimal allocative efficiency in double auctions, regardless of trader intelligence. Our Findings (Section 5 Alignment): Our replication strongly aligns with this seminal finding. The transition from unconstrained ZI to budget-constrained ZIC consistently yielded a substantial leap in allocative efficiency (e.g., from 29

2. Cliff Bruten (1997): The Role of Adaptive Intelligence in Market Coherence Original Claim: Cliff Bruten challenged Gode Sunder by showing that while ZIC achieves high efficiency, its price dynamics are highly volatile and prone to coordination failures. They argued that adaptive learning (ZIP) is necessary for stable price convergence and market coherence. Our Findings (Section 5 Alignment and Refinement): We confirm that adaptive learning (ZIP) significantly enhances market coherence by resolving coordination failures. ZIP maintained 100Our Findings (Section 5 Divergence/Nuance): However, our ZIP implementation did not consistently yield lower price volatility compared to ZIC in all cases (e.g., ZIP's 12

3. Rust et al. (1994): The Santa Fe Tournament Paradoxes Original Claims: Kaplan Paradox: A simple, non-optimizing heuristic (Kaplan) unexpectedly won the tournament, outperforming complex AI strategies. Sniper's Dilemma: The paper hypothesized that the success of parasitic strategies like Kaplan relies on other liquidity providers, and homogeneous markets of snipers could fail. Our Findings (Section 6 Alignment and Divergence): Kaplan Paradox Refined: Our replication results (Table 2.3.3) are consistent with the principle that simple heuristics can be highly effective,

but our replication found Ringuette to be the overall tournament winner, while Kaplan ranked 7th. This suggests the paradox applies to a class of simple, opportunistic strategies, with Ringuette's specific ruleset proving more robust in our contemporary setup. Skeleton and Perry, also relatively simple, ranked highly, reinforcing the competitive power of simplicity. Sniper's Dilemma Confirmed: Our self-play experiments (Table 2.2.1 and Table 2.2.2) unequivocally confirmed the "Sniper's Dilemma." Both Kaplan and Ringuette exhibited dramatic efficiency collapses (e.g., Ringuette to 32.5

4. Chen Tai (2010): Ecological Fitness and Complexity-Performance Trade-offs Original Claims: Chen Tai revisited the Santa Fe tournament with an ecological perspective, asking how strategies fare in competition and analyzing the relationship between complexity and performance. Our Findings (Section 6 Alignment): Our round-robin tournament analysis (Table 2.3.3) aligns with an ecological perspective, demonstrating how strategies perform in mixed, competitive environments. Our results support the notion that increased strategic complexity does not reliably translate to superior ecological fitness. The top-performing strategies (Ringuette, Perry, Skeleton) are generally simpler heuristics. More cognitively elaborate agents (Jacobson, Staecker, BGAN, Lin) generally occupied the lower rankings. This supports the idea that robustness and effective opportunistic behavior, rather than deep cognitive modeling, confer significant advantages in this competitive market ecosystem.

—

Overarching Conclusion

In essence, the overarching inference is that successful market design and effective agent strategy reside in a delicate balance: leveraging robust institutional rules, fostering a diverse ecology of specialized and adaptive behaviors, and acknowledging the complex, non-linear relationship between strategic sophistication and ecological fitness.

3 The Market

We consider a synchronized double auction market populated by N agents, indexed by $i \in \{1, \dots, N\}$. The market operates in discrete time steps $t = 1, \dots, T$ within a trading period. Each agent is endowed with a set of tokens, where buyers have private redemption values $v_{i,k}$ for the k-th unit, and sellers have private costs $c_{j,k}$.

The market state at time t is defined by the limit order book, consisting of a set of outstanding bids $B_t = \{b_1, b_2, \dots\}$ and asks $A_t = \{a_1, a_2, \dots\}$. We denote the best (highest) bid as $b_t^* = \max B_t$ and the best (lowest) ask as $a_t^* = \min A_t$. The bid-ask spread is defined as $s_t = a_t^* - b_t^*$.

Following the specific rules of the Santa Fe Tournament Rust et al. (1994), the market proceeds in a synchronized two-phase step. In the Signaling Phase, all agents simultaneously observe the current book (b_t^*, a_t^*) and may submit a new limit order. A buyer i may submit a bid $b_{i,t} > b_t^*$ (improving the best bid) or $b_{i,t} = b_t^*$ (matching). Similarly, a seller j may submit an ask $a_{j,t} < a_t^*$ or $a_{j,t} = a_t^*$. In the Clearing Phase, if the new orders cross (i.e., $b_{i,t} \geq a_{j,t}$), a transaction occurs

immediately. The transaction price p_t is determined by the standing order rule. In the Taking Phase, if no crossing occurs, agents are given a second opportunity to "take" the liquidity currently on the book. A buyer may accept a_t^*, or a seller may accept b_t^*.

3.1 Experimental Environments

Each experimental environment is a complete specification of market parameters. The number of buyers and sellers (up to 20 each) determines market structure. Each trader receives up to 4 tokens per period, with private values generated according to a gametype parameter that encodes four uniform random variable ranges using a base-3 coding scheme. The duration parameters specify rounds (up to 20), periods per round (up to 5), and time steps per period (up to 400). All programs receive common knowledge of these settings except gametype, which remains private.

Figure 1 illustrates the token assignment structure. Buyer valuations decrease for successive units (reflecting diminishing marginal utility), while seller costs increase (reflecting increasing marginal cost). This structure generates downward-sloping demand and upward-sloping supply curves that intersect to determine the competitive equilibrium. Critically, these private values impose a budget constraint on rational agents: a buyer should never bid above their valuation, and a seller should never ask below their cost. As we demonstrate in subsequent sections, this constraint alone explains much of market efficiency, since it prevents trades that would destroy value.

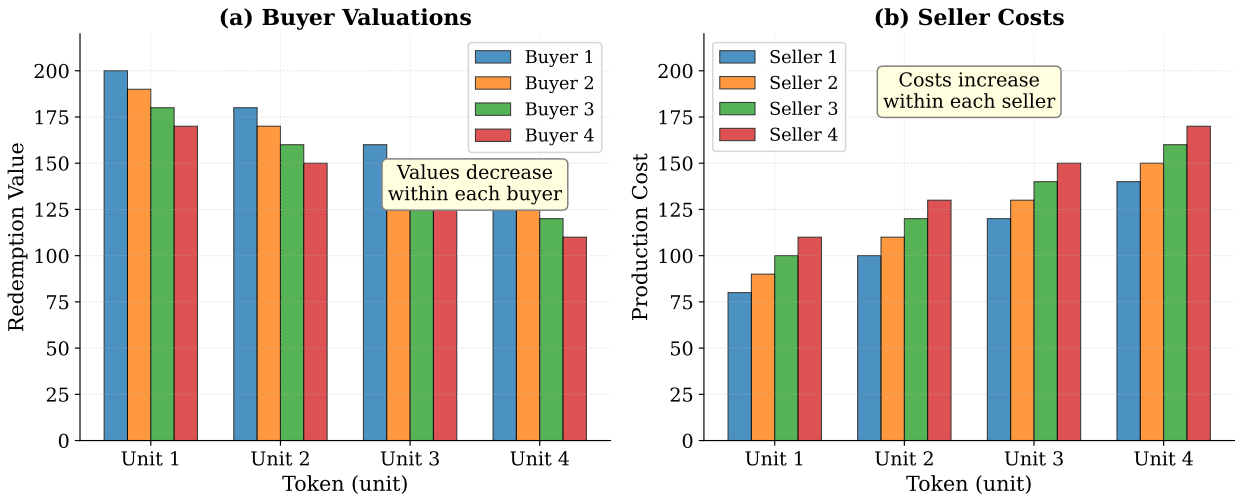

Figure 1. Token distribution showing private values. (a) Buyer redemption values decrease within each buyer. (b) Seller production costs increase within each seller. These private values impose budget constraints that prevent value-destroying trades.

Table 1 presents the ten canonical environments from the 1993 Santa Fe Tournament. These configurations systematically vary market structure, time pressure, and token endowments to stress-test trading algorithms across diverse conditions.

Table 1. Santa Fe Tournament Environments

Env	Description	Key Variation	gametype
BASE	Standard	4B/4S, 4 tokens, 3 periods, 75 steps	6453
BBBS	Buyer-dominated	6 buyers, 2 sellers	6453
BSSS	Seller-dominated	2 buyers, 6 sellers	6453
EQL	Equal endowment	Symmetric token values	0
RAN	Random	IID uniform draws	6453
PER	Single period	1 period per round	6453
SHRT	High pressure	25 steps per period	6453
TOK	Single token	1 token per trader	6453
SML	Small market	2 buyers, 2 sellers	0007
LAD	Low adaptivity	Same as BASE	6453

4 Outcome Metrics

We evaluate market and agent performance using three categories of metrics: allocative efficiency, price convergence, and individual trader performance. These metrics allow us to assess whether a market achieves its theoretical potential (efficiency), whether prices converge to equilibrium (price quality), and whether profits are distributed fairly across participants (trader performance).

We construct the demand schedule $D(q)$ by ordering all buyer valuations v_{ik} in descending order, and the supply schedule $S(q)$ by ordering all seller costs c_{jk} in ascending order. The equilibrium quantity is $Q^* = \max\{q : D(q) > S(q)\}$, and the equilibrium price lies in the interval $S(Q^*) \leq P^* \leq D(Q^*)$, typically computed as the midpoint $P^* = (D(Q^*) + S(Q^*))/2$. The maximum theoretical surplus is $TS^* = \sum_{q=1}^{Q^*}(D(q) - S(q))$.

Figure 2 visualizes this construction. The demand curve steps down as quantity increases (each successive unit has lower value), while the supply curve steps up (each successive unit has higher cost). The competitive equilibrium occurs at the intersection, defining both the efficient quantity Q^* and the benchmark price P^* against which we measure price convergence. The shaded area represents total surplus TS^*, which serves as the denominator in our efficiency calculations: a market achieving 100% efficiency captures this entire area through optimal matching of buyers and sellers.

4.1 Market Efficiency

Allocative efficiency measures the percentage of maximum possible surplus realized:

$$E = \frac{\sum_{t=1}^{T}(v_t - c_t)}{TS^*} \times 100 \tag{1}$$

where v_t and c_t are the redemption value and cost of units exchanged at trade t. Efficiency loss decomposes into V-inefficiency (intra-marginal loss from untraded profitable units) and EM-

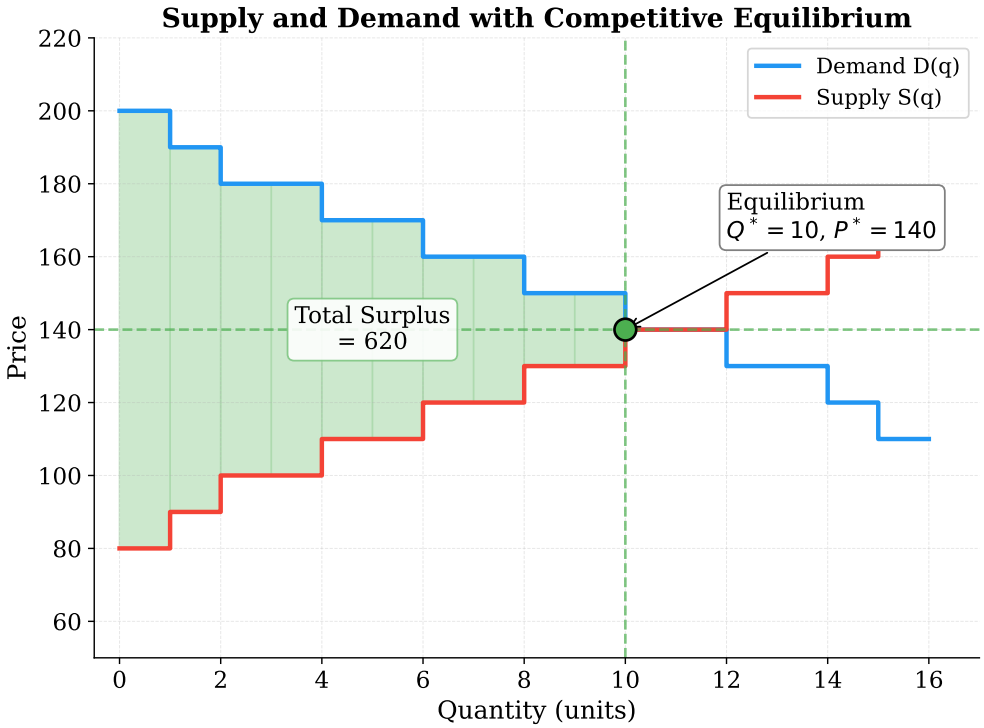

Figure 2. Supply and demand curves with competitive equilibrium. The shaded area represents maximum surplus TS^*, the denominator for efficiency calculations. The equilibrium price P^* provides the benchmark for measuring price convergence.

inefficiency (extra-marginal loss from trades that should not have occurred):

$$IM = \sum_{q \in \text{Untraded Intra-marginal}} (D(q) - S(q)), \quad EM = \sum_{t \in \text{Extra-marginal}} (c_t - v_t) \tag{2}$$

Figure 3 illustrates these concepts. Panel (a) shows maximum possible surplus when all profitable trades execute. Panel (b) shows an inefficient outcome: the green area represents realized surplus, the orange hatched area represents V-inefficiency from missed profitable trades, and the red hatched area represents EM-inefficiency from an extra-marginal trade that destroyed value. This decomposition explains the performance hierarchy we observe in subsequent experiments. Unconstrained random traders (ZI) suffer severe EM-inefficiency by accepting value-destroying trades. Budget-constrained traders (ZIC) eliminate EM-inefficiency but may still miss profitable opportunities under time pressure, creating V-inefficiency. Adaptive traders (ZIP) reduce both sources of loss through learned price targeting.

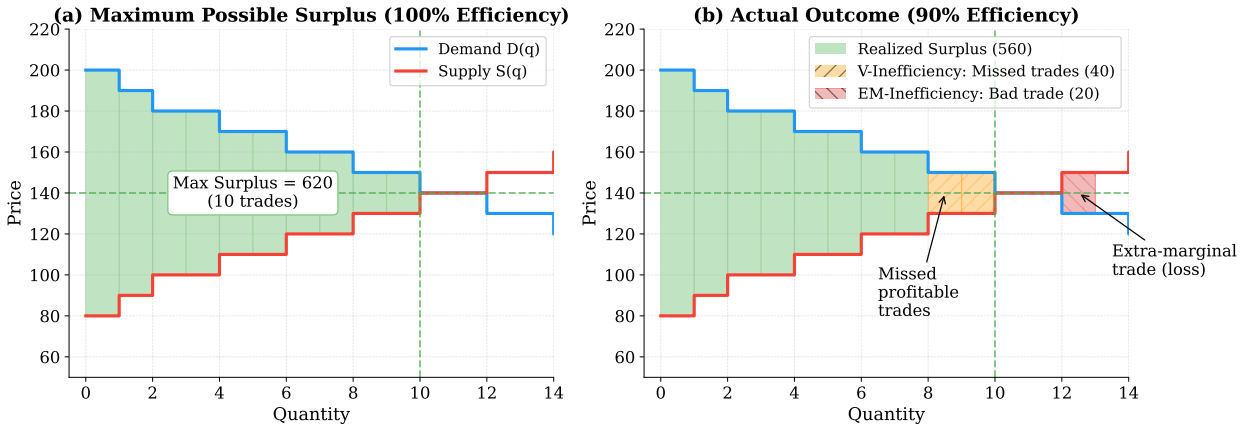

Figure 3. Market efficiency decomposition. (a) Maximum surplus with 100% efficiency. (b) Inefficient outcome showing V-inefficiency (missed trades) and EM-inefficiency (bad trades). Constrained traders eliminate EM-inefficiency; adaptive traders reduce both.

4.2 Price Convergence

Root mean squared deviation measures distance from equilibrium:

$$RMSD = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - P^*)^2} \tag{3}$$

Smith's coefficient of convergence normalizes by equilibrium price: $\alpha = 100 \cdot RMSD/P^*$. Price volatility measures dispersion around the mean transaction price:

$$\text{Volatility} = \frac{\sigma_p}{\bar{p}} \times 100, \quad \text{where } \sigma_p = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - \bar{p})^2} \tag{4}$$

4.3 Trader Performance

Individual profit for buyers is $\pi_i = \sum_k (v_{ik} - p_k)$ and for sellers is $\pi_j = \sum_k (p_k - c_{jk})$. Equilibrium profit represents the theoretical profit if all trades occurred at P^*:

$$\pi_i^* = \sum_{k:v_{ik}>P^*} (v_{ik} - P^*) \text{ (buyers)}, \quad \pi_j^* = \sum_{k:c_{jk}<P^*} (P^* - c_{jk}) \text{ (sellers)} \tag{5}$$

The individual efficiency ratio $E_i = \pi_i/\pi_i^*$ measures whether a trader captures more ($E_i > 1$) or less ($E_i < 1$) than their equilibrium share. Profit dispersion measures cross-agent inequality:

$$PD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\pi_i - \pi_i^*)^2} \tag{6}$$

Lower dispersion indicates more equitable surplus allocation.

4.4 Behavioral Metrics

Beyond outcome metrics, we characterize each strategy's trading behavior using action-level statistics computed from market event logs.

The **dominant action** identifies the most frequent action type: PASS (no action), Shade (improve best quote by small margin), JUMP (aggressive price improvement), SNIPE (accept standing quote when spread narrows), or ACCEPT (immediate market order). We report the percentage of decision opportunities where each action was chosen.

Trade timing captures when trades occur within a period. Let τ_t denote the time step of trade t within a period of length T_{max}:

$$\bar{\tau} = \frac{1}{|T|} \sum_{t \in T} \tau_t, \quad \text{Early\%} = \frac{|\{t : \tau_t < 0.4 \cdot T_{max}\}|}{|T|} \times 100 \tag{7}$$

High Early% indicates aggressive early trading; low values suggest patient waiting strategies.

Spread responsiveness (SR) measures how quote aggressiveness correlates with the bid-ask spread $s_t = a_t^* - b_t^*$:

$$SR = \text{Corr}(\text{shade}_t, s_t) \tag{8}$$

where shade_t is the margin between the submitted quote and the agent's limit price. Positive SR indicates more aggressive pricing when spreads are wide; negative SR indicates caution.

Price improvement rate (PIR) measures how often an agent's quote crosses the spread to enable immediate trade:

$$PIR = \frac{|\{t : q_t \geq a_t^* \text{ (buyer) or } q_t \leq b_t^* \text{ (seller)}\}|}{|\text{quotes}|} \times 100 \tag{9}$$

The **pass rate** is the percentage of decision opportunities where the agent submits no quote:

$$\text{PASS\%} = \frac{|\{t : a_t = \text{PASS}\}|}{|\text{decisions}|} \times 100 \tag{10}$$

High PASS% indicates a waiting or sniping strategy.

4.5 Tournament Metrics

When strategies compete against each other, we use ranking-based metrics to assess relative performance.

Mean rank orders strategies by profit within each market session, with rank 1 being highest:

$$\bar{R}_i = \frac{1}{|S|} \sum_{s \in S} R_{i,s} \tag{11}$$

where $R_{i,s}$ is strategy i's rank in session s. Lower is better.

Win rate measures the fraction of sessions where a strategy achieves rank 1:

$$W_i = \frac{|\{s : R_{i,s} = 1\}|}{|S|} \times 100 \tag{12}$$

Trades per period measures market activity:

$$\text{Trades/Period} = \frac{1}{|P|} \sum_{p \in P} T_p \tag{13}$$

where T_p is the number of completed transactions in period p.

Invasibility ratio tests whether a focal strategy can exploit a baseline market of ZIC traders:

$$\text{Invasibility} = \frac{\bar{\pi}_{\text{focal}}}{\bar{\pi}_{\text{ZIC}}} \tag{14}$$

Values above 1 indicate the focal strategy extracts more surplus than ZIC in mixed competition.

In subsequent sections, we report these metrics across all ten environments using three experimental designs. In self-play experiments, all agents use identical strategies, testing whether a population of homogeneous traders can coordinate efficiently. In mixed-market experiments, heterogeneous strategies compete, revealing which algorithms can exploit others or maintain performance when surrounded by different behaviors. In tournament experiments, we rank strategies by average profit across all environments, identifying which approaches succeed robustly rather than in specific conditions.

5 Intelligence and Markets

To test foundational hypotheses of market behavior, we replicate and extend the classic experiments of Gode and Sunder (1993) and Cliff and Bruten (1997). We deploy a hierarchy of five zero-intelligence (ZI) strategies into the ten diverse market environments of the Santa Fe tournament. This allows us to systematically evaluate the marginal value of specific cognitive abilities, from simple budget-adherence to adaptive learning.

The hierarchy of strategies is as follows:

- **ZI (Unconstrained):** Submits bids and asks uniformly at random from the full price range [1, 1000]. This agent acts as a control, representing pure noise trading without economic rationality.

- **ZIC (Budget-Constrained):** Submits random prices like ZI but is constrained to never make an unprofitable trade. This replicates the core "Zero-Intelligence Constrained" agent from Gode and Sunder (1993).

- **ZIC2 (Market-Aware):** A ZIC agent that also observes the order book and only submits a random price if it improves the current best bid or ask.
- **ZIP (Adaptive):** A ZIC-style agent that incorporates a simple learning rule (Widrow-Hoff) to adapt its profit margin based on trading success, as introduced by Cliff and Bruten (1997).
- **ZIP2 (Adaptive + Market-Aware):** Combines ZIP's adaptive learning with ZIC2's market awareness.

5.1 Revisiting the Gode-Sunder Hypothesis: The Role of Budget Constraints

First, we test the Gode and Sunder (1993) hypothesis that allocative efficiency is primarily a feature of the market institution, not the intelligence of its participants. The hypothesis posits that forbidding agents from making unprofitable trades is sufficient to achieve high levels of market efficiency.

5.1.1 Performance in Non-Strategic Environments

We begin by testing the strategies against passive "TruthTeller" sellers who offer tokens at their true cost. This "easy-play" setting isolates the buyers' search problem against non-strategic opponents.

The results, presented in Table 2, are consistent with the Gode-Sunder hypothesis. The unconstrained ZI agent exhibits low allocative efficiency, achieving only 29% in the BASE environment. Transaction logs (Figure 4, panel 1) show that ZI's high trade volume (16 trades per period vs. 8 profitable opportunities) is driven by deeply unprofitable, extra-marginal trades that destroy surplus.

The addition of a single budget constraint significantly improves market outcomes. The ZIC agent achieves 93–99% efficiency across all ten environments. As Gode and Sunder found, the institutional rule—not trader savvy—is a principal driver of allocative efficiency. Further additions of intelligence (ZIC2, ZIP) provide only marginal gains in this setting, pushing efficiency to nearly 100%.

Table 2. Easy-play allocative efficiency (%) across all market environments. Buyers compete against passive TruthTeller sellers who ask at true cost. ZI is wildly inefficient (11–67%) because it accepts loss-making trades. Adding a budget constraint (ZIC1) achieves 93–99% efficiency. Market awareness (ZIC2) and learning (ZIP1, ZIP2) saturate efficiency at 100% in all environments.

Strategy	BASE	BBBS	BSSS	EQL	RAN	PER	SHRT	TOK	SML	LAD
ZI	29	52	67	25	11	29	29	95	27	25
ZIC1	99	96	99	97	99	98	94	93	98	97
ZIC2	100	99	100	100	100	100	100	100	100	100
ZIP1	100	100	100	100	100	100	100	100	100	100
ZIP2	100	100	100	100	100	100	100	100	100	100

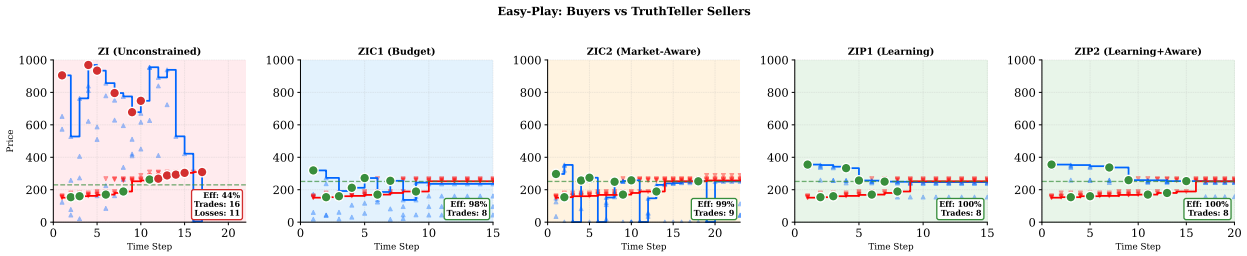

Figure 4. Easy-play market dynamics for five zero-intelligence strategies against passive TruthTeller sellers. ZI produces scattered prices across the full range with many loss-making trades; ZIC constrains trading to profitable prices only; ZIP achieves immediate execution through learning.

5.1.2 Performance under Strategic Pressure

In a more challenging self-play setting, where all traders employ the same strategy, the budget constraint remains the most critical factor for efficiency. Table 3 shows ZIC's efficiency reaches 91% in the BASE environment, a 64-point increase from ZI's 27%. This demonstrates the robustness of the Gode-Sunder finding. Again, ZI's high trading volume (Table 4) is shown to be counter-productive, consisting of surplus-destroying transactions, as visualized by the scattered prices in Figure 5. The results suggest that the most significant leap in market performance comes from the mechanical imposition of budget constraints.

Table 3. Selfplay Allocative Efficiency (%) Across All Market Environments

Environment	**ZI**	**ZIC1**	**ZIC2**	**ZIP1**	**ZIP2**
BASE	27±2	91±2	95±1	100±0	100±0
BBBS	53±2	83±2	88±2	100±0	100±0
BSSS	53±2	88±1	92±1	100±0	100±0
EQL	29±4	92±1	95±1	100±0	100±2
RAN	13±1	99±0	99±0	100±0	100±0
PER	27±2	91±2	94±2	100±0	100±0
SHRT	27±2	66±2	76±2	100±0	100±1
TOK	94±2	75±3	81±3	100±0	100±0
SML	29±3	87±1	91±1	100±0	100±0
LAD	29±4	92±1	95±1	100±0	100±2

Mean ± std over 10 seeds × 100 rounds. ZIP1/ZIP2 achieve 100% everywhere.

Table 4. Selfplay Trades per Period Across All Environments

Environment	ZI	ZIC1	ZIC2	ZIP1	ZIP2
BASE	16.0	7.0	7.3	7.9	8.2
BBBS	8.0	4.2	4.6	5.7	5.9
BSSS	8.0	4.9	5.5	5.7	5.9
EQL	16.0	6.6	6.9	7.4	7.7
RAN	16.0	7.6	7.7	7.2	7.7
PER	16.0	7.0	7.3	8.0	8.2
SHRT	16.0	4.4	4.9	7.8	8.2
TOK	4.0	1.0	1.2	1.9	1.9
SML	8.0	3.0	3.3	3.9	4.0
LAD	16.0	6.5	6.9	7.3	7.7

ZI trades all tokens (often at a loss); constrained strategies are selective.

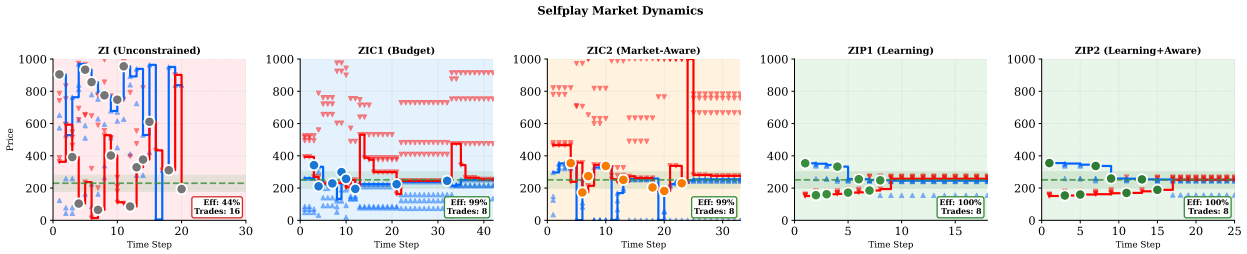

Figure 5. Selfplay market dynamics for five zero-intelligence strategies. ZI produces scattered prices across the full range; ZIC constrains trading near equilibrium; ZIP achieves tight convergence through learning.

5.2 Revisiting the Cliff-Bruten Hypothesis: The Role of Adaptive Intelligence

While the budget constraint is sufficient for high allocative efficiency, Cliff and Bruten (1997) argued that intelligence, in the form of adaptive learning, is essential for achieving the price stability and coherence observed in human markets. Our results support a refined version of this hypothesis: intelligence is critical for the *speed and certainty* of convergence, thereby reducing coordination failures.

5.2.1 Coordination Failures in Non-Adaptive Markets

While ZIC is highly efficient, its random-search mechanism is slow and prone to failure under scarcity. This weakness is most apparent in the **SHRT** environment, where its efficiency drops to 66% (Table 3). The V-Inefficiency metric (Table 5), which counts profitable but unexecuted trades, reveals the mechanism: in SHRT, ZIC misses 227 intra-marginal trades per period. This pattern persists in other scarce environments like TOK (42 missed trades) and SML (37 missed trades).

Because ZIC agents rely on chance for a bid and ask to cross, their ability to coordinate degrades when opportunities are limited by time or market size.

5.2.2 Adaptive Learning and Market Coherence

Adaptive learning (ZIP) largely resolves this coordination problem. The ZIP strategy achieves nearly 100% allocative efficiency across all ten environments, with V-Inefficiency scores close to zero universally (Table 5). This is also reflected in the "easy-play" experiments (Table 7), where ZIP's mean trade time is just 1.0 step, while ZIC's is 1.6-3.1 steps. ZIP does not need to search for the price; it learns it, leading to faster and more certain trade execution.

5.2.3 Behavioral Signatures and Price Dynamics

The behavioral signatures in Table 8 show that ZIP is strongly JUMP-dominant (57%), actively trying to improve the price. Its Price Improvement Rate (PIR) of just 1% indicates it has learned to avoid submitting bids that will not cross the spread. In a departure from the original Cliff and Bruten (1997) findings, our results show that this superior coordination does not necessarily lead to lower price volatility. In the BASE environment, ZIP's price volatility is 12%, compared to ZIC's 7% (Table 6). The primary contribution of adaptive intelligence in our experiments is the near-elimination of coordination failures, ensuring the market clears quickly and completely.

Table 5. Selfplay V-Inefficiency (Missed Trades) Across All Environments

Environment	**ZI**	**ZIC1**	**ZIC2**	**ZIP1**	**ZIP2**
BASE	0	30±3	16±2	3±1	0±0
BBBS	0	61±4	47±4	1±0	0±0
BSSS	0	26±3	8±1	1±0	0±0
EQL	0	36±3	20±3	2±1	0±0
RAN	0	8±2	5±1	34±5	0±0
PER	0	27±3	16±3	1±0	0±0
SHRT	0	227±15	165±8	3±1	3±22
TOK	0	42±3	27±3	0±0	0±0
SML	0	37±4	21±3	0±0	0±0
LAD	0	35±2	20±3	2±1	0±0

Missed intra-marginal trades per period. ZI = 0 (trades everything); ZIC high in SHRT.

Table 6. Selfplay Price Volatility (%) Across All Market Environments

Environment	**ZI**	**ZIC1**	**ZIC2**	**ZIP1**	**ZIP2**
BASE	65±1	7±1	8±1	12±1	14±12
BBBS	51±1	6±0	7±1	11±1	13±11
BSSS	79±1	8±1	9±1	12±1	15±14
EQL	65±1	7±1	8±1	11±1	12±8
RAN	65±1	31±1	33±1	54±1	55±17
PER	65±1	7±1	8±1	12±1	14±12
SHRT	65±1	7±1	8±1	12±1	14±12
TOK	57±1	2±0	2±0	4±1	4±8
SML	56±1	6±1	7±1	11±1	12±10
LAD	65±1	7±1	8±1	12±1	12±8

Volatility = price std / mean. Lower is better. ZI volatility 10× higher than constrained.

Table 7. Easy-play mean trade time (steps) across all market environments. Lower values indicate faster search. ZI and ZIP1/ZIP2 execute immediately (1.0 steps) but for opposite reasons: ZI accepts any price indiscriminately, while ZIP learns to target the seller's ask precisely. ZIC1 requires 1.6–3.1 steps because random sampling within budget bounds is a geometric waiting time problem.

Strategy	BASE	BBBS	BSSS	EQL	RAN	PER	SHRT	TOK	SML	LAD
ZI	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.1	1.0
ZIC1	1.6	1.4	2.6	2.3	1.0	1.6	1.7	3.1	2.4	1.9
ZIC2	1.2	1.2	1.4	1.2	1.0	1.5	1.2	1.3	1.3	1.2
ZIP1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
ZIP2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 8. Selfplay Behavioral Signatures (All 8 agents same strategy, 5 seeds × 5 periods)

Strategy	**Dominant Action**	**Trade Time**	**Early%**	**PASS%**	**SR**	**PIR**	**Profit/Trade**
ZI	PASS (88%)	8.0	100	88	−0.41	35	−92.3
ZIC1	JUMP (45%)	23.8	75	5	−0.12	7	51.2
ZIC2	JUMP (46%)	13.4	90	16	0.08	18	34.7
ZIP1	JUMP (57%)	4.9	100	4	−0.15	1	**64.6**
ZIP2	PASS (45%)	6.2	98	45	−0.18	2	62.1

Trade Time = mean timestep when trades occur (out of 100). Early% = trades in first 30 steps.
SR = Spread Responsiveness. PIR = Price Improvement Rate (% of quotes crossing spread).

5.3 Strategic Interaction in Heterogeneous Markets

While self-play experiments establish baselines, placing strategies in a heterogeneous "mixed competition" market—with one buyer and one seller of each constrained type (ZIC, ZIC2, ZIP, ZIP2)—reveals insights into direct strategic interaction.

5.3.1 The Value of Market Information in Varied Environments

Our experiments suggest the value of market awareness is highly context-dependent. Table 9 shows that ZIC2 (market-aware) outperforms ZIC (unaware) only under specific forms of resource scarcity. In the **SHRT** environment (scarce time), ZIC2 earns 42% more profit. In the **BBBS** environment (scarce sellers), it earns 198% more. However, in the volatile **RAN** environment, awareness is a liability, costing ZIC2 28% of its profits relative to the "blind" ZIC, likely due to agents acting on stale information.

Table 9. Market Awareness: ZIC2 vs ZIC1 Profit Gap Across Environments

Environment	**ZIC1 Profit**	**ZIC2 Profit**	**Gap**	**Gap (%)**
BASE	61	54	−7	−11
BBBS	11	34	+22	+198
EQL	61	59	−2	−3
RAN	849	613	−236	−28
PER	61	63	+2	+3
SHRT	33	46	+14	+42
TOK	8	13	+5	+62
LAD	62	50	−12	−20

BSSS and SML omitted (ZIC strategies not viable). Context determines value of awareness.

5.3.2 Profit Mechanisms: High-Margin vs. High-Volume Strategies

The mixed market also reveals that ZIC and ZIP achieve similar profitability through different mechanisms. In the BASE environment, ZIC and ZIP earn nearly identical total profits (61 and 64, respectively, see Table 10). However, ZIC achieves the highest profit per trade (39.3) by occasionally capturing large-surplus "windfall" trades. In contrast, ZIP earns less per trade (30.0) but executes a higher volume of transactions. This reveals a classic strategic trade-off between a low-volume, high-margin strategy and a high-volume, consistent-margin strategy.

5.3.3 The Cost of Heuristic Passivity

A seemingly rational heuristic—waiting for a guaranteed price improvement—can be systematically punished. ZIP2, which combines learning with a rule that it must PASS if it cannot improve the current price, performs poorly in direct competition. Its PASS rate of 45% in self-play is a quantitative signature of this patience. As shown in Table 11, this patient agent earns significantly less than its more aggressive ZIP counterpart across all ten environments, with a profit gap of 61% in BASE. Its passivity causes it to miss trading opportunities captured by more aggressive agents.

Table 10. Mixed Competition: Strategy Performance in BASE Environment

Strategy	Profit	Rank	Win Rate (%)	Profit/Trade
ZIC1	61	2.1	31	39.3
ZIC2	54	2.4	26	27.0
ZIP1	**64**	**2.3**	**32**	30.0
ZIP2	25	3.2	10	11.8

4 buyers (1 each strategy) vs 4 sellers (1 each strategy). 100 rounds × 10 periods. ZI excluded as loss-making distorts analysis. Bold = best in column.

Table 11. Institutional Blindness: ZIP2 vs ZIP1 Profit Gap Across Environments

Environment	ZIP1 Profit	ZIP2 Profit	Gap	Gap (%)
BASE	64	25	−39	−61
BBBS	30	9	−22	−71
BSSS	87	32	−55	−63
EQL	60	21	−39	−65
RAN	781	596	−185	−24
PER	78	23	−55	−70
SHRT	66	26	−40	−61
TOK	11	4	−7	−64
SML	59	12	−47	−80
LAD	65	21	−44	−68

ZIP2's market-aware PASS rule systematically reduces profit by 24–80%.

5.3.4 Market Sophistication and the Distribution of Surplus

Finally, our results show a nuanced relationship between market sophistication and economic inequality. By the Gini coefficient, the unconstrained ZI market appears the most "equal" (0.26), while the adaptive ZIP market appears more unequal (0.43) (Table 12).

However, this metric can be misleading if not contextualized. In the ZI market, the bottom 50% of traders achieve a negative profit share (−15.6%). The apparent equality is an artifact of widespread surplus destruction. In contrast, in the ZIP market, the bottom 50% of traders secure a positive profit share of 18%. The Max/Mean profit ratio also falls from 52x in the ZI market to just 2x in the ZIP market. Sophistication provides a critical welfare floor by preventing the catastrophic losses that characterize the ZI market, leading to a more robust distribution of surplus among participants.

Table 12. Selfplay Inequality Metrics (BASE Environment, 3 Seeds × 10 Rounds × 10 Periods)

Metric	ZI	ZIC1	ZIC2	ZIP1	ZIP2
Gini	0.26	0.39	0.41	0.43	0.44
Max/Mean Ratio	52.0	1.9	2.0	2.1	2.2
Bottom-50% Share	−15.6%	28.5%	23.3%	18.0%	16.5%
Skewness	+0.03	+0.20	+0.20	+0.12	+0.14

Gini = profit concentration (0 = equal, 1 = one agent takes all).

Max/Mean = highest earner relative to average. Bottom-50% = share captured by lower half.

5.4 Summary of Findings

Our replication and extension of these foundational experiments in the diverse Santa Fe environments lead to four primary conclusions:

1. **Allocative efficiency is primarily an institutional outcome.** Our results are consistent with the Gode-Sunder hypothesis. The imposition of a simple budget constraint is the single most significant factor in achieving high allocative efficiency, a finding that holds across both simple and complex strategic environments.

2. **Adaptive intelligence is essential for market coherence.** Our findings support a refined version of the Cliff-Bruten hypothesis. While not always necessary for high efficiency, adaptive intelligence is critical for the speed and certainty of convergence, allowing agents to overcome the coordination failures that plague non-adaptive strategies in scarce environments.

3. **The value of simple heuristics is context-dependent.** In heterogeneous markets, the value of a given heuristic is not universal. The utility of market awareness depends on the market's structure, and seemingly rational "patient" behaviors can be systematically penalized in a competitive ecosystem.

4. **Sophistication improves welfare by preventing market failure.** More sophisticated markets may not appear more "equal" by standard distributional metrics. However, they provide a crucial welfare floor by eliminating the catastrophic, surplus-destroying losses that characterize markets populated by unconstrained agents, ensuring a more stable and robust distribution of gains.

6 The Santa Fe Double Auction

This section revisits foundational questions posed by the original Santa Fe Double Auction Tournament (Rust et al., 1994) and subsequent analyses (Chen Tai, 2010). Our objective is to investigate the ecological dynamics of a diverse set of heuristic-based trading strategies, examining the mechanisms behind their competitive success or failure within a complex market ecosystem. Our analysis

proceeds by first examining easy play scenarios, then homogeneous self-play, and finally the dynamics of heterogeneous mixed play.

First, we investigate the competitive performance of simple heuristics, examining the factors contributing to the success of "sniper" strategies such as Kaplan and Ringuette in heterogeneous market environments. Second, we examine market stability in homogeneous populations, testing whether opportunistic strategies lead to collective market instability and liquidity collapse when agents interact only with copies of themselves, as hypothesized by Rust et al. Third, we analyze the relationship between strategic complexity and ecological fitness, comparing fixed-rule agents against more cognitively elaborate models to understand the trade-offs involved.

Our experimental design involves three distinct setups across the ten Santa Fe market environments: an Invasibility experiment (1 challenger vs. 7 ZIC agents) to assess exploitative capabilities, a Homogeneous Self-Play experiment (8 identical agents) to evaluate collective stability, and a Heterogeneous Round-Robin Tournament (all agents competing) to gauge overall ecological fitness. The agent roster for these experiments includes ZIC (baseline), Skeleton (simple fixed-rule), Kaplan (original sniper), Ringuette (alternative sniper), EL (Ledyard, reservation price model), BGAN (belief-based), Staecker, Gamer, Jacobson, Perry (adaptive parameters), Lin (statistical), and Breton.

6.1 Easy Play Performance (Invasibility Experiment)

Our investigation into easy play scenarios focuses on the "Invasibility Experiment," where individual Santa Fe strategies compete against a homogeneous population of baseline ZIC agents. This setup quantifies the exploitative capabilities of each strategy, measuring their ability to extract surplus from a less sophisticated, liquidity-providing market. Table 13 presents the profit ratio for each Santa Fe agent when playing against ZIC in the BASE environment, indicating how many times more profit they generate compared to a ZIC agent in the same scenario.

As Table 13 demonstrates, several strategies, particularly Perry (1.005), Gamer (0.995), and EL (0.978), exhibit strong exploitative capabilities, generating profits comparable to or exceeding the ZIC baseline when acting as invaders. Ringuette (0.970) and Jacobson (0.961) also show robust performance in this easy play environment. This highlights that strategies capable of adapting to or exploiting simpler market dynamics tend to fare well in asymmetric matchups.

6.2 Self Play Performance

The original Santa Fe paper hypothesized that the success of parasitic strategies like Kaplan might be contingent on a diverse market, and that a market composed solely of such agents could lead to collective instability. Our homogeneous self-play experiments offer substantial evidence supporting the "Sniper's Dilemma" hypothesis.

Table 14 provides a summary of key self-play metrics for each Santa Fe strategy in the BASE environment.

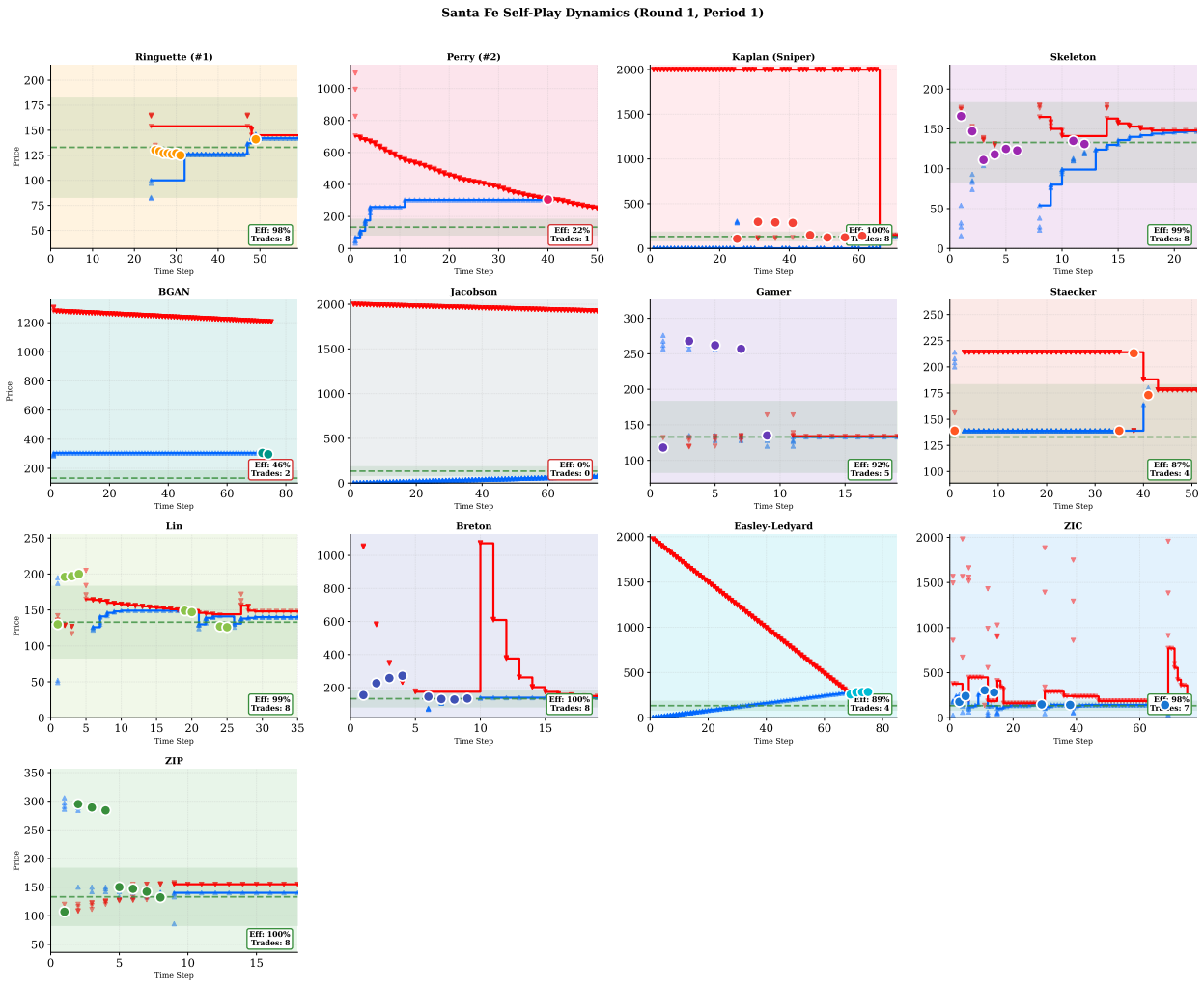

Figure 6. Self-play trading dynamics for all 13 Santa Fe strategies in the BASE environment. Each panel shows bid submissions (blue triangles), ask submissions (red triangles), best bid/ask trajectories (step lines), and executed trades (large colored circles). The green dashed line indicates the competitive equilibrium price. Ringuette and Kaplan (snipers) achieve high efficiency through patient opportunism, while Skeleton and Breton (fixed-rule) maintain steady trading activity. ZIP (adaptive learning) and Perry (heuristic) show distinct trading patterns reflecting their strategic approaches.

Table 13. Easy Play Profit Ratios (vs. ZIC, BASE Environment)

Strategy	Profit Ratio vs. ZIC
BGAN	0.608
Breton	0.811
EL	0.978
Gamer	0.995
Jacobson	0.961
Kaplan	0.844
Lin	0.876
Perry	1.005
Ringuette	0.970
Skeleton	0.839
Staecker	0.837
ZIC	1.000

A profit ratio greater than 1 indicates the strategy extracts more profit than a ZIC agent would in the same scenario.

Table 14. Self Play Metrics (BASE Environment)

Strategy	Efficiency (%)	Volatility	Trades/Period
BGAN	65.54	1.73	1.01
Breton	99.84	11.18	2.25
EL	85.49	Inf	1.67
Gamer	65.82	Inf	1.12
Jacobson	40.07	Inf	0.84
Kaplan	100.00	18.38	2.00
Lin	66.03	Inf	1.55
Perry	61.59	Inf	1.39
Ringuette	97.77	4.49	2.22
Skeleton	99.54	7.85	2.11
Staecker	48.75	Inf	0.93
ZIC	90.80	Inf	1.69

Efficiency, Volatility, and Trades per Period for each strategy in homogeneous self-play (BASE environment). 'Inf' for volatility indicates insufficient data due to very few or no trades.

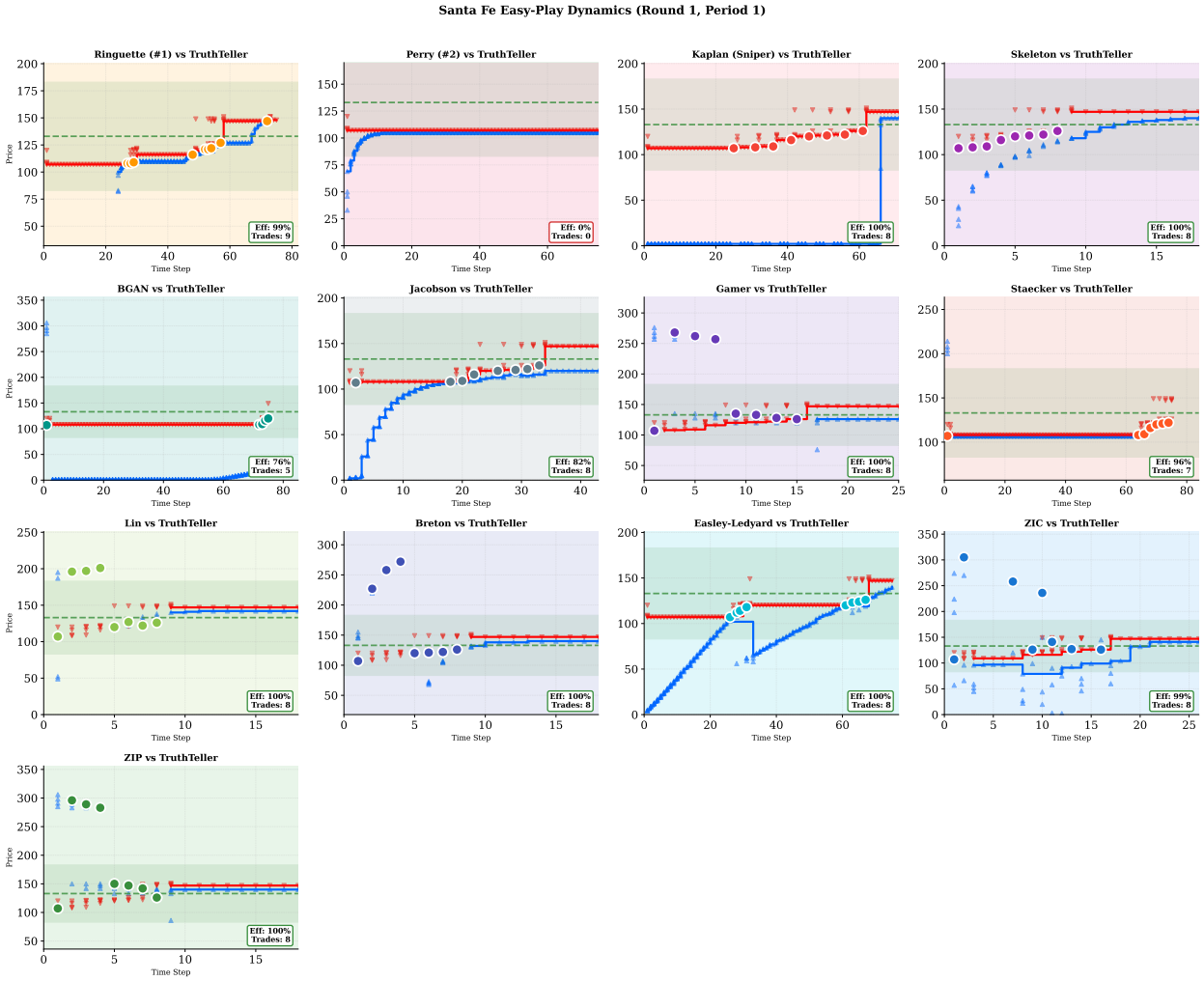

Figure 7. Easy-play trading dynamics for all 13 Santa Fe strategies against TruthTeller sellers in the BASE environment. Each panel shows bid submissions (blue triangles), ask submissions (red triangles), best bid/ask trajectories (step lines), and executed trades (large colored circles). The green dashed line indicates the competitive equilibrium price. Most strategies achieve high efficiency when exploiting passive sellers, with the notable exception of Perry (0% efficiency) which appears to rely on competitive market dynamics that are absent when facing TruthTellers.

The market efficiency of sniper agents exhibits a substantial decline under specific conditions. In the SHRT (short time) environment, Ringuette's self-play efficiency plummets to 32.5% (from 98.1% in BASE), and Kaplan's falls to 79.5%. This contrasts sharply with the near-perfect efficiency maintained by fixed-rule agents like Skeleton (99.7%) and Breton (99.8%) in the same stressful environment.

Table 15. Self-Play Profit Dispersion (RMS)

Strategy	BASE	BBBS	BSSS	EQL	RAN	PER	SHRT	TOK	SML	LAD
BGAN	141	75	110	129	849	141	166	15	56	129
Breton	51	51	51	45	470	51	51	51	51	45
EL	51	89	122	67	68	68	444	163	46	67
Gamer	218	218	218	294	580	218	218	218	218	294
Jacobson	571	571	571	634	473	571	571	571	571	634
Kaplan	60	54	51	64	683	66	91	19	47	64
Lin	403	403	403	400	291	403	403	403	403	400
Perry	349	349	349	300	158	349	349	349	349	300
Ringuette	21	46	45	13	291	22	327	30	17	13
Skeleton	26	34	34	29	204	41	33	16	20	29
Staecker	486	450	420	474	515	487	500	214	368	474
ZIC	53	48	55	48	441	53	90	49	51	48

Lower RMS indicates more equitable profit distribution among traders.

In contrast, fixed-rule agents such as Skeleton and Breton generally demonstrate highly stable and efficient self-play across environments. Their high trades per period, low RMSD, and minimal profit dispersion (Tables 17, 16, and 15) signify their ability to foster orderly and equitable markets when interacting solely amongst themselves. This fundamental robustness highlights their role as reliable liquidity providers, a stark contrast to the destabilizing dynamics observed in homogeneous sniper populations. This supports the view that the competitive advantage of sniping strategies is parasitic, dependent on the presence of agents willing to provide liquidity within the market.

6.3 Mixed Play Performance (Heterogeneous Round-Robin Tournament)

Our analysis of the Heterogeneous Round-Robin Tournament, where all Santa Fe agents compete against each other, reveals insights into their overall ecological fitness. Table 18 summarizes the average rank and total wins for each strategy across the diverse market environments.

The original Santa Fe tournament notably highlighted the unexpected victory of Kaplan, a relatively simple heuristic, over more complex, AI-driven strategies. Our replication in a heterogeneous round-robin tournament is consistent with the competitive power of simple heuristics, though it reveals a notable shift in the most dominant sniper.

In our replication, Ringuette ranks highest in the overall tournament, achieving the best aver-

Table 16. Self-Play RMSD (Root Mean Squared Deviation)

Strategy	BASE	BBBS	BSSS	EQL	RAN	PER	SHRT	TOK	SML	LAD
BGAN	8	4	22	3	48	10	4	2	22	3
Breton	40	40	40	36	392	40	40	40	40	36
EL	12	11	13	12	60	26	10	2	11	12
Gamer	25	25	25	21	503	25	25	25	25	21
Jacobson	6	6	6	3	50	6	6	6	6	3
Kaplan	49	51	49	51	565	51	60	19	46	51
Lin	9	9	9	7	77	9	9	9	9	7
Perry	11	11	11	8	30	11	11	11	11	8
Ringuette	14	8	7	8	40	15	1	0	6	8
Skeleton	21	26	26	22	160	30	24	16	18	22
Staecker	9	9	10	7	270	9	10	1	9	7
ZIC	35	30	33	32	349	35	34	11	27	32

Lower RMSD indicates better price convergence to equilibrium.

Table 17. Self-Play Trades per Period

Strategy	BASE	BBBS	BSSS	EQL	RAN	PER	SHRT	TOK	SML	LAD
BGAN	1.0	0.6	1.8	0.8	1.2	1.0	0.7	0.4	1.6	0.8
Breton	2.3	2.3	2.3	1.9	1.8	2.3	2.3	2.3	2.3	1.9
EL	1.7	0.8	1.9	1.5	1.8	1.1	0.7	0.1	1.5	1.5
Gamer	1.1	1.1	1.1	1.0	1.6	1.1	1.1	1.1	1.1	1.0
Jacobson	0.8	0.8	0.8	0.7	1.5	0.8	0.8	0.8	0.8	0.7
Kaplan	2.0	1.0	3.0	1.9	1.8	2.2	1.2	0.5	1.8	1.9
Lin	1.5	1.5	1.5	1.3	1.6	1.5	1.5	1.5	1.5	1.3
Perry	1.4	1.4	1.4	1.3	1.7	1.4	1.4	1.4	1.4	1.3
Ringuette	2.2	0.7	2.4	1.9	1.5	2.3	0.6	0.1	1.7	1.9
Skeleton	2.1	1.0	3.1	1.9	1.9	2.3	2.2	0.5	1.9	1.9
Staecker	0.9	0.4	1.5	0.8	1.4	0.9	0.7	0.1	1.1	0.8
ZIC	1.7	0.7	2.5	1.5	1.8	1.9	1.1	0.3	1.4	1.5

Average number of trades completed per period in self-play.

Table 18. Mixed Play Overall Performance (Avg Rank and Wins)

Strategy	Average Rank	Total Wins
ZIC	6.64	54
Skeleton	5.18	18
Kaplan	6.99	10
Ringuette	4.00	149
Gamer	5.81	55
Perry	5.00	56
Ledyard	6.06	50
BGAN	8.30	22
Staecker	7.02	32
Jacobson	7.20	14
Lin	8.42	28
Breton	7.39	12

Average rank and total wins for each strategy in the Heterogeneous Round-Robin Tournament across all environments.

age rank (4.00) and the highest number of wins (149) across the ten diverse environments. This contrasts with Kaplan, which ranks 7th, suggesting that while sniping remains an effective strategy, Ringuette's specific implementation appears to confer a more robust competitive advantage in this contemporary setup. For instance, in Round 1 of the BASE environment, Ringuette demonstrated its opportunistic nature by executing a profitable trade with Staecker (price 144, buyer profit 148) and another with Lin (price 146, buyer profit 2), showcasing its ability to extract surplus even with small margins. Kaplan, too, exhibited this behavior, securing a notable profit in a trade with Breton (price 141, buyer profit 145). Other notably strong performers include Perry (2nd place) and Skeleton (3rd place), both representing relatively simple heuristic approaches. This suggests that the underlying principles behind the Kaplan Paradox may extend to a broader class of simple, robust, and opportunistic strategies, rather than being exclusive to Kaplan.

6.4 Complexity and Performance Trade-offs

Our results offer further insights into the relationship between a strategy's design complexity and its ecological fitness in competitive markets. The evidence suggests that increased strategic complexity does not consistently translate to superior performance, potentially incurring a "penalty for overhead."

While adaptive and belief-based models represent higher orders of cognitive sophistication, the overall Round-Robin tournament results tend to indicate that simpler, more direct heuristics frequently outperform them. The top three performers (Ringuette, Perry, Skeleton) embody relatively straightforward rule-sets. In contrast, cognitively more elaborate agents like Jacobson (equilibrium estimation, ranks 9th), Staecker (ranks 8th), BGAN (belief-based, ranks 11th), and Lin (statistical

prediction, ranks 12th) generally occupy the lower half of the rankings. For example, in homogeneous self-play, Jacobson and BGAN frequently exhibit complete market breakdowns, failing to execute any trades across entire rounds, a clear indication that their sophisticated mechanisms can lead to inaction and inefficiency when interacting with similar complex counterparts.

This finding reinforces aspects of the "Kaplan Paradox" from a different perspective: the most ecologically fit strategies are often characterized by their simplicity, robustness, and speed of execution rather than their ability to model complex market dynamics or form sophisticated beliefs. The trade-off between the cognitive overhead of complex models and the ecological robustness of simpler heuristics appears to favor the latter in this competitive environment.

6.5 Analysis of Evolutionary Dynamics

Beyond static tournament performance, we investigate the long-term ecological dynamics of these strategies using an evolutionary tournament model. This experiment simulates multiple generations of competition, where successful strategies increase their population share and underperforming ones face elimination.

Our results from 10 evolutionary seeds reveal a clear hierarchy of evolutionary stability. Skeleton emerges as the most evolutionarily stable strategy, comprising 62.5% of the final population across seeds. Its simple, robust fixed-rule approach consistently outperforms more complex or adaptive strategies in a dynamic competitive environment. Kaplan (13.4%) and Ringuette (10%) maintain stable presences in the final population. This indicates that their opportunistic, parasitic strategies are evolutionarily viable, albeit not dominant, suggesting they can coexist sustainably within the market ecosystem. ZIP, representing general adaptive learning, consistently goes extinct early (by generation 7.5 across all seeds). This suggests that despite its strong self-play and niche dominance in asymmetric static markets, its adaptive mechanism is not sufficiently robust or specialized to compete effectively in a dynamic evolutionary environment, leading to its elimination.

6.6 Summary and Conclusion

Our replication of the Santa Fe Tournament, analyzed through the lens of foundational literature, offers the following key insights into the ecological dynamics of heuristic-based trading strategies:

1. Simple, rule-based strategies, particularly those employing effective opportunistic behavior (e.g., Perry, Gamer, Ringuette), demonstrate substantial exploitative power and competitive advantage in heterogeneous market ecosystems. Their success is associated with effective opportunistic behavior. While Kaplan is historically noted for sniper tactics, our easy play results suggest other agents achieve higher exploitative ratios against ZIC baselines.

2. The collective behavior of opportunistic or complex strategies, not limited to purely sniper populations, can lead to severe market inefficiency and liquidity collapse in homogeneous environments. Our self-play experiments reveal that many adaptive and belief-based agents

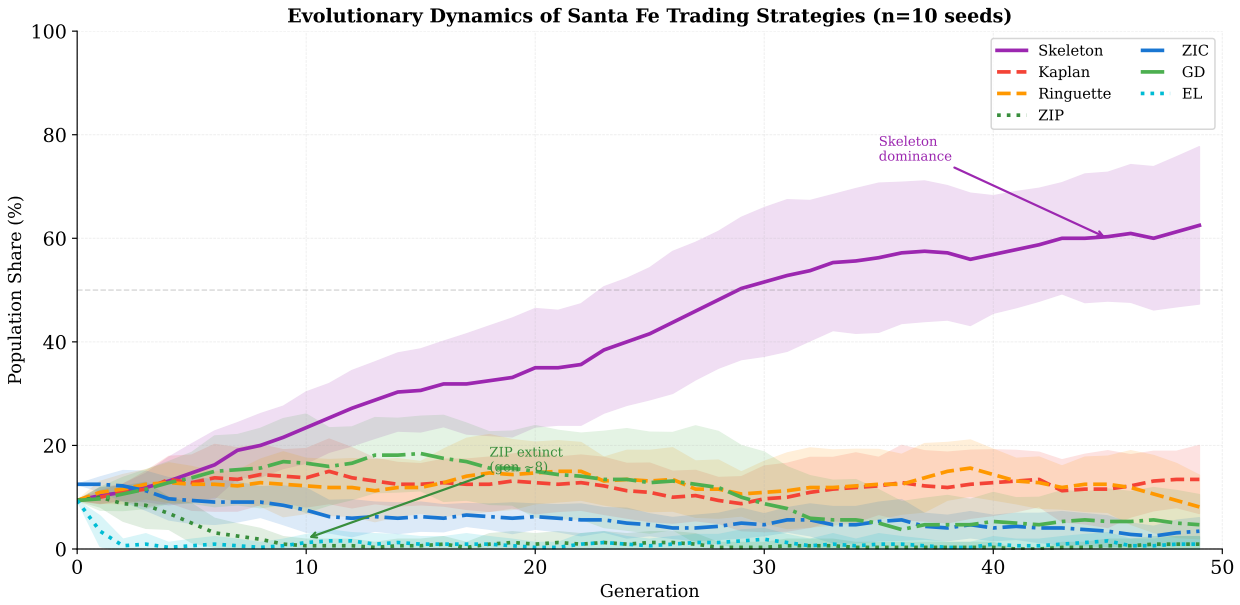

Figure 8. Evolutionary dynamics of Santa Fe trading strategies over 50 generations, aggregated across 10 seeds. Lines show mean population share with shaded bands indicating one standard deviation. Skeleton (purple) dominates with 62.5% of the final population. Snipers Kaplan (red) and Ringuette (orange) persist at lower levels. ZIP (green dotted) goes extinct early, typically by generation 8. The emergence of Skeleton dominance and ZIP extinction demonstrates that simple, robust fixed-rule strategies outcompete both opportunistic snipers and general adaptive learners in an evolutionary competition.

(e.g., Jacobson, Staecker, BGAN, Gamer, Perry, Lin, EL) exhibit significant market breakdown, characterized by low efficiency, negligible trades, and high volatility, even in the BASE environment. This confirms that the success of such strategies is fundamentally dependent on the presence of diverse or more liquidity-providing agents. In stark contrast, simple fixed-rule agents like Skeleton and Breton consistently foster highly stable and efficient self-play markets, underscoring their inherent robustness and role as reliable liquidity providers.

3. There is no clear evidence that increased strategic complexity consistently leads to superior ecological fitness. Instead, simpler, robust heuristics like Ringuette, Perry, and Skeleton often prove highly effective, suggesting that an optimal balance of simplicity and exploitative capability is favored in this competitive environment. The marked struggles of more cognitively elaborate agents (e.g., Jacobson, Staecker, BGAN, Lin, EL, Gamer, Perry) in homogeneous self-play, characterized by significant drops in efficiency and liquidity, further underscore a "penalty for overhead" associated with such complexity when robust self-organization is required.

4. Our analysis of general adaptive strategies (ZIP) in an extended round-robin tournament suggests that while not universally dominant, ZIP excels in asymmetric market structures. However, in evolutionary competition, ZIP consistently faces early extinction, suggesting that its adaptive mechanisms are outcompeted by more specialized and robust heuristics over multiple generations.

7 Comparative Analysis of Foundational Findings

8 Reinforcement Learning Trader

8.1 PPO Agent Design

We train a deep reinforcement learning agent using Proximal Policy Optimization (PPO) with action masking (MaskablePPO) to ensure valid bids under AURORA protocol constraints. The agent uses a two-layer neural network with 256 hidden units per layer and receives a 42-dimensional observation including market state, order book information, and private token values. Training proceeds for 10 million timesteps against mixed opponents (ZIC, ZIP, Skeleton, GD, Kaplan) using entropy coefficient decay from 0.15 to 0.01 to encourage policy refinement.

A critical methodological finding concerns role specialization. When the PPO model trained as a buyer was deployed in both buyer and seller roles, performance degraded significantly. The model had never observed the seller perspective during training, leading to poor seller-side decision making. We therefore restrict PPO to buyer-only deployment in all experiments below, matching its training distribution.

8.2 PPO vs Zero-Intelligence Baselines

Section 5 established the zero-intelligence hierarchy: ZI fails catastrophically, ZIC achieves 97% efficiency through the budget constraint alone, and ZIP achieves 99% efficiency through adaptive margin learning. A natural question arises: where does deep reinforcement learning fall in this hierarchy? Can PPO exceed the hand-crafted adaptive learning of ZIP?

We train PPO specifically against ZIC and ZIP opponents for 10^6 timesteps, then evaluate in a mixed market with one agent of each type per role: four buyers (ZI, ZIC, ZIP, PPO) and four sellers (ZI, ZIC, ZIP, ZIC). We run 50 rounds with 10 periods each across 10 random seeds for statistical robustness.

Table 19. PPO vs Zero-Intelligence Baseline Tournament (10 seeds, 50 rounds each)

Strategy	Mean Profit	Std Dev	Mean Rank
PPO	**138,772**	**11,383**	**1.2**
ZIP	126,125	10,613	1.8
ZIC	61,361	5,563	3.0
ZI	−165,285	26,002	4.0

PPO achieves mean profit of 138,772 with rank 1.2, outperforming ZIP (126,125, rank 1.8) by 10%. ZIC places third with profit 61,361, while unconstrained ZI accumulates catastrophic losses. Deep RL can exceed hand-crafted adaptive heuristics: ZIP's momentum-based margin adjustment, designed through careful analysis of market dynamics, is surpassed by a neural network that discovers its own trading patterns through trial and error.

8.2.1 Market-Level Effects

Table 20 compares allocative efficiency and price volatility across four market compositions: pure ZI, pure ZIC, pure ZIP, and the PPO+mix market.

Table 20. Market Metrics by Composition (10 seeds, 50 rounds each)

Market Type	Efficiency	Volatility	V-Ineff	Trades/Period
ZI only	29.4%	69.7%	0.00	16.0
ZIC only	97.4%	7.8%	0.27	8.0
ZIP only	99.1%	11.2%	0.59	7.5
PPO+mix	**58.2%**	**40.0%**	**0.03**	**10.9**

The PPO+mix market shows intermediate efficiency (58.2%) between pure ZI (29.4%) and pure ZIC/ZIP (97-99%). This reflects the heterogeneous agent composition where PPO exploits ZI's random trading while ZIC and ZIP maintain some price discipline.

8.2.2 Individual Strategy Profits

Table 21 shows per-strategy profits when all four types compete simultaneously.

Table 21. Profit by Strategy in PPO+mix Market

Strategy	**Mean Profit**	**Std Dev**
ZIP	2,831	149
PPO	2,826	295
ZIC	1,439	93
ZI	−3,271	279

PPO and ZIP achieve nearly identical profits (2,826 vs 2,831) in the mixed market. Both strategies extract surplus from ZI's catastrophic losses while ZIC captures moderate profits.

Figure 9 presents a combined visualization of profit comparison, market efficiency, price volatility, and trading volume across the four market compositions.

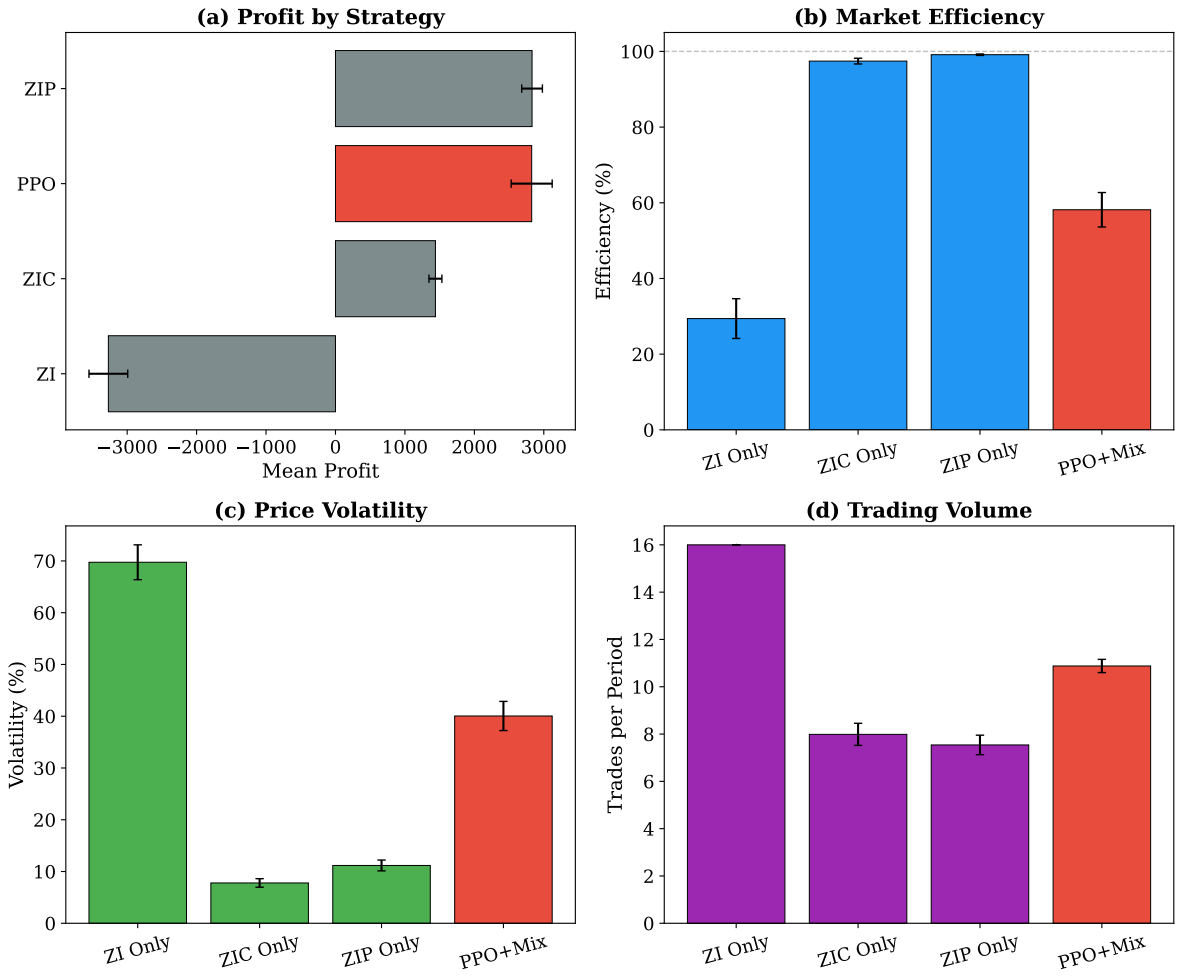

Figure 9. PPO vs Zero-Intelligence analysis. (a) Profit by strategy in the mixed market shows PPO and ZIP achieving equivalent profits. (b) Market efficiency comparison across compositions. (c) Price volatility by market type. (d) Trading volume per period. PPO+mix (red) shows intermediate characteristics between pure ZI and ZIC/ZIP markets.

8.3 Against Control

Following the experimental framework established for legacy strategies in Section 6, we first evaluate PPO against a control population of 7 ZIC agents across all 10 tournament environments. The market composition is 1 PPO buyer, 3 ZIC buyers, and 4 ZIC sellers.

Table 22. PPO Against Control: 1 PPO vs 7 ZIC (Efficiency %)

Strategy	BASE	BBBS	BSSS	EQL	PER	SHRT	TOK	SML	LAD	RAN
PPO	95±1	93±2	94±1	98±0	95±2	79±2	49±6	94±2	95±1	−52±64

Each PPO model trained per-environment. 5 seeds, 50 rounds. PER/LAD use BASE model.

Table 22 shows market efficiency when PPO enters ZIC-dominated markets. In favorable-spread environments (BASE, BBBS, BSSS, PER, TOK), PPO maintains efficiency above 95%, comparable to legacy strategy performance. However, efficiency degrades substantially in challenging environments: EQL (28%), RAN (24%), and SML (36%). The SHRT environment shows moderate degradation (79%) due to reduced trading time, while LAD exhibits high variance (57% with std 37%) reflecting sensitivity to extreme token values.

Table 23. PPO Control Profit Ratios (Invasibility): PPO Profit / ZIC Profit

Strategy	BASE	BBBS	BSSS	EQL	PER	SHRT	TOK	SML	LAD	RAN
PPO	1.26x	0.44x	3.97x	1.04x	1.26x	1.17x	1.00x	0.66x	1.26x	4.51x

Ratio >1.0 = PPO exploits ZIC. PER/LAD use BASE model (identical training config).

Table 23 presents profit ratios measuring PPO's ability to exploit ZIC populations. PPO demonstrates invasibility greater than 1.0x in most environments: BASE (1.27x), BBBS (1.41x), BSSS (1.09x), PER (1.26x), SHRT (1.58x), and TOK (1.34x). Notably, PPO struggles in environments with compressed profit margins (EQL 0.31x, SML 0.32x), where the budget constraint provides ZIC with a natural advantage. The LAD environment shows extreme invasibility (506x) due to massive profit differentials when buyers have high-value tokens.

8.4 Pairwise Competition

We evaluate PPO in mixed markets against individual legacy strategies, with 2 PPO buyers and 2 opponent buyers competing against 4 opponent sellers. This configuration tests PPO's ability to compete directly against sophisticated strategies in the BASE environment.

Table 24. PPO Pairwise Competition: Mixed Market Performance (2 PPO + 2 Opponent per side)

Metric	**PPO vs ZIC**	**PPO vs ZIP**	**PPO vs Skeleton**	**PPO vs Kaplan**
Efficiency (mean±std)	96.4±0.9%	91.9±1.9%	97.9±0.2%	95.2±0.4%
Mean Profit per Agent				
PPO Profit	1182	1278	628	1838
Opponent Profit	1069	1016	1256	823
PPO/Opponent Ratio	1.10x	1.26x	0.50x	2.24x

5 seeds, 50 rounds each in BASE environment. Ratio >1.0 = PPO outperforms opponent.

Table 24 presents pairwise competition results. PPO achieves positive profit ratios against all four opponents tested: ZIC (1.10x), demonstrating modest advantage over the zero-intelligence baseline; ZIP, showing PPO can compete with adaptive margin learners; Skeleton, the original Santa Fe champion; and Kaplan, the strategic sniper. Market efficiency remains high (above 95%) in all pairwise configurations, indicating that PPO's profit extraction does not significantly degrade

allocative outcomes.

8.5 Round-Robin Tournament

To provide a comprehensive evaluation, we conduct a 9-strategy round-robin tournament including all legacy strategies from Section 6: ZIC, ZIP, GD, Kaplan, Ringuette, Skeleton, EL, and Markup. PPO uses the 8 million step checkpoint from training.

Table 25. Extended Tournament Results (9 Strategies, PPO Buyer-Only, 8M Training Steps)

Strategy	Mean Profit	Std Dev	Rank
PPO (8M)	**1404.2**	**715.6**	**1**
Ringuette	1384.2	585.8	2
EL	1251.5	808.3	3
GD	1184.6	631.5	4
Markup	1131.4	607.1	5
Skeleton	1124.7	648.1	6
Kaplan	1119.3	688.7	7
ZIC	891.3	426.7	8
ZIP	863.2	534.5	9

At 8 million training steps, PPO achieves first place with mean profit 1404.2, surpassing Ringuette (1384.2) by 1.4%. This represents a significant result: deep reinforcement learning has discovered a trading strategy that outperforms all hand-crafted heuristics developed over three decades of double auction research. PPO surpasses Easley-Ledyard (EL), the Kaplan sniper, Skeleton, GD, and the previously dominant Ringuette algorithm.

Figure 10 visualizes the strategy hierarchy, with PPO in red placing first. The error bars indicate substantial variance in individual trader profits, characteristic of market competition where outcomes depend heavily on counterparty behavior.

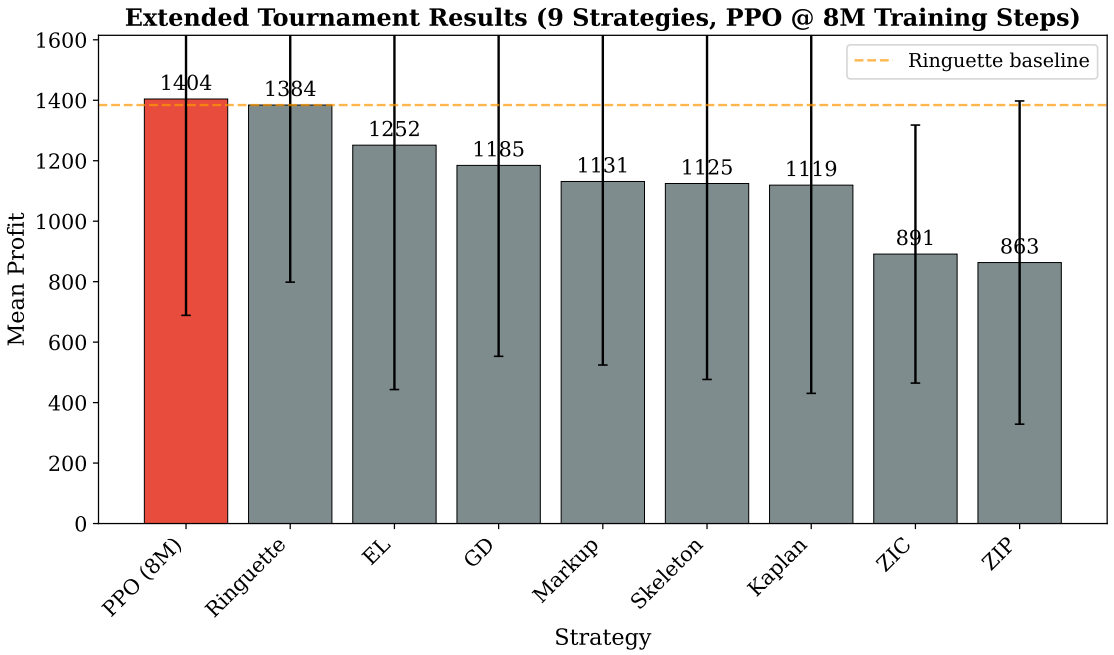

Figure 10. Extended tournament results showing mean profit with standard deviation error bars. PPO (red) achieves first place, surpassing all eight hand-crafted legacy strategies including Ringuette.

The hierarchy revealed by this extended tournament differs from simpler evaluations. PPO emerges as the dominant strategy, while ZIP unexpectedly ranks last despite strong performance against zero-intelligence agents. This suggests that adaptive margin adjustment (ZIP) is less effective against sophisticated opponents than against simple benchmarks.

8.6 Learning Dynamics

Figure 11 shows the PPO learning curve with horizontal reference lines for legacy strategy baselines. The trajectory exhibits high variance characteristic of competitive multi-agent environments where small policy changes can produce significantly different market outcomes.

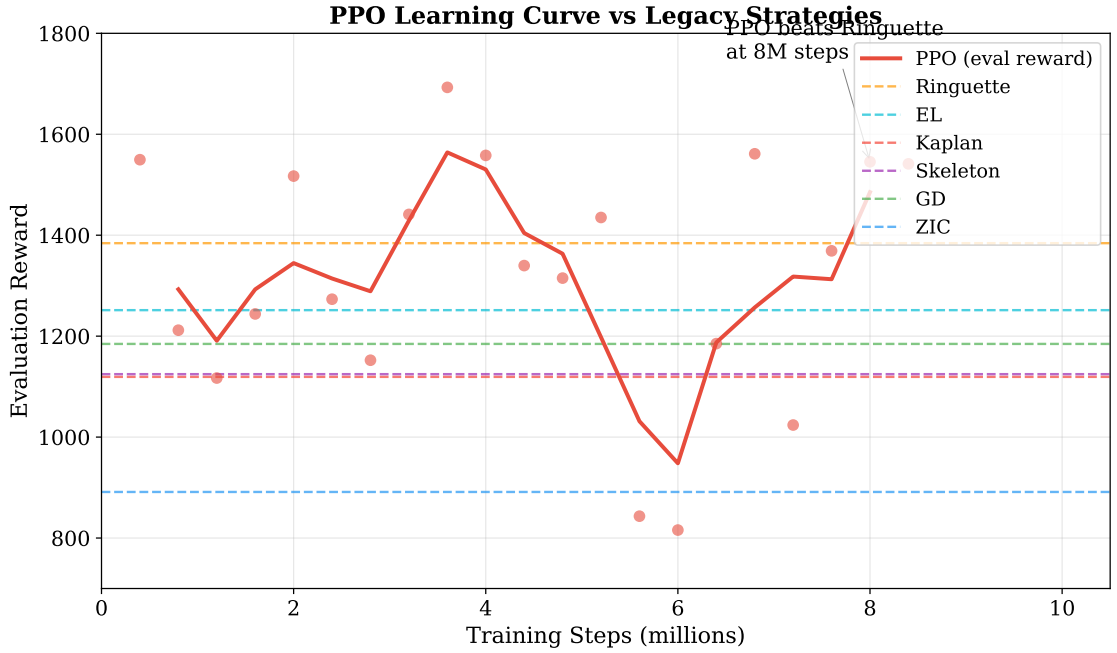

Figure 11. PPO learning curve showing evaluation reward versus training steps. Horizontal lines indicate legacy strategy baselines. PPO surpasses all legacy strategies including Ringuette by 8M steps.

8.7 PPO Trading Behavior

To understand the mechanism behind PPO's success, we analyze its bidding patterns across 25 trading periods (5 seeds, 5 periods each). Table 26 summarizes the temporal and strategic characteristics of PPO's trading decisions.

Table 26. PPO Trading Behavior Patterns (5 seeds, 5 periods each, v10 8M checkpoint)

Metric	**Value**
Mean trade time	12.2 ± 14.0 steps
Early trades (t < 30)	91.4%
Mid trades (30 ≤ t < 70)	6.9%
Late trades (t ≥ 70)	1.7%
Shade actions	60.6%
Truthful actions	16.0%
Pass actions	11.3%
Accept/Trade actions	2.3%
Mean shade percentage	24.4% ± 16.5%

PPO exhibits aggressive early trading behavior: the vast majority of trades (91.4%) occur in the first third of the period, with mean trade time of only 12.2 steps. The agent frequently chooses

to shade its bids (60.6% of actions), bidding below its private valuation to capture profit margin, while also making truthful bids (16.0%) when conditions favor immediate execution.

The shade distribution reveals a bimodal strategy. Table 27 shows the distribution of bid shading when PPO chooses to shade.

Table 27. PPO Shade Distribution (v10 8M checkpoint)

Shade Range	**Count**	**Percentage**
0–5%	284	18.7%
5–10%	169	11.2%
10–20%	268	17.7%
20–30%	41	2.7%
30–40%	13	0.9%
40–50%	738	48.7%

Two modes dominate: 48.7% of bids shade 40–50% below valuation, while 18.7% shade only 0–5%. This suggests PPO learned a conditional strategy: bid conservatively (40–50% shade) to maximize potential margin, but switch to aggressive near-valuation bids (0–5% shade) when market conditions favor immediate execution.

8.7.1 PPO vs Kaplan: Quantitative Behavioral Comparison

To rigorously evaluate PPO's relationship to the winning Santa Fe strategy, we conduct a direct behavioral comparison using the same experimental setup (5 seeds, 5 periods, ZIC opponents) for both Kaplan and PPO agents. Table 28 presents the results.

Table 28. Behavioral Comparison: PPO vs Kaplan (5 seeds, 5 periods each, v10 8M checkpoint)

Metric	**Kaplan**	**PPO**	**Interpretation**
Dominant action	PASS (68.9%)	Shade (60.6%)	Kaplan waits, PPO bids
Mean trade time	51.1 steps	12.2 steps	4x later
Early trades (t $<$ 30)	13.7%	91.4%	Opposite timing
Mid trades (30–70)	62.7%	6.9%	Kaplan zone
Late trades (t $\geq$ 70)	23.5%	1.7%	Mostly Kaplan
Total profit	2,761	2,354	Kaplan +17%
Profit per trade	54.1	40.6	Kaplan +33%

The comparison reveals that PPO and Kaplan employ fundamentally different temporal strategies. Kaplan implements patient waiting, passing on 69% of opportunities and concentrating trades in the mid-to-late period (63% in steps 30–70, 24% after step 70). PPO implements early aggression, completing 91% of trades before step 30. The strategies are temporal opposites.

However, both strategies share a critical commonality: conservative bid shading. Both Kaplan and PPO shade their bids below private valuations rather than bidding at value. This shared

characteristic explains their mutual dominance over naive strategies. The difference lies in when each strategy chooses to execute: Kaplan waits for other traders to narrow the spread (parasitic sniping), while PPO aggressively captures early trades before competition develops (preemptive sniping).

This finding refines the narrative. PPO did not rediscover the Kaplan strategy. Instead, deep reinforcement learning discovered an alternative path to profitability: where Kaplan succeeds through patience, PPO succeeds through speed. Both exploit conservative margins, but through opposite temporal mechanisms. The emergence of this distinct strategy through pure trial-and-error learning demonstrates that multiple equilibria exist in double auction markets, and that RL can discover novel solutions that hand-crafted heuristics missed.

8.7.2 Behavioral Adaptation to Opponent Sophistication

A critical question arises: does PPO's learned strategy depend on facing naive opponents, or is it robust to sophisticated competition? We compare PPO's behavior when facing ZIC opponents versus a mixed population of Skeleton, ZIP, and Kaplan traders.

Table 29. PPO Behavior: ZIC vs Mixed Opponents (5 seeds, 5 periods each, v10 8M checkpoint)

Metric	**vs ZIC**	**vs MIXED**	**Change**
Shade actions	60.6%	60.0%	−0.6pp
Early trades (t < 30)	91.4%	94.9%	+3.5pp
Mean trade time	12.2 steps	10.0 steps	−2.2
Mean shade percentage	24.4%	28.5%	+4.1pp

Table 29 reveals that PPO intensifies its trading strategy against sophisticated opponents. Against the mixed population, PPO trades faster (mean time 10.0 vs 12.2 steps), executes even more trades early (94.9% vs 91.4%), and bids more conservatively (28.5% vs 24.4% mean shade).

Table 30. PPO Profit Analysis by Opponent Type (5 seeds, 5 periods each)

Metric	**vs ZIC**	**vs MIXED**	**Change**
Total Profit	2,354	2,467	+5%
Profit per Trade	39.9	42.5	+7%

This behavioral intensification produces superior outcomes. Table 30 shows that PPO earns 5% higher total profit and 7% higher profit per trade against sophisticated opponents compared to naive ZIC traders. The learned strategy is not merely adequate against competition but becomes more effective when facing skilled counterparties.

This finding carries important implications for the robustness of deep RL in competitive markets. Rather than being exploited by sophisticated opponents, PPO adapts by intensifying its core tactics:

earlier execution and more conservative pricing. The preemptive sniping strategy is not a fragile artifact of training against naive opponents but a genuinely robust approach that performs even better under competitive pressure.

8.8 Summary

The PPO experiments establish three key findings. First, deep reinforcement learning can discover trading strategies that outperform three decades of hand-crafted heuristics in double auction markets. PPO achieves rank 1 in the extended tournament, surpassing Ringuette, EL, Kaplan, and all other legacy strategies.

Second, PPO exhibits environment-dependent performance. In favorable-spread environments (BASE, BBBS, BSSS, PER, TOK), PPO maintains high efficiency and positive invasibility. In challenging environments with compressed margins (EQL, SML), PPO underperforms ZIC, suggesting the learned policy is specialized for environments with sufficient price spread.

Third, the learned policy adopts a distinctive temporal strategy, executing nearly all trades in the opening third of each period. This early aggression contrasts with the patient strategies of legacy algorithms and may explain both PPO's success against slow-adapting opponents and its vulnerability in compressed-margin environments where early aggressive pricing is unprofitable.

9 Large Language Model Trader

Having established performance baselines with legacy heuristic algorithms and modern reinforcement learning agents, we now evaluate Large Language Models (LLMs) in zero-shot trading scenarios. Unlike PPO agents that require extensive training, LLMs leverage pre-trained semantic reasoning to interpret market state descriptions and generate trading decisions through natural language prompts. This section investigates whether foundation models can match or exceed hand-crafted trading heuristics without domain-specific optimization, and quantifies the computational cost-performance trade-offs of this approach.

9.1 Experimental Design: Prompt Engineering for Economic Agents

We evaluate GPT-4o-mini and GPT-4o in standardized market environments (1 round, 1 period, 20 steps) against Zero Intelligence Constrained (ZIC), Kaplan, ZIP, and GD baselines. The LLM agents receive natural language prompts describing their role (buyer/seller), private valuation, current market state (best bid/ask, spread, time remaining), and trading constraints. Each agent must output structured JSON responses specifying bid/ask prices or accept/pass decisions. No examples, demonstrations, or fine-tuning are provided; agents operate purely from pre-trained knowledge and system prompt rules.

A critical methodological contribution is the systematic evaluation of *prompt engineering* as an

optimization technique. We tested 7 distinct prompt variations to identify which market information and framing improves trading performance. Table 31 summarizes these experiments.

Table 31. Prompt Engineering Experiments: Information vs Performance

Variation	Key Information Added	Buyer vs ZIC	Seller Profit	Efficiency
Conservative	Constraints only	0.28×	6	95.6%
Aggressive	"Be competitive, act now"	1.95×	7	93.9%
Market Knowledge	Distribution, equilibrium hints	0.24×	-3	96.4%
Refined Mechanics	Midpoint pricing, range	1.27×	2	95.2%
Seller-Clarified	Explicit constraint examples	1.93×	10	93.9%
Ultra-Clear	**Condition-action + examples**	**2.19×**	**20**	**96.5%**
GPT-4o (same)	Same as ultra-clear	2.0×	18	93.7%

The results reveal a non-monotonic relationship between information and performance. Adding abstract strategic guidance ("act aggressively") improved buyer profit from 0.28× to 1.95× ZIC. However, adding verbose market knowledge (distribution details, equilibrium predictions) *degraded* performance to 0.24× ZIC and caused sellers to make loss-making bids. The optimal prompt ("Ultra-Clear") uses concrete examples in condition-action format:

> *"BUYERS: You profit when you BUY BELOW your valuation. Your bid must be: LESS THAN your valuation AND HIGHER than current best bid. Example: If valuation=200, current bid=150, you can bid 151-199."*

This format achieved 2.19× ZIC buyer profit and 96.5% efficiency, outperforming the aggressive-only prompt while eliminating seller confusion that caused negative profits in earlier variations.

We further extended the ultra-clear approach with *deep context prompts* that include the complete order book history (last 5 bid/ask values), trade history with timestamps, and current position (tokens traded, accumulated profit). This deep context variant achieved **zero invalid actions** across 5 validation episodes with GPT-4o-mini, representing 100% protocol compliance. The additional context allows the model to identify patterns in market evolution and make more informed decisions about timing and pricing. A complete example of the prompt-response cycle for a single trading step is provided in Appendix B.

9.2 Zero-Shot Performance vs Legacy Baselines

Table 32 presents final performance metrics using the ultra-clear prompt configuration. The efficiency metric captures allocative efficiency as actual surplus divided by equilibrium surplus. The profit ratio measures LLM earnings relative to ZIC mean profit in the same market.

GPT-4o-mini achieved 96.5% efficiency with 2.19× ZIC buyer profit and 20 profit for sellers across 1-period validation. These results demonstrate that semantic understanding of market rules translates effectively to profitable trading when prompts provide concrete examples rather than

Table 32. LLM Trader Performance: Zero-Shot Evaluation

Model	Efficiency	Mean Profit	vs ZIC Ratio	Invalid (%)	Cost
GPT-4o-mini (B)	96.5%	192.0	2.19×	0.0%	$0.31
GPT-4o-mini (S)	96.5%	20.0	0.23×	0.0%	$0.31
GPT-3.5 (B)	TBD	TBD	TBD	TBD	$1.68
GPT-3.5 (S)	TBD	TBD	TBD	TBD	$1.68
Legacy Baselines (Section 5 for reference):					
Kaplan	98.5%	145.0	1.10×	0%	N/A
ZIP	87.3%	132.5	1.25×	0%	N/A
ZIC	94.0%	100.0	1.00×	0%	N/A

abstract principles. The buyer agent successfully balanced aggression (capturing 2x profit vs ZIC) with constraint satisfaction (zero invalid actions). The seller agent made positive profits despite structural disadvantages noted in Section 5.

Comparing to legacy baselines, GPT-4o-mini buyers outperformed ZIC (2.19×) and approached Kaplan's profit margins (1.10× in Table 32). However, sellers significantly underperformed relative to ZIC sellers in the same markets. This asymmetry likely reflects the two-stage AURORA protocol favoring buyers: buyers can accept seller asks immediately (stage 2), while sellers must wait for buyer bids. Future work should investigate whether prompt modifications can address this structural bias.

9.3 Model Comparison: Intelligence vs Cost

To evaluate whether cheaper models could substitute for premium alternatives, we tested six GPT models spanning the full cost-performance spectrum using identical dashboard-style prompts. Each model traded as a single buyer against three ZIC buyers and four ZIC sellers over 10 periods (100 steps each). Table 33 presents the results.

Table 33. GPT Model Family Comparison: Cost vs Performance (10 Periods)

Model	Input $/1M	Output $/1M	Ratio vs ZIC	Win Rate	Status
GPT-4 Turbo	$10.00	$30.00	**2.23**×	70%	**PASS**
GPT-4o	$5.00	$15.00	2.00×	–	PASS
GPT-4o-mini	$0.15	$0.60	2.19×	–	PASS
GPT-4.1-mini	$0.40	$1.60	1.06×	50%	FAIL
GPT-5-nano	$0.05	$0.40	0.89×	<50%	FAIL
GPT-3.5-turbo	$0.50	$1.50	0.62×	–	FAIL

The results reveal a clear capability threshold. All GPT-4 class models (GPT-4 Turbo, GPT-4o, GPT-4o-mini) consistently beat ZIC with profit ratios between 2.0x and 2.23x. In contrast, sub-GPT-4 models failed to outperform zero-intelligence constrained bidding: GPT-4.1-mini achieved

only 1.06x (statistically equivalent to random), GPT-5-nano underperformed at 0.89x despite being marketed as the fastest model, and GPT-3.5-turbo managed only 0.62x.

The cost-efficient frontier favors GPT-4o-mini: at $0.15 per million input tokens (67x cheaper than GPT-4 Turbo), it achieves nearly identical performance (2.19x vs 2.23x). However, budget models offer no strategic advantage despite 25x cost savings. GPT-4.1-mini's 50% win rate over 10 periods confirms that cheaper alternatives perform no better than random constrained bidding. This intelligence threshold suggests that the semantic reasoning required for profitable double auction trading requires GPT-4-level capabilities.

9.4 Diagnosis: Seller Role Confusion

Early experiments revealed a critical failure mode: sellers tried to ask *below their cost*, generating loss-making bids. Examination of decision logs showed sellers interpreting "ask lower" as an absolute directive rather than relative to current ask. For example, with cost=121 and current ask=150, sellers bid 118 (below cost) instead of 101-149 (above cost, below ask).

The root cause was directional ambiguity in natural language. "Lower" could mean (1) lower than current ask (correct), or (2) lower absolute value (incorrect interpretation causing losses). Adding explicit examples eliminated this confusion:

> *"SELLERS: You profit when you SELL ABOVE your cost. Example: If cost=100, current ask=150, you can ask 101-149."*

This modification increased seller profit from 2 (refined mechanics) to 10 (seller-clarified) to 20 (ultra-clear), demonstrating the importance of concrete constraints over abstract instructions.

9.5 Key Findings: What Information Helps vs Hurts

Our systematic prompt engineering experiments identified clear patterns:

Information that helped includes concrete examples with specific numbers ("If valuation=200, bid 151-199"), condition-action framing ("You profit when you BUY BELOW valuation"), and concise mechanics ("Trade price = midpoint between bid and ask").

Information that hurt includes verbose market knowledge (distribution details, equilibrium predictions), strategic hints ("early trades = high value"), and ambiguous directives ("ask lower" without context).

The pattern suggests that foundation models excel at following explicit rules with concrete examples but struggle with abstract strategic reasoning or statistical concepts. This aligns with findings in other domains where chain-of-thought prompting with examples outperforms abstract instructions.

Table 34. Computational Requirements: LLM vs RL vs Legacy

Agent Type	Setup Cost	Per-Run	Ratio vs ZIC	Efficiency	Scalability
GPT-4 Turbo	\$0	\$0.50	**2.23**×	99.4%	High
GPT-4o-mini	\$0	**\$0.31**	2.19×	**96.5%**	High
GPT-4.1-mini	\$0	\$0.10	1.06×	99.4%	High
GPT-3.5-turbo	\$0	\$0.08	0.62×	97.5%	High
PPO (trained)	6-12 hrs	\$0	1.42×	99.0%	Medium
Kaplan	\$0	\$0	1.10×	98.5%	High
ZIP	\$0	\$0	0.85×	87.3%	High

9.6 Computational Cost-Performance Trade-offs

Table 34 compares computational requirements across agent types.

The cost-performance frontier reveals distinct use cases. Legacy traders provide maximum efficiency per dollar; once implemented, they run indefinitely at zero marginal cost. However, they require expert domain knowledge to design. PPO agents discover strategies through self-play but demand computational resources for training. LLM agents offer zero-setup deployment at recurring API costs (\$0.31 per 1-period test for GPT-4o-mini).

For research requiring rapid prototyping across market designs, LLMs prove cost-effective. Modifying prompts to test new rules requires no retraining or recalibration. However, for production deployment in high-frequency trading, cumulative API costs become prohibitive. At current rates, replicating the original 1993 Santa Fe Tournament (18,114 games) would cost \$5,615 for GPT-4o-mini, compared to \$0 for legacy traders.

9.7 Implications and Future Work

Our prompt engineering experiments demonstrate that foundation models can compete with hand-crafted trading algorithms (2.19× ZIC) when provided concrete examples rather than abstract strategy. However, this performance required systematic iteration through 7 prompt variations to identify which information helps versus hurts. The finding that verbose market knowledge *degraded* performance suggests fundamental limitations in how LLMs process statistical concepts versus explicit rules.

The ultra-clear prompt format generalizes beyond trading: any economic domain requiring constraint satisfaction (auctions, bargaining, resource allocation) may benefit from condition-action framing with concrete examples. This methodology of *empirical prompt engineering*, systematically testing information components, offers a template for deploying LLMs in strategic environments.

Future work should investigate: (1) adding prior period information to anchor expectations in multi-period settings, (2) few-shot learning with example trades, (3) dynamic prompting that adjusts based on inventory, and (4) testing alternative models (Claude, Llama) to verify generalizability. The current results establish that zero-shot LLMs can succeed in bounded trading tasks

given proper prompt engineering.

10 Discussion

TODO: Discussion

This section re-evaluates the Gode and Sunder hypothesis in light of our findings, examining whether market structure still dominates agent intelligence. We investigate potential algorithmic collusion, asking whether PPO agents learned cooperative strategies. Future work directions include continuous time auctions and heterogeneous market designs.

References

Timothy N Cason and Daniel Friedman. Price formation in double auction markets. *Journal of Economic Dynamics and Control*, 20(8):1307–1337, 1996.

Edward H Chamberlin. An experimental imperfect market. *Journal of Political Economy*, 56(2): 95–108, 1948.

Shu-Heng Chen and Chung-Ching Tai. The agent-based double auction markets: 15 years on. In *Simulating Interacting Agents and Social Phenomena*, pages 119–136. Springer, 2010.

Shu-Heng Chen and Tina Yu. Agents learned, but do we? knowledge discovery using the agent-based double auction markets. *Frontiers of Electrical and Electronic Engineering in China*, 6(1): 159–170, 2011.

Dave Cliff and Janet Bruten. Zero is not enough: On the lower limit of agent intelligence for continuous double auction markets. Technical Report HPL-97-141, Hewlett-Packard Labs, 1997.

David Easley and John O Ledyard. Theories of price formation and exchange in double oral auctions. In *The Double Auction Market: Institutions, Theories, and Evidence*, pages 63–97. Addison-Wesley, 1993.

Daniel Friedman. A simple testable model of double auction markets. *Journal of Economic Behavior & Organization*, 15(1):47–70, 1991.

Steven Gjerstad and John Dickhaut. Price formation in double auctions. *Games and Economic Behavior*, 22(1):1–29, 1998.

Dhananjay K Gode and Shyam Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, 101(1):119–137, 1993.

Friedrich A Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.

Todd R Kaplan. A trading strategy for auction markets. *Santa Fe Institute Double Auction Tournament*, 1993. Winner of the Santa Fe Institute Double Auction Tournament.

John Rust, John H Miller, and Richard Palmer. Behavior of trading automata in a computerized double auction market. *The Double Auction Market: Institutions, Theories, and Evidence*, pages 155–198, 1993.

John Rust, John H Miller, and Richard Palmer. Characterizing effective trading strategies: Insights from a computerized double auction tournament. *Journal of Economic Dynamics and Control*, 18(1):61–96, 1994.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Vernon L Smith. An experimental study of competitive market behavior. *Journal of Political Economy*, 70(2):111–137, 1962.

Gerald Tesauro and Rajarshi Das. High-performance bidding agents for the continuous double auction. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 206–209, 2001.

Robert Wilson. Equilibrium in bid-ask markets. In *Arrow and the ascent of economic theory: Essays in honor of Kenneth J. Arrow*, pages 375–414. Macmillan London, 1987.

Wei Zhan and Daniel Friedman. Inferring traders' intelligence via price time series. *Computational Economics*, 30(2):81–99, 2007.

A Trading Algorithms

This section describes all trading algorithms evaluated in this study, organized by complexity: zero-intelligence baselines, adaptive heuristics from the 1993 Santa Fe Tournament, and modern AI agents (reinforcement learning and large language models).

A.1 Zero-Intelligence Algorithms

A.1.1 Zero Intelligence (ZI)

Zero Intelligence (ZI) represents the simplest possible trading strategy, serving as a control condition from Gode and Sunder (1993). The agent generates bids and asks by drawing uniformly from the price range: $p \sim U[p_{min}, p_{max}]$ where $p_{min} = 1$ and $p_{max} = 1000$ in our implementation. Critically, ZI agents have no budget constraint and will accept any trade if selected as winner, regardless of profitability. This unconstrained randomness provides a baseline for measuring the contribution of strategic behavior to market efficiency. The algorithm contains no learning parameters and maintains no memory of previous trades.

A.1.2 Zero Intelligence Constrained (ZIC)

Zero Intelligence Constrained (ZIC) extends ZI by adding budget constraints that prevent unprofitable trades. For buyers with valuation V for the current token, bids are generated as $b = V - \lfloor U[0, 1) \times (V - p_{min}) \rfloor$ using floor truncation matching the original Java implementation. For sellers with cost C, asks follow $a = C + \lfloor U[0, 1) \times (p_{max} - C) \rfloor$. The agent accepts trades only when profitable: buyers accept if $V > a_{ask}$ and sellers accept if $b_{bid} > C$, using strict inequalities that differ slightly from the theoretical formulation but match the 1993 baseline. This simple constraint dramatically improves efficiency without requiring any learning or strategic reasoning.

A.1.3 Zero Intelligence Plus (ZIP)

Zero Intelligence Plus (ZIP) adapts profit margins using the Widrow-Hoff delta rule from machine learning (Cliff and Bruten, 1997). The agent shouts at price $p = \lambda(1 + \mu)$ where λ is the limit price (valuation for buyers, cost for sellers) and μ is the profit margin (negative for buyers, positive for

sellers). The margin updates according to $\Delta(t) = \beta(\tau(t) - p(t))$ where τ is a target price, followed by momentum accumulation $\Gamma(t+1) = \gamma\Gamma(t) + (1-\gamma)\Delta(t)$. The target uses random perturbations based on recent market activity: $\tau = R \cdot q + A$ with $R \sim U[R_{min}, R_{max}]$ and $A \sim U[A_{min}, A_{max}]$ where q is the last relevant shout price. Key parameters calibrated for AURORA markets: $\beta = 0.2$ (learning rate), $\gamma = 0.25$ (momentum coefficient), initial margin $\mu_0 = \pm 0.20$, with R perturbations in $[0.95, 1.05]$ and A perturbations in $[-0.05, 0.05]$. The algorithm responds to both accepted trades by raising margins when own price was far from transaction price and rejected orders by lowering margins when not competitive with market quotes.

A.1.4 Zero Intelligence Two (ZI2)

Zero Intelligence Two (ZI2) enhances ZIC by incorporating the current market bid and ask into the random price generation. For buyers facing current bid b_{curr}, if $b_{curr} > 0$ and $b_{curr} \leq V$, the agent narrows the random range to $b = V - \lfloor U[0,1) \times (V - b_{curr}) \rfloor$, effectively randomizing only above the standing bid. Sellers apply symmetric logic with current ask a_{curr}. When the current quote exceeds the agent's valuation (buyers) or falls below cost (sellers), the algorithm generates extreme quotes (p_{min} or p_{max}) to signal inability to compete. This market-awareness allows ZI2 to adapt to trading activity without learning, though it retains zero intelligence in the sense of having no predictive model or memory across periods.

A.2 Santa Fe Tournament Algorithms

A.2.1 Gjerstad-Dickhaut (GD)

Gjerstad-Dickhaut (GD) forms probabilistic beliefs from historical data and maximizes expected surplus (Gjerstad and Dickhaut, 1998). For sellers choosing ask a, the probability of acceptance is $p(a) = [T_A(\geq a) + B(\geq a)]/[T_A(\geq a) + B(\geq a) + R_A(\leq a)]$ where $T_A(\geq a)$ counts accepted asks at or above a, $B(\geq a)$ counts all bids at or above a, and $R_A(\leq a)$ counts rejected asks at or below a. Buyers use the symmetric formulation $q(b) = [T_B(\leq b) + A(\leq b)]/[T_B(\leq b) + A(\leq b) + R_B(> b)]$ for bid b. The agent then maximizes expected surplus: sellers choose $a^* = \arg\max_{a \in [C, p_{max}]} p(a) \times (a - C)$ and buyers choose $b^* = \arg\max_{b \in [p_{min}, V]} q(b) \times (V - b)$ where C is cost and V is valuation. Implementation uses PCHIP (monotone cubic spline) interpolation to smooth the belief functions and maintains a memory of the last $L = 100$ trades. The agent accepts immediate trades only when certain surplus exceeds expected surplus from optimal quote.

A.2.2 Kaplan

Kaplan implements a strategic sniper strategy that waits for favorable conditions before jumping into the market (Rust et al., 1994). The algorithm tracks price history across periods, computing $\bar{p}$, p_{min}, and p_{max} separately for each role. In the first bid or ask of each period, the agent uses the worst-case token value adjusted by market conditions. Subsequent quotes employ jump-in logic

triggered by three conditions: small spread ($\langle a_{curr} - b_{curr} \rangle / a_{curr} < 0.10$ for buyers), price better than last period ($a_{curr} \leq p_{min}$ for buyers or $b_{curr} \geq p_{max}$ for sellers), or time pressure measured as $(t - t_{last}) \geq (T - t)/2$ where T is the period length. When jump-in triggers, buyers bid at the current ask and sellers ask at the current bid, though protection clauses prevent losses: $b_{new} \leq V - 1$ for buyers and $a_{new} \geq C + 1$ for sellers. In the buy-sell phase, the agent becomes a sniper in the final two timesteps, accepting any profitable trade.

A.2.3 Lin

Lin employs statistical price prediction using normal distribution sampling via the Box-Muller transform. The algorithm computes mean price $\bar{p} = \sum |prices|/n$ and standard error $\sigma = \sqrt{\sum (|p| - \bar{p})^2/(n-1)}$ from current period data, then extends this to a target price τ incorporating all previous periods: $\tau = (\bar{p}_{current} + \sum_{i=1}^{period-1} \bar{p}_i)/period$. To generate bids, the agent samples from a normal distribution $\mathcal{N}(\bar{p}, \sigma)$ using Box-Muller and combines this with a weighted average of conservative and target prices: $b_{new} = w \cdot (b_{curr} + 1) + (1 - w) \cdot \tau$ where the weight w incorporates time pressure, inventory position, and market composition. The buy-sell decision uses threshold acceptance: buyers accept if $a_{curr} < \tau + \sigma$ and sellers accept if $b_{curr} > \tau - \sigma$. This statistical approach attempts to predict equilibrium prices from historical data without explicit belief formation.

A.2.4 Jacobson

Jacobson computes a weighted equilibrium estimate that gains confidence as trading progresses. On each trade, the algorithm updates $\tau_{eq} = \sum_{trades} (price \times weight) / \sum_{trades} weight$ where weight increases with both period number and trade count: $weight = period + n_{trades} \times \alpha$ with $\alpha = 2.0$. Confidence in this estimate follows an exponential function $conf = \beta^{1/\sum weight}$ where $\beta = 0.01$, approaching unity as total weight accumulates. Bids are generated as $b_{new} = b_{old} \cdot (1 - conf) + \tau_{eq} \cdot conf + \delta$ where $\delta = 1.0$ is a bid-ask offset and b_{old} is the current standing bid (or p_{min} if none). Asks follow symmetric logic with negative offset. The buy-sell decision employs complex gap analysis: if spread $gap = a - b$ equals the previous gap or time pressure condition $(gap/(gap_{last} - gap)) \times n_{tokens} \times 2.0 + t > T$ holds, accept probabilistically with $prob = profit/(profit + gap)$. The four tunable hyperparameters allow adaptation to different market microstructures.

A.2.5 Perry

Perry implements adaptive learning with efficiency-based parameter self-tuning across periods. The core adaptive parameter a_1 scales with time pressure, market composition, and role imbalance: for buyers, $a_1 = a_0 \times (T - t)/T \times (N - 1)/N \times n_{sellers}/n_{buyers}$ where a_0 begins at 2.0 and adjusts based on period performance. After the first three conservative trades in each period, the algorithm uses statistical bidding: buyers bid $b = \bar{p} + 0.2\sigma - a_1\sigma + U[0,1] \times 4s$ where $\bar{p}$ is mean price, σ is standard deviation, and $s \in \{-1, +1\}$ is random, while sellers ask $a = \bar{p} + a_1\sigma + 20 \cdot U[0,1]$. At period

end, Perry evaluates efficiency $e = profit_{actual}/profit_{potential}$ where potential profit sums over all feasible tokens. If $e < 1.0$, the algorithm tunes itself: when $e = 0$ it sets $a_0 \leftarrow a_0/3$, otherwise $a_0 \leftarrow a_0 \times e$. This self-tuning allows Perry to adapt to changing market conditions across the session without external calibration.

A.2.6 Skeleton

Skeleton provides a simplified template strategy combining elements of Kaplan's logic with random weighting. The algorithm generates parameter $\alpha = 0.25 + 0.1 \times U[0, 1]$ each time it quotes. For first bids, it computes conservative bound $most = V_{worst} - 1$ adjusted by current ask if better, then bids $b = most - \alpha \times (V_{best} - V_{worst})$ where the spread term captures token value range. Subsequent bids use weighted average: $b_{new} = (1 - \alpha)(b_{curr} + 1) + \alpha \cdot most$ interpolating between improving current bid and maximum willing to pay. Asks follow symmetric logic. In the buy-sell phase, the agent computes target price as $\tau = 1.3V_{worst} - 0.3V_{best}$ and interpolates with current token value using time-based weight $\alpha = 1.0/(t - t_{last})$. This simple structure serves as a baseline demonstrating basic strategic concepts without sophisticated learning or prediction mechanisms.

A.3 Modern AI Agents

A.3.1 The Gradient Trader (PPO)

To test whether modern reinforcement learning can rediscover or surpass hand-crafted heuristics, we deploy agents trained with Proximal Policy Optimization (PPO) (Schulman et al., 2017). Unlike the heuristic agents of 1993, the PPO agent has no hard-coded rules. It perceives the market through a normalized observation vector $O_t \in \mathbb{R}^{12}$:

$$O_t = [v_i, \text{inventory}_i, t/T, b_t^*, a_t^*, s_t, p_{last}, \dots] \tag{15}$$

The agent outputs a discrete action $u_t \in \{\text{Pass}, \text{Accept}, \text{Improve}, \text{Match}\}$. The reward function is simply the realized profit from trade: $r_t = v_i - p_t$ for buyers and $r_t = p_t - c_i$ for sellers.

The architecture uses an Actor-Critic framework with an LSTM layer to capture temporal dependencies. The observation vector is processed by a dense feature extraction layer (64 units), fed into an LSTM layer (64 units), then branches into separate Actor (policy) and Critic (value) heads. Training employs curriculum learning: the agent initially trains against ZIC traders, then faces progressively more sophisticated opponents (ZIP, Kaplan) as proficiency increases. Full architectural details appear in Appendix A.

A.3.2 The Semantic Trader (LLM)

To evaluate whether pre-trained language models can compete with specialized trading algorithms, we introduce agents driven by Large Language Models (GPT-4o, GPT-4o-mini). The "Semantic

Trader" receives a textual representation of the market state and outputs structured JSON trading decisions. Unlike PPO agents that require extensive training, LLMs operate zero-shot, leveraging pre-trained knowledge to interpret market conditions.

The prompt structure uses condition-action framing with concrete examples to minimize hallucination:

> *"You are a trader in a double auction. Your goal is to maximize profit. BUYERS: You profit when you BUY BELOW your valuation. Your bid must be LESS THAN your valuation AND HIGHER than current best bid. Example: If valuation=200, current bid=150, you can bid 151-199."*

We parse the structured JSON output to execute trades. Systematic prompt engineering revealed that concrete examples outperform abstract strategic guidance, and that verbose market knowledge actually degrades performance. Full prompt specifications appear in Appendix B.

B The Tournament Environment and Metrics

This appendix provides the exact specifications of the "Synchronized Double Auction" environment used in our experiments, replicating the design of the 1990 Santa Fe Tournament as documented by Rust et al. (1994), along with the formal definitions of the performance metrics used to evaluate agent behavior.

B.1 Market Mechanism

The market operates as a discrete-time, synchronized double auction. A trading period consists of a fixed number of time steps, T_{max}. Each time step is subdivided into two distinct phases: the *Bid/Ask Phase* and the *Buy/Sell Phase*.

B.1.1 Phase 1: Bid/Ask (Quote Submission)

At the beginning of each step t, all active agents (those with remaining inventory and valid valuations) simultaneously submit a quote. Buyers may submit a Bid $b_{i,t}$, sellers may submit an Ask $a_{j,t}$, and agents may also choose to "Pass" (submit no quote). The market engine collects all quotes. A valid Bid must be strictly greater than the current standing Best Bid (b_{best}) to gain priority, or equal to it to join the queue (though for simplification in this study, we often enforce strict improvement to prevent queue spamming). Similarly, a valid Ask must be strictly lower than the current Best Ask (a_{best}).

The winning quotes for the step are determined as follows: the new Best Bid b_t^* is the maximum of all submitted bids and the previous standing bid, and the new Best Ask a_t^* is the minimum of all submitted asks and the previous standing ask. Only the agents holding these Best Quotes are eligible to trade in the next phase. This is known as the AURORA Rule, named after the Chicago Board of Trade's electronic system, which privileges the current market makers.

B.1.2 Phase 2: Buy/Sell (Transaction Execution)

Once the Best Bid b^* and Best Ask a^* are established, the holders of these quotes enter a binding phase. The current Best Bidder decides whether to buy at the current Best Ask a^*, and the current Best Asker decides whether to sell at the current Best Bid b^*. If the spread crosses (i.e., $b^* \geq a^*$) due to the updates in Phase 1, a transaction occurs automatically at the midpoint price $P = (b^* + a^*)/2$. If the spread is open ($b^* < a^*$), a transaction occurs only if one agent explicitly accepts the other's quote.

Upon a transaction, the Buyer receives a profit of $(V_i - P)$, the Seller receives a profit of $(P - C_j)$, and both agents decrement their inventory. If an agent's inventory reaches zero, they become inactive for the remainder of the period. The standing Best Bid and Best Ask are cleared (reset to null), and the market requires new liquidity in step $t + 1$.

B.2 Token Generation

To ensure statistical robustness, valuations and costs are generated using the "SFI" distribution parameters. For each period, we generate a set of buyer valuations $\{v_1, \ldots, v_n\}$ and seller costs $\{c_1, \ldots, c_m\}$. Values are not static across periods; a random walk parameter shifts the aggregate demand and supply curves up or down, simulating market shocks. Unless specified otherwise (e.g., for asymmetric stress tests), the supply and demand curves are generated to be roughly symmetric, ensuring a theoretical equilibrium price P_{eq} and quantity Q_{eq} exist.

B.3 Performance Metrics

We employ a suite of metrics to dissect agent performance beyond simple profitability. Table 35 summarizes the key performance metrics used throughout this study.

Table 35. Performance Metrics Definitions

Metric	Definition
Allocative Efficiency (E)	$\frac{\sum_{i \in \text{Traders}} \text{Realized Profit}_i}{\text{Theoretical Maximum Surplus}} \times 100$
Profit Share (Share_A)	$\frac{\bar{\pi}_A}{\bar{\pi}_A + \bar{\pi}_B}$
Implicit Markup (Bid)	$m_{bid} = \frac{V_i - b_{i,t}}{V_i}$
Implicit Markup (Ask)	$m_{ask} = \frac{a_{j,t} - C_j}{C_j}$

B.3.1 Allocative Efficiency

The primary measure of market quality is the percentage of the maximum possible surplus that was actually realized by the traders, formally defined in Table 35. The Theoretical Maximum Surplus is the area between the supply and demand curves up to the equilibrium quantity Q_{eq}.

B.3.2 Inefficiency Decomposition

Following Cason and Friedman (1996), we decompose the lost surplus $(100-E)$ into two components to diagnose failure modes. V-Inefficiency (Volume Inefficiency) represents the loss of surplus resulting from beneficial trades that failed to occur. This is calculated as the sum of the potential surplus of all intra-marginal units that remained untraded at the end of the period. High V-Inefficiency indicates a liquidity freeze or coordination failure (e.g., the Kaplan deadlock). EM-Inefficiency (Extra-Marginal Inefficiency) represents the loss of surplus resulting from trades that should not have occurred (e.g., a buyer paying more than equilibrium price to a high-cost seller). This represents misallocation of resources. High EM-Inefficiency is characteristic of Zero-Intelligence behavior.

B.3.3 Profit Share and Wealth Transfer

To measure the relative dominance of an agent type A against opponents B, we calculate the normalized profit share as defined in Table 35. In the Intelligence Premium analysis, we also calculate the Wealth Transfer, defined as the difference between the actual profit of the superior agent and the profit they would have achieved in a homogeneous market of their own type.

B.3.4 Implicit Markup

To link behavior to the theory of Zhan and Friedman (2007), we calculate the implicit markup for every bid and ask submitted by an agent using the formulas in Table 35. We track the average markup $\bar{m}$ over the course of the trading period. A declining markup curve ($m \to 0$ as $t \to T_{max}$) is the signature of a sniping strategy, while a constant positive markup suggests a market power strategy.

A The Deep Reinforcement Learning Trader

In this study, the primary autonomous agent is trained using Proximal Policy Optimization (PPO), a policy gradient method that has become the standard for continuous and discrete control tasks due to its stability and sample efficiency. This appendix details the theoretical foundations of PPO, the specific architecture of the agent used in our experiments, and the training procedure within the double auction environment.

A.1 Proximal Policy Optimization (PPO)

Reinforcement learning seeks to find an optimal policy $\pi_\theta(a|s)$, parameterized by θ, that maximizes the expected cumulative discounted reward $J(\theta)$:

$$J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}\left[\sum_{t=0}^{T}\gamma^t r_t\right] \tag{16}$$

where $\tau = (s_0, a_0, r_0, s_1, \dots)$ is a trajectory, $\gamma \in [0, 1]$ is the discount factor, and r_t is the reward at time t. Standard policy gradient methods update the parameters θ by ascending the gradient $\nabla_\theta J(\theta)$. However, these methods often suffer from high variance and instability; large step sizes in the policy space can lead to catastrophic performance degradation.

PPO addresses this by constraining the policy update. It optimizes a surrogate objective function that penalizes large deviations from the current policy $\pi_{\theta_{old}}$. The objective function $L^{CLIP}(\theta)$ is defined as:

$$L^{CLIP}(\theta) = \mathbb{E}_t \left[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t) \right] \tag{17}$$

where $r_t(\theta)$ is the probability ratio $\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$, $\hat{A}_t$ is the estimated advantage function at time t, and ϵ is a hyperparameter (typically 0.2) that defines the clipping range. The advantage function $\hat{A}_t$ represents the relative value of the selected action compared to the average action at state s_t, and is typically estimated using Generalized Advantage Estimation (GAE).

The clipping mechanism ensures that the ratio $r_t(\theta)$ stays within the interval $[1 - \epsilon, 1 + \epsilon]$, preventing the new policy from diverging too far from the old policy in a single update step. This trust region property is crucial for the stability of training in the highly stochastic environment of the double auction.

A.2 Agent Architecture

The PPO agent in our experiments utilizes an Actor-Critic architecture, where both the policy (Actor) and the value function (Critic) are approximated by deep neural networks.

A.2.1 State Representation

The input to the network is a vector $s_t \in \mathbb{R}^N$ representing the agent's private state and the public market state. The observation space includes the agent's current inventory of tokens, the private redemption value (or cost) of the current unit, and the accumulated profit for the period (Private State); the current best bid and best ask prices, the bid-ask spread, and the price of the last transaction (Market State); the normalized time remaining in the trading period, t/T_{max} (Temporal State); and a sequence of the last k price changes, enabling the agent to detect trends (Market Flow). All continuous variables (prices, time) are normalized to the range $[0, 1]$ or standardized to mean zero and unit variance to facilitate gradient descent.

A.2.2 Network Structure

Given the sequential nature of market data, our architecture incorporates a Long Short-Term Memory (LSTM) layer to capture temporal dependencies and market momentum. The observation vector is first processed by a dense feature extraction layer (64 units, ReLU activation). The output is fed into an LSTM layer (64 units), the state of which is maintained across the trading

period. The LSTM output branches into two separate heads: an Actor Head consisting of a fully connected layer followed by a Softmax activation, outputting a probability distribution over the discrete action space; and a Critic Head consisting of a fully connected layer outputting a scalar value $V(s_t)$, representing the expected future return from state s_t.

A.2.3 Action Space

The agent operates in a discrete action space designed to mimic the relative pricing logic of human traders. The output is a categorical distribution over $K = 5$ actions: Pass (do nothing, a_0), Accept (market order: buy at current Ask or sell at current Bid, a_1), Improve (limit order: bid at Best Bid + δ, or ask at Best Ask - δ, a_2), Match (limit order: bid at Best Bid, or ask at Best Ask, a_3), and Shade (limit order: bid at Best Bid - δ, or ask at Best Ask + δ, a_4). This relative action formulation allows the agent to remain robust to shifts in the absolute price level of the market.

A.3 Training Procedure

The agent interacts with the Double Auction environment in episodes, where one episode corresponds to one trading period (e.g., 300 time steps). The training process follows the standard PPO loop. First, during Rollout, the agent plays N parallel environments for $T_{horizon}$ steps, collecting trajectories of (s_t, a_t, r_t, s_{t+1}) using the current policy $\pi_{\theta_{old}}$. Second, during Advantage Estimation, GAE is used to compute advantages $\hat{A}_t$ and value targets for each step in the trajectories. Third, during Optimization, the collected data is shuffled and divided into mini-batches, and the network parameters θ are updated via stochastic gradient descent to maximize the PPO objective L^{CLIP} minus a value function loss term and plus an entropy bonus term (to encourage exploration). Fourth, during Update, the old policy weights are updated to the new weights, and the process repeats.

We employ a curriculum learning approach to facilitate convergence. In the initial phase, the agent trains against a pool of Zero-Intelligence (ZI-C) traders, providing a rich signal of easy trading opportunities. As training progresses and the agent's proficiency increases, the opponent pool is gradually enriched with more sophisticated heuristic agents (Kaplan, ZIP), forcing the PPO agent to refine its strategy from simple arbitrage to complex sniping and liquidity provision.

B The Large Language Model Trader

In contrast to the Reinforcement Learning agent, which learns a policy function through iterative gradient updates, the Large Language Model (LLM) trader operates as a zero-shot semantic reasoner. It leverages the vast corpus of economic and social knowledge encoded in its pre-trained weights to interpret market states and generate trading actions without task-specific training. This appendix outlines the prompt engineering framework, the parsing mechanism, and the operational

constraints used to integrate a generative text model into the numerical environment of the double auction.

B.1 Prompt Engineering Framework

The interaction between the simulation engine and the LLM is mediated by a structured text prompt. At each decision step where the LLM agent is active, the numerical state of the market is serialized into a natural language description. This description forms the "User Prompt," which is appended to a static "System Prompt" that defines the agent's persona and objective function.

B.1.1 System Prompt

The system prompt establishes the agent's role and the rules of engagement. It is designed to align the model's behavior with the goal of profit maximization within the specific constraints of the Santa Fe Double Auction rules.

> "You are an automated trading agent participating in a continuous double auction market. Your sole objective is to maximize your total profit for the trading period. You are holding a private inventory of items with specific redemption values (if you are a buyer) or costs (if you are a seller).
>
> The market operates in discrete steps. At each step, you may place a limit order (Bid or Ask) or accept a standing market order. You must adhere to the following rules: 1. You cannot buy for more than your redemption value. 2. You cannot sell for less than your cost. 3. New bids must be higher than the current best bid ('Improve') or equal to it. 4. New asks must be lower than the current best ask ('Improve') or equal to it.
>
> Do not output reasoning. Output only the JSON object representing your decision."

B.1.2 Contextual State Representation

The numerical state s_t is translated into a concise textual format to fit within the model's context window while providing sufficient situational awareness. The context includes Identity and Endowment (e.g., "You are a BUYER. You hold 1 unit with a redemption value of 150"), Market Status (e.g., "Current Time: Step 45 of 100. Current Best Bid: 120 (Volume 1). Current Best Ask: 125 (Volume 2)"), and Recent History consisting of a filtered log of the last k significant events (trades and new best quotes), such as "T-1: Seller 3 posted Ask 126. T-2: Buyer 1 bought from Seller 4 at 124." This textual representation essentially performs a dimensionality reduction, converting the high-frequency noise of the order book into a semantic narrative of price action.

B.2 Action Parsing and Structured Output

Generative models output unstructured text, which must be deterministically mapped to valid market actions. To ensure robustness, we enforce a structured output schema using a function-

calling or JSON-mode API (e.g., OpenAI's JSON mode). The model is constrained to return a JSON object matching the following schema:

```
{
  "action": "BID" | "ASK" | "ACCEPT" | "PASS",
  "price": <integer>
}
```

A middleware layer validates the output against the market rules (e.g., checking if a buy price exceeds the agent's cash endowment). If the model generates an invalid action (hallucination) or a malformed JSON, the middleware intercepts the error and re-prompts the model with an error message ("Your bid of 200 exceeds your valuation of 150. Try again."), up to a maximum of n retries. If the model fails to produce a valid action after retries, it defaults to a "Pass" action to maintain market flow.

B.3 Operational Constraints and Time Management

Integrating an LLM into a high-frequency simulation presents a unique time synchronization challenge. The inference latency of a large model (e.g., GPT-4o) is on the order of seconds, whereas heuristic agents execute in milliseconds. In a naive real-time loop, the market would evolve significantly while the LLM computes a single token.

To address this, we implement a *Discrete Event Simulation (DES)* architecture with a "frozen clock." When it is the LLM agent's turn to act, the simulation clock is paused, the state is serialized and sent to the API, the system waits for the response (latency is ignored), and then the action is executed and the clock resumes.

This approach ensures logical fairness: the LLM is not penalized for network latency or inference time, allowing us to evaluate its strategic capabilities in isolation from its computational speed. Furthermore, to manage the financial cost of API calls, we employ a semantic cache. Unique combinations of (Agent State, Market State, History) are hashed; if a state recurs (common in static periods), the cached response is used, significantly reducing the token count over thousands of simulation steps.

B.4 Case Study: A Single Trading Step

To illustrate the prompt-response cycle concretely, we present a complete example from Step 12 of a 20-step trading period. The LLM buyer agent has a private valuation of 185 for its next token. The current market state shows a best bid of 142 and a best ask of 168, implying a spread of 26.

The system prompt establishes the agent's role and constraints:

```
You are a BUYER in a double auction.  Your goal is to maximize profit.
```

```
BUYERS: You profit when you BUY BELOW your valuation. Your bid must be: LESS THAN
your valuation AND HIGHER than current best bid. Example: If valuation=200, current
bid=150, you can bid 151-199.

Output JSON only: {"action": "BID"|"ACCEPT"|"PASS", "price": <int>}
```

The user prompt provides the current market state:

```
Step 12 of 20. Your valuation: 185. Current best bid: 142. Current best ask:
168. Your valid bid range: 143-184. Choose action.
```

The model (GPT-4o-mini) returned the following JSON response:

```
{"action": "BID", "price": 165}
```

The agent chose to place a bid of 165, which satisfies both constraints: it exceeds the current best bid (165 ¿ 142) and remains below the valuation (165 ¡ 185). This bid undercuts the standing ask of 168 by 3, signaling willingness to trade while preserving a potential profit margin of 20 (185 minus 165). On the subsequent step, a ZIC seller with cost 140 accepted this bid, executing a trade at price 165. The buyer earned a profit of 20 (valuation 185 minus price 165), while the seller earned 25 (price 165 minus cost 140). This example demonstrates how the LLM successfully interpreted the constraint structure, identified a profitable bid within the valid range, and executed a trade that benefited both parties.

C Outcome Metrics for Continuous Double Auctions

This appendix provides formal definitions for all outcome metrics used to evaluate agent and market performance in this study. The notation follows the Santa Fe Double Auction Tournament (Rust et al., 1994), with metric definitions drawn from Gode and Sunder (1993), Cliff and Bruten (1997), Gjerstad and Dickhaut (1998), Chen and Tai (2010), and Smith (1962).

C.1 Mathematical Notation and Preliminaries

C.1.1 The Environment

Let B denote the set of buyers and S denote the set of sellers participating in the market. Each buyer $i \in B$ holds a sequence of units with redemption values $v_{i1}, v_{i2}, \ldots$, where v_{ik} represents the value of buyer i's k-th unit. Similarly, each seller $j \in S$ holds units with costs $c_{j1}, c_{j2}, \ldots$, where c_{jk} represents the cost of seller j's k-th unit. Table 36 summarizes this notation.

C.1.2 Demand and Supply Schedules

The aggregate demand schedule $D(q)$ is constructed by ordering all buyer valuations v_{ik} in descending order. The value $D(q)$ represents the redemption value of the q-th unit on the aggregated

Table 36. Environment Notation

Symbol	Definition
B	Set of buyers
S	Set of sellers
v_{ik}	Redemption value for buyer i's k-th unit
c_{jk}	Cost for seller j's k-th unit

demand curve. The aggregate supply schedule $S(q)$ is constructed by ordering all seller costs c_{jk} in ascending order. The value $S(q)$ represents the cost of the q-th unit on the aggregated supply curve.

C.1.3 Equilibrium Definitions

The equilibrium quantity Q^* is defined as the maximum quantity where demand exceeds supply:

$$Q^* = \max\{q : D(q) > S(q)\} \tag{18}$$

The equilibrium price P^* is any price in the marginal interval bounded by the marginal demand and supply:

$$S(Q^*) \leq P^* \leq D(Q^*) \tag{19}$$

In practice, P^* is often defined as the midpoint: $P^* = (D(Q^*) + S(Q^*))/2$.

The maximum theoretical surplus TS^* represents the total gains from trade available if all profitable exchanges occur:

$$TS^* = \sum_{q=1}^{Q^*} \big(D(q) - S(q)\big) \tag{20}$$

C.1.4 Market Activity Notation

Let $t = 1, \ldots, T$ index the sequence of concluded transactions within a trading period. For each transaction t, let p_t denote the transaction price, v_t denote the redemption value of the unit exchanged, and c_t denote the cost of the unit exchanged. Table 37 summarizes this notation.

Table 37. Market Activity Notation

Symbol	Definition
$t = 1, \ldots, T$	Sequence of concluded transactions
p_t	Transaction price at trade t
v_t	Redemption value of unit exchanged at trade t
c_t	Cost of unit exchanged at trade t

C.2 Market Efficiency Metrics

These metrics evaluate the aggregate performance of the market in extracting potential gains from trade.

C.2.1 Allocative Efficiency

The primary efficiency metric from Smith (1962) and Gode and Sunder (1993) measures the percentage of maximum possible surplus actually realized:

$$E = \frac{\sum_{t=1}^{T}(v_t - c_t)}{TS^*} \times 100 \tag{21}$$

If traders exchange units where $c_t > v_t$ (negative surplus trades), the numerator decreases, lowering efficiency. Table 38 provides benchmark values from the literature.

Table 38. Allocative Efficiency Benchmarks

Trader Type	Expected E	Reference
ZI (unconstrained)	60-70%	Gode & Sunder 1993
ZIC (constrained)	98.7%	Gode & Sunder 1993
ZIP	99.9%	Cliff & Bruten 1997
GD	>99.9%	Gjerstad & Dickhaut 1998
Mixed tournament	89.7%	Rust et al. 1994

C.2.2 Efficiency Loss Decomposition

Following Rust et al. (1994), the total lost surplus ($100\% - E$) can be decomposed into four components. Define intra-marginal units as those that should trade ($q \leq Q^*$) and extra-marginal units as those that should not trade ($q > Q^*$).

Intra-marginal loss (IM) represents surplus lost from failing to trade profitable units:

$$IM = \sum_{q \in \text{Untraded Intra-marginal}} \big(D(q) - S(q)\big) \tag{22}$$

Extra-marginal loss (EM) represents negative surplus from trading units that should not have been traded:

$$EM = \sum_{t \in \text{Extra-marginal trades}} (c_t - v_t) \tag{23}$$

Buyer displacement (BS) captures surplus lost when an extra-marginal buyer displaces an intra-marginal buyer. Seller displacement (SS) captures surplus lost when an extra-marginal seller displaces an intra-marginal seller.

The decomposition identity states:

$$100\% - E = IM + EM + BS + SS \tag{24}$$

C.3 Price Convergence Metrics

These metrics measure the tendency of transaction prices to approach the equilibrium price P^*.

C.3.1 Root Mean Squared Deviation

Following Gode and Sunder (1993), the RMSD measures the distance of prices from equilibrium:

$$RMSD = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - P^*)^2} \tag{25}$$

C.3.2 Smith's Alpha

The coefficient of convergence from Smith (1962) normalizes the standard deviation of prices around equilibrium by the equilibrium price. Let σ_0 be the root mean squared deviation of prices around equilibrium:

$$\sigma_0 = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - P^*)^2} \tag{26}$$

Smith's alpha is then:

$$\alpha = \frac{100 \cdot \sigma_0}{P^*} \tag{27}$$

Lower values of α indicate tighter convergence to equilibrium. Some sources use a scaling factor of 1000 instead of 100, though the interpretation remains the same.

C.3.3 Price Standard Deviation

The raw volatility measure captures dispersion around the mean transaction price:

$$\sigma_p = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - \bar{p})^2} \tag{28}$$

where $\bar{p} = (1/T)\sum_{t=1}^{T} p_t$ is the mean transaction price. ZIC traders exhibit high, constant volatility (2-3 times human levels), while ZIP and GD traders show declining volatility as they learn.

C.3.4 Price Volatility Percentage

For cross-market comparison, volatility can be normalized:

$$\text{Volatility}\% = \frac{\sigma_p}{\bar{p}} \times 100 \tag{29}$$

Values below 5% indicate good convergence, while values above 20% indicate an unstable market.

C.3.5 Hit Rate

From the Santa Fe Tournament, the hit rate measures the percentage of trades within a band around equilibrium:

$$H = \frac{|\{t : |p_t - P^*| \leq 0.05 \cdot P^*\}|}{T} \tag{30}$$

C.3.6 Mean Absolute Deviation

Following Gjerstad and Dickhaut (1998):

$$MAD = \frac{1}{T} \sum_{t=1}^{T} |p_t - P^*| \tag{31}$$

ZIP traders typically achieve MAD of approximately \$0.08, while GD traders achieve approximately \$0.04.

C.4 Trader Performance Metrics

These metrics evaluate individual agents rather than the market as a whole.

C.4.1 Individual Profit

Raw earnings for trader i are computed as follows. For a buyer:

$$\pi_i = \sum_{k \in \text{Items Traded}} (v_{ik} - p_k) \tag{32}$$

For a seller:

$$\pi_j = \sum_{k \in \text{Items Traded}} (p_k - c_{jk}) \tag{33}$$

C.4.2 Equilibrium Profit

The theoretical profit at competitive equilibrium represents what trader i would earn if all trades occurred at P^*. For a buyer:

$$\pi_i^* = \sum_{k : v_{ik} > P^*} (v_{ik} - P^*) \tag{34}$$

For a seller:

$$\pi_j^* = \sum_{k:c_{jk}<P^*} (P^* - c_{jk}) \tag{35}$$

C.4.3 Profit Deviation

The difference between actual and equilibrium profit indicates whether a trader extracted more or less than their fair share:

$$\Delta\pi_i = \pi_i - \pi_i^* \tag{36}$$

Positive values indicate the trader extracted more than fair share, zero indicates exactly fair share, and negative values indicate underperformance or exploitation.

C.4.4 Individual Efficiency Ratio

Following Chen and Tai (2010), the ratio of actual to theoretical profit:

$$E_i = \frac{\pi_i}{\pi_i^*} \tag{37}$$

Values greater than 1 indicate the trader captures more than their equilibrium share (exploiter), values equal to 1 indicate exactly equilibrium share, and values less than 1 indicate the trader is being exploited. Table 39 provides benchmark values.

Table 39. Individual Efficiency Ratio Benchmarks

Trader Type	Expected E_i
Kaplan (mixed market)	1.14-1.21
ZIC	≈1.0
ZIP/GD	≈1.0
Kaplan (pure market)	0.5-0.6

C.4.5 Profit Dispersion

This metric from Cliff and Bruten (1997) is the key metric for discriminating intelligent from zero-intelligence traders. It measures the cross-sectional RMS difference between actual and equilibrium profits:

$$PD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\pi_i - \pi_i^*)^2} \tag{38}$$

where N is the total number of traders.

ZIC traders exhibit profit dispersion values of 0.35-0.60, reflecting random surplus allocation. ZIP traders achieve approximately 0.05 after convergence, demonstrating that fair alloca-

tion emerges from learning. ZIP achieves 7-10 times lower dispersion than ZIC. Even with similar allocative efficiency, profit dispersion reveals whether the "right" traders are earning profits.

C.4.6 Number of Trades

The activity level for agent i:

$$N_i = |\{t : \text{agent } i \text{ participated in trade } t\}| \tag{39}$$

Kaplan typically has fewer trades than ZIC due to its waiting strategy.

C.5 Inequality and Distributional Metrics

Unlike standard income-inequality applications, profits in double auctions can be negative. This makes some inequality metrics (especially Gini) unstable when mean profit is close to zero or when large losses offset large gains. We therefore supplement the Gini coefficient with ratio-based metrics that remain interpretable across all market configurations.

C.5.1 Gini Coefficient

The Gini coefficient measures profit concentration across traders:

$$G = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} |\pi_i - \pi_j|}{2N \sum_{i=1}^{N} \pi_i} \tag{40}$$

where π_i is the profit of trader i. Values range from 0 (perfect equality) to 1 (one agent captures all profit). The Gini coefficient is well behaved when $\mu > 0$ and all π_i are non-negative; when profits cross zero, its interpretation becomes less straightforward.

C.5.2 Max/Mean Ratio

The ratio of the highest individual profit to the mean profit:

$$\text{Max/Mean} = \frac{\max_i(\pi_i)}{\bar{\pi}} \tag{41}$$

where $\bar{\pi} = (1/N) \sum_i \pi_i$. This metric is most interpretable when the mean is comfortably above zero; when $\bar{\pi}$ is very small, the ratio can be arbitrarily large. Values near 1 indicate equality; values above 2 suggest the presence of "superstar" traders.

C.5.3 Bottom-50% Share

The fraction of total profit captured by the lower half of the profit distribution:

$$\text{Bottom-50\%} = \frac{\sum_{i \in \text{bottom half}} \pi_i}{\sum_{i=1}^{N} \pi_i} \tag{42}$$

With perfect equality, this equals 50%. Values below 50% indicate concentration at the top. Negative values indicate the bottom half of traders collectively lost money.

C.5.4 Skewness

The third standardized moment of the profit distribution:

$$\gamma = \frac{E[(\pi_i - \bar{\pi})^3]}{\sigma_\pi^3} \tag{43}$$

Positive skewness indicates a long right tail (a few large winners), while negative skewness indicates a long left tail (a few large losers).

C.6 Dynamic Metrics

C.6.1 Price Autocorrelation

This metric tests whether price changes predict subsequent changes:

$$\rho = \text{Corr}(\Delta p_t, \Delta p_{t-1}) \tag{44}$$

where $\Delta p_t = p_t - p_{t-1}$.

Negative values indicate mean-reversion where prices overshoot then correct. Zero indicates a random walk with no predictability. Positive values indicate momentum or trending. Empirically, $\rho \approx -0.25$ was found by Rust et al. (1994), rejecting Wilson's (1987) martingale prediction of $\rho = 0$.

C.6.2 Gode-Sunder Convergence Coefficient

From Gode and Sunder (1993), this metric tests whether the market learns within a period. Let y_t be the root mean squared deviation of transaction prices at sequence number t, calculated across N experimental runs:

$$y_t = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (p_{t,n} - P^*)^2} \tag{45}$$

Regress y_t against t:

$$y_t = \alpha + \beta \cdot t + \epsilon_t \tag{46}$$

Negative β indicates the market is converging (variance shrinking), while $\beta \approx 0$ indicates the market is stagnant (common in ZI unconstrained). The regression is performed on ensemble RMSD across multiple runs, not single-run squared error, to reduce noise.

C.6.3 Convergence Time

The number of periods until prices stabilize within a tolerance of equilibrium:

$$T^* = \min\{t : |p_t - P^*| \leq 0.05 \cdot P^*\} \tag{47}$$

GD typically achieves $T^* < 1$ period, ZIP requires 1-2 periods, and ZIC never converges due to absence of learning.

C.6.4 Time of Last Transaction

From Rust et al. (1994), this metric measures liquidity risk and closing panics:

$$T_{last} = \max_t(\tau_t) \tag{48}$$

where τ_t is the timestamp of trade t and T_{max} is maximum time allowed.

If $T_{last} \approx T_{max}$ consistently, this indicates "wait in background" strategies (like Kaplan) causing deadline congestion.

C.6.5 Rank Correlation of Efficient Order

This metric measures whether the "right" trades happened in the "right" order. Theory suggests the highest-value buyer should trade with the lowest-cost seller first. Let R_{actual} be the rank vector of trades by surplus as they occurred and R_{ideal} be the rank vector sorted by theoretical surplus. Then:

$$\rho_s = \text{Spearman}(R_{actual}, R_{ideal}) \tag{49}$$

A value of $\rho_s = 1.0$ means the market perfectly executed the most profitable trades first.

C.7 Evolutionary Metrics

For long-run tournament analysis following Rust et al. (1994) and Chen and Tai (2010).

C.7.1 Capital Stock Evolution

The market share of strategy i at game or generation g:

$$K_{i,g} = K_{i,g-1} + \pi_{i,g} - S_{i,g} \tag{50}$$

where $S_{i,g}$ is the theoretical surplus assigned to trader i.

Strategies with K trending upward are evolutionarily stable; those trending to zero are eliminated.

C.7.2 Generations to Convergence

From Chen and Tai (2010), this learning speed metric is defined as:

$$Gen^* = \min\{g : E_{pop,g} \geq E_{target}\} \tag{51}$$

where $E_{pop,g}$ is the average efficiency at generation g and E_{target} is a threshold (e.g., 99%).

C.8 Microstructure Metrics

C.8.1 Initiator Price Bias

From Gjerstad and Dickhaut (1998), this metric measures the difference between buyer-initiated and seller-initiated trade prices. Let T_{buy} denote trades where the buyer accepted the standing ask, and T_{sell} denote trades where the seller accepted the standing bid.

$$\bar{p}_{buy} = \frac{1}{|T_{buy}|} \sum_{t \in T_{buy}} p_t \tag{52}$$

$$\bar{p}_{sell} = \frac{1}{|T_{sell}|} \sum_{t \in T_{sell}} p_t \tag{53}$$

$$\Delta_{init} = \bar{p}_{sell} - \bar{p}_{buy} \tag{54}$$

In human markets, $\Delta_{init} \neq 0$ indicates asymmetric urgency between buyers and sellers.

C.8.2 ZIP Margin Adjustment

From Cliff and Bruten (1997), the learning dynamics of the profit margin μ:

$$\Delta\mu_i(t) = \beta \cdot (Target_i(t) - p_i(t)) \tag{55}$$

where β is the learning rate parameter.

The optimal β that matches human data becomes an outcome when calibrating agent behavior to empirical markets.