

MASTERARBEIT | MASTER'S THESIS

Titel | Title

An Experimental Test of Learning Rules: Evidence from
Benchmark and Game-Embedded Tasks

verfasst von | submitted by

Xiaopeng Dong

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 953

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Research in Economics and Finance

Betreut von | Supervisor:

Univ.-Prof. Dipl.-Math. Karl Schlag PhD

AI Disclosure Statement

This thesis was written entirely by the author. Artificial intelligence tools (specifically, large language models) were used in the following limited and auxiliary capacities only:

1. **Code review:** Reviewing and debugging code (oTree, Python).
2. **L^AT_EX formatting:** Assisting with document formatting, table layout, and figure placement.
3. **Grammar correction:** Proofreading English text for grammatical accuracy.
4. **Literature search:** Searching and identifying relevant academic references.

All AI-assisted outputs were verified and validated by the author prior to inclusion.

Abstract

Diese Masterarbeit untersucht drei Eigenschaften von Lernregeln (Börgers et al., 2004): Unverzerrtheit, Expediency und Monotonizität. Sie fragt, ob diese Eigenschaften in einem Drei-Arm-Bandit-Experiment empirisch nachgewiesen werden können und ob ein gamifiziertes Framing das Verhalten im Vergleich zu einem Standard-Framing verändert. Die Ergebnisse zeigen, dass Unverzerrtheit im Wesentlichen erfüllt ist, während Expediency und Monotonizität unter keinem der beiden Framings zuverlässig nachgewiesen werden können. Gamifizierung verbessert die Leistung nicht robust, erhöht jedoch die Reaktionszeit und die Interaktion. Zudem zeigt sie das Potenzial, die erwarteten Anreizkosten zu senken, ohne die zugrunde liegenden Auszahlungsverteilungen zu verändern.

Schlagwörter: Experimentelle Wirtschaftsforschung; Mehrarmige Banditen; Lernregeln; Gamifizierung

This thesis studies three properties of learning rules (Börgers et al., 2004): unbiasedness, expediency, and monotonicity. It asks whether they can be detected in a three-armed bandit experiment, and whether a gamified framing changes behavior relative to a standard framing. Unbiasedness broadly holds, while expediency and monotonicity cannot be reliably detected under either framing. Gamification does not robustly improve performance, but it increases response time and interaction. It also shows potential to reduce expected incentive costs without changing the underlying payoff distributions.

Keywords: Experimentation; Multi-armed bandits; Learning rules; Gamification

1 Introduction

The multi-armed bandit problem is a canonical model of decision-making under uncertainty, which captures the exploration-exploitation trade-off. The metaphor comes from a casino, where people face a row of different slot machines but do not know the payoff distributions, and they have to trade off between exploiting machines that have paid well and exploring untried ones. The same structure arises in many real-world settings, such as online recommendation systems, clinical trials, and resource allocation.

Existing work often studies bandit problems under Bayesian updating, where decision makers hold priors about payoff distributions and update beliefs after observing outcomes. In contrast, this thesis focuses on non-Bayesian learning rules, which do not require the decision maker to hold or update a prior over payoff distributions. Some of these rules update choice probabilities directly from realized payoffs; others adjust behavior based on simple sample statistics such as observed frequencies or mean payoffs. Both classes substantially reduce the cognitive demands of the learning task and reflect bounded rationality. Börgers et al. (2004) identify three properties that a learning rule worth studying should satisfy. Unbiasedness means that, when all arms have the same expected payoff, the learning rule preserves choice probabilities rather than changing the choice probability of any particular arm. Expediency means that the decision maker's expected payoff weakly increases from one period to the next. Monotonicity means that, when some arms have higher expected payoffs than others, the learning rule increases in expectation the probability of choosing the payoff-maximizing arm or arms.

In this thesis, we explore the decision-making patterns and performance in multi-armed bandit problems and ask two questions: the main question is whether human participants exhibit learning dynamics consistent with unbiasedness, expediency, and monotonicity. A secondary methodological question is whether gamified framing affects learning behavior, engagement, and experimental cost. This question is motivated by well-known problems in online experiments: inattentiveness, fatigue, and weak engagement, which gamified framing may help mitigate.

We design an experiment to analyze participants' behavioral patterns under bounded rationality. The experiment uses a within-subject design. Each participant completes two 30-round three-armed bandit tasks. The two tasks differ in two respects. First, in payoff environment: an unequal-payoff environment and an equal-payoff environment. Second, in framing: one task uses a standard experimental format, while the other embeds the task in a mini-game with optional weapon upgrades. The framing is counterbalanced against task position across participant groups, so that the framing effect can be identified within each environment. In addition to choices and realized payoffs, the experiment records response times, survey responses, and elicited beliefs about payoff probabilities.

The empirical analysis combines aggregate learning property tests with participant-level comparison of behavioral patterns under different framings. For the property tests, we examine whether choice dynamics satisfy the three theoretical benchmarks in the appropriate payoff environments: unbiasedness in the equal-payoff environment, and expediency and monotonicity in the unequal-payoff environment. For the comparison exercise, we compare standard and gamified framings using response-time measures, performance, and expected cost.

To complement the experimental analysis, we run simulations in which agents follow the

Cross rule (Cross, 1973), a non-Bayesian learning rule shown to satisfy all three properties by Börgers et al. (2004). We first ask whether, under an experimentally realistic number of rounds, the three properties can be reliably detected even in a population composed entirely of Cross-rule agents. When detection fails, we extend the simulation as a counterfactual exercise, varying the number of rounds far beyond the experimental horizon to test whether each property becomes statistically detectable as the learning dynamics approach their asymptotic behavior.

We find that the three properties differ sharply in how difficult they are to detect. In the equal-payoff environment, the standard framing is consistent with temporal invariance of arm-choice frequencies. The gamified framing remains consistent with temporal invariance but differs on the auxiliary uniform-level diagnostic, indicating a stable change in choice levels rather than a clear change in trajectory. In the unequal-payoff environment, evidence for expediency and monotonicity remains weak: trends are negative and insignificant under the standard framing and only marginally positive under the gamified framing, and the between-framing comparisons detect no significant difference on either property. The simulations show that this weakness is largely due to limited statistical power: even a population composed entirely of Cross-rule agents which satisfy the properties yields detection in only about half of replications at the 30-round horizon and about two thirds at 1,000 rounds. As a result, this weak evidence should not be read as evidence against the theory. Separately, because participants spend in-game currency on optional weapon upgrades during the gamified task, the resulting reduction in the lottery pool lowers the experiment’s expected incentive cost by approximately 15 percent.

This thesis contributes to two main strands of literature. The first is the experimental literature on multi-armed bandit problems. Early work by Banks et al. (1997) provided a laboratory test of the bandit problem but was limited to two arms and did not classify individual strategies. A parallel non-Bayesian tradition, building on Cross (1973), instead models learning as direct adjustments to choice probabilities based on realized payoffs, which is the tradition our thesis extends. Most recently, Hudja and Woods (2025) classify behavioral strategies in two- and three-armed bandits with finite and indefinite horizons, finding that most participants are best described by either win-stay-lose-shift or reinforcement learning. Our thesis extends this literature in three directions. First, we provide the first experimental test of the three theoretical properties of learning rules proposed by Börgers et al. (2004). We complement the experiment with simulations that explore the conditions under which these properties are statistically detectable. Second, we elicit participants’ beliefs to examine whether stated beliefs are consistent with observed choices, revealing substantial belief-action gaps. Third, we classify participants into learning types via maximum likelihood estimation over a family of non-Bayesian rules, extending the behavioral classification approach of Hudja and Woods (2025) to a setting that combines benchmark and game-embedded tasks.

The second strand is the methodology of online experiments and experimental design. On-line experiments have become a valuable tool for behavioral research, offering access to larger and more diverse participant pools at lower cost (Horton et al., 2011). However, they introduce methodological challenges that are largely absent in laboratory settings: participants may read instructions inattentively, become distracted, or drop out before completing the task (Paolacci and Chandler, 2014). Arechar et al. (2018) document this problem in a repeated public goods experiment, where only about half of the experimental groups retained all members throughout

the session. Gamification has been proposed as a potential solution. A systematic review by Lumsden et al. (2016) finds that incorporating gamified elements into cognitive tasks can improve participant engagement and data quality without compromising scientific validity. Yet the evidence is nuanced: Lumsden et al. (2017) show that superficial game elements such as points and cosmetic themes alone do not significantly reduce attrition. These findings suggest that how one gamifies matters more than whether one gamifies. While gamification has been studied in psychology and the serious-games literature, to our knowledge, no economic experiment has embedded a fully functional game into its task framing. Our mini-game treatment contributes to this literature by providing direct evidence on the effects of embedding a full game on three outcomes: behavioral performance, data quality, and expected incentive cost.

2 Experimental Design

2.1 Multi-Armed Bandit Problem

In this thesis, we implement a three-armed bandit task in an online experiment. In each task, participants choose among three arms over a fixed number of rounds. Participants are not informed in advance about either the prize amount or the winning probability of any arm. In each round, the selected arm yields either a fixed number of tickets or zero. The participant’s objective is to maximize total earnings within a given number of rounds.

2.2 Treatments and Parameters

The experiment uses a within-subject design in which each participant completes two tasks of 30 rounds each, for a total of 60 rounds. Two features vary across the two tasks: the payoff environment and the framing. The payoff environment changes in a fixed sequence: every participant faces the equal-payoff environment first and the unequal-payoff environment second. The framing is counterbalanced: the standard-first group receives the standard framing in the first task and the gamified framing in the second, while the game-first group receives the reverse. Counterbalancing the framing against task position ensures that, within each environment, half of the participants face each framing, so that each environment–framing combination is represented by balanced, comparable samples.

The two tasks differ first in their payoff environments. In both, each of the three arms is a two-outcome lottery that returns a fixed prize with a given probability and zero otherwise. The notation $p \times [x]$ denotes that a payoff of x tickets is realized with probability p , so that an arm’s expected value is simply $p \cdot x$.

In the equal-payoff environment (E1), the three arms are

$$A : 0.8 \times [10] + 0.2 \times [0], \quad B : 0.5 \times [16] + 0.5 \times [0], \quad C : 0.4 \times [20] + 0.6 \times [0],$$

which by construction share an identical expected value of $EV = 8$ but differ in how that value is decomposed into prize size and win probability: Arm A pays a small prize frequently, Arm C a large prize rarely, and Arm B an intermediate prize with even odds.

In the unequal-payoff environment (E2), the arms are

$$A : 0.5 \times [22] + 0.5 \times [0], \quad B : 0.35 \times [28] + 0.65 \times [0], \quad C : 0.65 \times [15] + 0.35 \times [0],$$

with expected values $EV_A = 11$, $EV_B = 9.8$, and $EV_C = 9.75$. Here Arm A is the unique optimal option, while Arms B and C are EV-inferior. Crucially, Arm A is not salient on any single dimension: it offers neither the largest prize nor the highest win probability. It is deliberately positioned between a high-prize, low-frequency arm (B) and a low-prize, high-frequency arm (C), so that neither a “chase-the-largest-prize” nor a “chase-the-most-frequent-win” heuristic identifies the optimal arm. Only by integrating prize size and win probability into expected value does Arm A emerge as optimal, which separates genuine expected-value learning from simpler payoff heuristics.

Participants are not informed of the prize amount or win probability of any arm in advance; after each round, only the realized outcome (tickets won, or zero) is displayed, so all knowledge of the environment must be acquired through exploration. Placing the equal-payoff environment first gives a clean baseline for unbiasedness: all arms share the same expected value, and participants enter without prior task experience. The unequal-payoff environment then tests expediency and monotonicity. Because it always follows the equal-EV task, some carryover of task experience is possible. However, since expediency and monotonicity are assessed as within-environment trends over rounds, the analysis does not rely on a clean comparison between environments.

Participants are paid through a Binary Lottery Procedure (BLP), first proposed by Smith (1961) and later applied by Roth and Malouf (1979). Each ticket earned in a task serves as a lottery ticket. At the end of each task, the computer draws a random number between 1 and 5,000. If the drawn number is equal to or less than the participant’s total number of tickets, the participant wins a prize of €5; otherwise, the participant receives nothing. Under expected utility, the BLP induces risk-neutral behavior over tickets, since maximizing the expected number of tickets is equivalent to maximizing the probability of winning the fixed monetary prize. At the same time, the BLP may introduce additional behavioral noise. Because each ticket has only a very small marginal monetary value and payment is realized through a lottery, participants may perceive the link between a single bandit choice and final payment as weak. The procedure is therefore incentive-compatible under expected utility, but in practice it may attenuate attention to payoff differences and make learning trends harder to detect.

2.3 Experiment

Instructions are displayed on the screen. After reading them, participants answer three comprehension questions before the standard task and four before the gamified task. The additional question in the gamified task covers the mini-game mechanics. Once all the questions are answered correctly, the experiment begins.

In the standard task, participants choose among three arms in each round. The only information available is the payoff from the previous round. Figure 1 shows the decision interface.

Decision Task

Round 2 of 30

Previous round payoff: 10 points

Please choose one option.

Option A

Option B

Option C

Powered by [gTree](#)

Figure 1: Bandit decision interface (standard task)

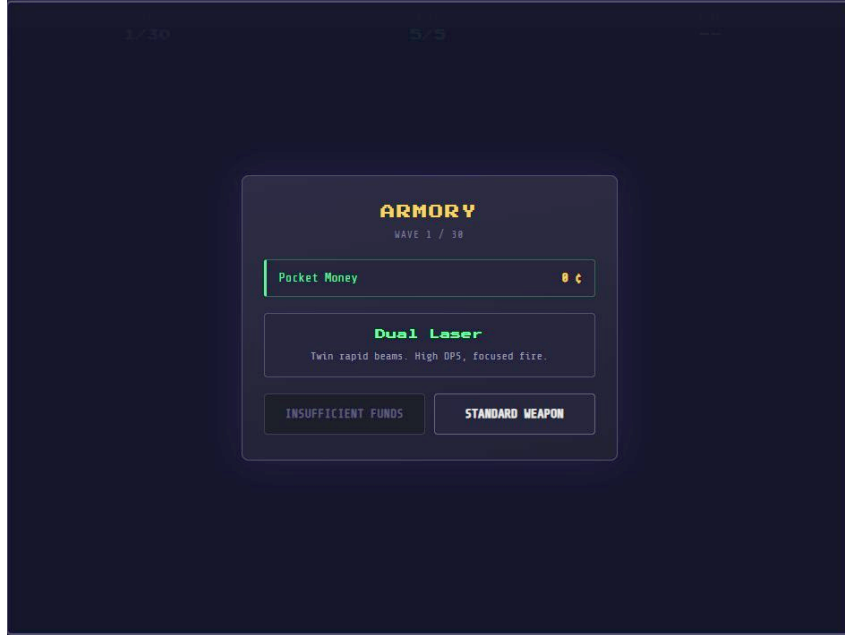
In the gamified task, each round has two stages. Participants first make the same bandit choice as in the standard task and observe whether they win tickets. If they win tickets, they can either keep them for the lottery or spend part of them on a temporary weapon upgrade before entering the mini-game.

The upgrade decision changes the use of tickets, but not the bandit problem itself. Spending tickets on upgrades lowers the participant’s final lottery balance, and therefore slightly lowers the chance of winning the monetary prize. However, it does not change the prize amounts or winning probabilities of the three arms.

The gamified framing therefore adds a non-monetary motive to the task. In the standard task, tickets matter only because they increase the lottery winning probability. In the gamified task, tickets can also buy an immediate in-game benefit. This distinction is important because the game layer may change how participants value wins, even when the underlying payoff distributions are unchanged.

Figure 2 shows the two screens that make up the gamified layer. Panel (a) shows the armory screen, which appears after each bandit round. The screen displays the participant’s available in-game tickets, the current wave, and the weapon option available for the next mini-game. Participants can either spend the tickets earned in the bandit task of the current round to upgrade their weapon or continue with the standard weapon. Panel (b) shows the mini-game itself. The game is an arcade-style space shooter: the participant controls a small spaceship at the bottom of the screen, fires projectiles at incoming enemies, and observes game status indicators such as wave, shield, and time at the top of the screen. In this sense, the interface resembles classic fixed-screen space-shooter games, where the player defends against enemies entering from above.

The purpose of this layer is to add game-like interaction, visual feedback, and short-term goals between bandit decisions. To sustain engagement, the mini-game features several types of enemies and weapons. Crucially, its outcome carries no material consequence: winning or losing produces only an on-screen notice and does not affect participants’ payoffs or the subsequent bandit task.



(a) Weapon upgrade interface



(b) Mini-game interface

Figure 2: Gamified task interface

After completing each task, participants report on a post-task survey their subjective estimate of the winning probability p_j for each of the three arms. Beliefs are elicited separately after each task, and the survey appears immediately after the 30th round of each task before the next task begins. This belief elicitation is unincentivized, following Charness et al. (2021), who find that introspective responses perform comparably to incentivized elicitation. We deliberately avoid incentivizing belief reports to prevent the hedging problem identified by Armantier and Treich (2013): in a multi-option environment, participants facing incentivized belief elicitation may strategically distort their reports to hedge against losses in the decision task, rather than

revealing their true beliefs.

2.4 Procedures

The experiment was programmed in oTree (Chen et al., 2016) and distributed through SurveyCircle and SurveySwap, which are two participant-exchange platforms for online academic research, supplemented by a student network at the University of Vienna. Data were collected in May 2026. Participants ranged in age from 18 to 48 ($M = 25.8$). The final sample comprises 129 participants (80 female, 46 male, and 3 who preferred not to disclose), with 64 in the standard-first group and 65 in the game-first group.

The BLP described above implies very low monetary stakes. With a single €5 prize drawn against 5,000 tickets, each ticket is worth an expected €0.001. A participant therefore expects about €0.24 per task in the equal-payoff environment, and between €0.29 (always choosing the lowest-EV arm) and €0.33 (always choosing the highest-EV arm) per task in the unequal-payoff environment, for a total expected payment below €0.60. Over the roughly 15 minutes the experiment takes, this is an implied hourly rate of about €2.40. The realized average payment was even lower, €0.19 per participant: only about 1.9% of the 258 lottery draws (129 participants over two tasks) returned a €5 win, reflecting both the randomness of the lottery and the fact that participants did not consistently choose the highest-expected-value arm.

This compensation is well below typical laboratory standards, which likely contributed to the participation and attrition difficulties. Although unintended, the low stakes also created a useful setting: they let us ask whether gamified framing can sustain participation and engagement when monetary incentives are weak.

3 Methodology and Simulation

3.1 Learning-property tests and hypotheses

The simulation benchmark is the Cross rule (Cross, 1973), a non-Bayesian learning rule shown by Börgers et al. (2004) to satisfy unbiasedness, expediency, and monotonicity. Let σ_i denote the current probability of choosing arm s_i , and let $x \in [0, 1]$ denote the normalized realized payoff. For all $i, j \in \{1, 2, \dots, n\}$ with $i \neq j$, the Cross rule updates choice probabilities as follows:

$$L(s_i, x)(\sigma_i) = \sigma_i + (1 - \sigma_i)x, \quad (1)$$

$$L(s_j, x)(\sigma_i) = \sigma_i - \sigma_i x. \quad (2)$$

Thus, after arm s_i is chosen, a higher realized payoff increases its future choice probability more strongly. If $x = 1$, the probability of choosing the successful arm moves to one; if $x = 0$, the probability vector remains unchanged.

We evaluate three learning properties: unbiasedness in the equal-payoff environment, and expediency and monotonicity in the unequal-payoff environment. Because the Cross rule satisfies all three properties, the simulations serve as a benchmark: any failure to detect a property reflects the limits of the test rather than absence of the property. They thus show how often each test recovers a property that is genuinely present.

Unbiasedness. In the equal-payoff environment, where $EV_A = EV_B = EV_C$, unbiasedness implies temporal invariance of arm-choice shares. Let $q_{a,t}$ denote the population choice share, defined as the proportion of participants choosing arm a in round t . The null hypothesis for H1 is

$$q_{a,t} = q_{a,s} \quad \text{for all } a \in \{A, B, C\} \text{ and all rounds } t, s.$$

Empirically, we test this implication using an early-late equality test. For each participant and each arm, we compute the change in that arm’s choice share between the early and late parts of the analyzed rounds. The null is that the mean early-late changes are jointly zero across arms. Because the three arm-share changes sum to zero, we use a joint F test on two independent components of the change vector.

The uniform-level test is reported as an auxiliary diagnostic. It asks whether average choice shares are equal to one third for each arm, which captures position bias or salience-driven concentration.

Expediency. Expediency requires that the decision maker’s expected payoff increase over time. In the unequal-payoff environment, let μ_a denote the true expected value of arm a , with $(\mu_A, \mu_B, \mu_C) = (11, 9.8, 9.75)$. For participant i in round t , define

$$v_{it} = \mu_{a_{it}},$$

where a_{it} is the arm chosen by participant i . Therefore, v_{it} is the expected payoff implied by the observed choice, not the realized payoff. The main test estimates

$$v_{it} = \alpha_i + \beta_E \log(t) + \varepsilon_{it}.$$

Expediency predicts $\beta_E > 0$.

Monotonicity. Monotonicity requires that the expected probability assigned to the payoff-maximizing arm or arms increases from one period to the next. In the unequal-payoff environment, Arm A is the unique payoff-maximizing arm, so we refer to it as the best arm. The empirical implication is therefore that the probability of choosing the best arm should increase over rounds. Define

$$b_{it} = \mathbb{1}\{a_{it} = A\}.$$

The main test estimates

$$b_{it} = \alpha_i + \beta_M \log(t) + \varepsilon_{it},$$

again with participant fixed effects and clustered standard errors. Monotonicity predicts $\beta_M > 0$.

The hypotheses follow directly:

H1 (Unbiasedness): In the equal-payoff environment, arm-choice shares are temporally invariant.

H2 (Expediency): In the unequal-payoff environment, the expected value of the chosen arm increases over rounds.

H3 (Monotonicity): In the unequal-payoff environment, the probability of choosing the best arm increases over rounds.

H4 (Framing): Compared with the standard framing, the gamified framing changes either the

level or the time trend of choice performance or response time.

3.2 Simulation results

We run 1000 simulations per property, each with 60 agents, approximately matching the per-framing sample size in the experiment (64 and 65 participants). Table 1 reports fractions of simulated samples. For unbiasedness and its auxiliary uniform-level diagnostic, consistency is defined as a failure to reject the corresponding null hypothesis: temporal invariance for the early-late equality test, and equal one-third choice shares for the uniform-level diagnostic. For expediency and monotonicity, the reported entries are the fraction of simulations in which the aggregate log-round trend test detects a statistically significant positive trend.

Table 1: Simulation recovery rates

Property / diagnostic	Test	Fraction of simulations
(a) Unbiasedness: equal-payoff environment		
Temporal invariance	Early-late equality test	Consistent: 95.0%
Uniform-level	Equal-share diagnostic	Consistent: 94.3%
(b) Expediency: unequal-payoff environment		
Decision maker’s expected payoff	Log-round trend	Detected: 57.9%
(c) Monotonicity: unequal-payoff environment		
Best-arm choice	Log-round trend	Detected: 57.8%

Table 1 shows how often the simulation tests recover the properties that are present in the simulated Cross-rule population. For unbiasedness, temporal invariance is recovered in 95.0% of simulations. The auxiliary uniform-level diagnostic is also consistent with the simulated data in more than 94% of simulations. These results indicate that the tests rarely classify the Cross-rule population as biased in the equal-payoff environment.

For expediency and monotonicity, recovery is much weaker. The log-round trend test detects a statistically significant increase in the decision maker’s expected payoff in only 57.9% of simulations. The result for monotonicity is almost identical: a statistically significant increase in best-arm choice is detected in 57.8% of simulations. Thus, even when all agents follow a rule that has been theoretically shown to satisfy the properties, the 30-round design detects those properties in only about 58% of simulated samples.

We then perform counterfactual simulations to examine how the detection rate scales with the horizon. Table 2 reports the fraction of simulations in which the log-round trend test detects a positive slope for horizons ranging from 30 to 1000 rounds. Extending the horizon raises the detection rate of expediency from 57.9% to 67.1% and of monotonicity from 57.8% to 66.5%.

Table 2: Counterfactual simulations: detection rates under varying horizons

Number of rounds	Expediency	Monotonicity
30 (baseline)	57.9%	57.8%
60	62.3%	62.5%
100	63.8%	63.5%
1000	67.1%	66.5%

4 Results

4.1 General Results

Figure 3 plots choice shares round by round, separately by environment and framing. The figure reveals a strong position bias toward Arm A in the opening rounds: in the equal-payoff environment, Arm A’s share exceeds 0.5 in both framings in the first two rounds, far above the unbiased benchmark of $1/3$. This early dominance is consistent with a primacy effect, in which the first option encountered receives disproportionate attention. The bias is further amplified in the gamified condition.

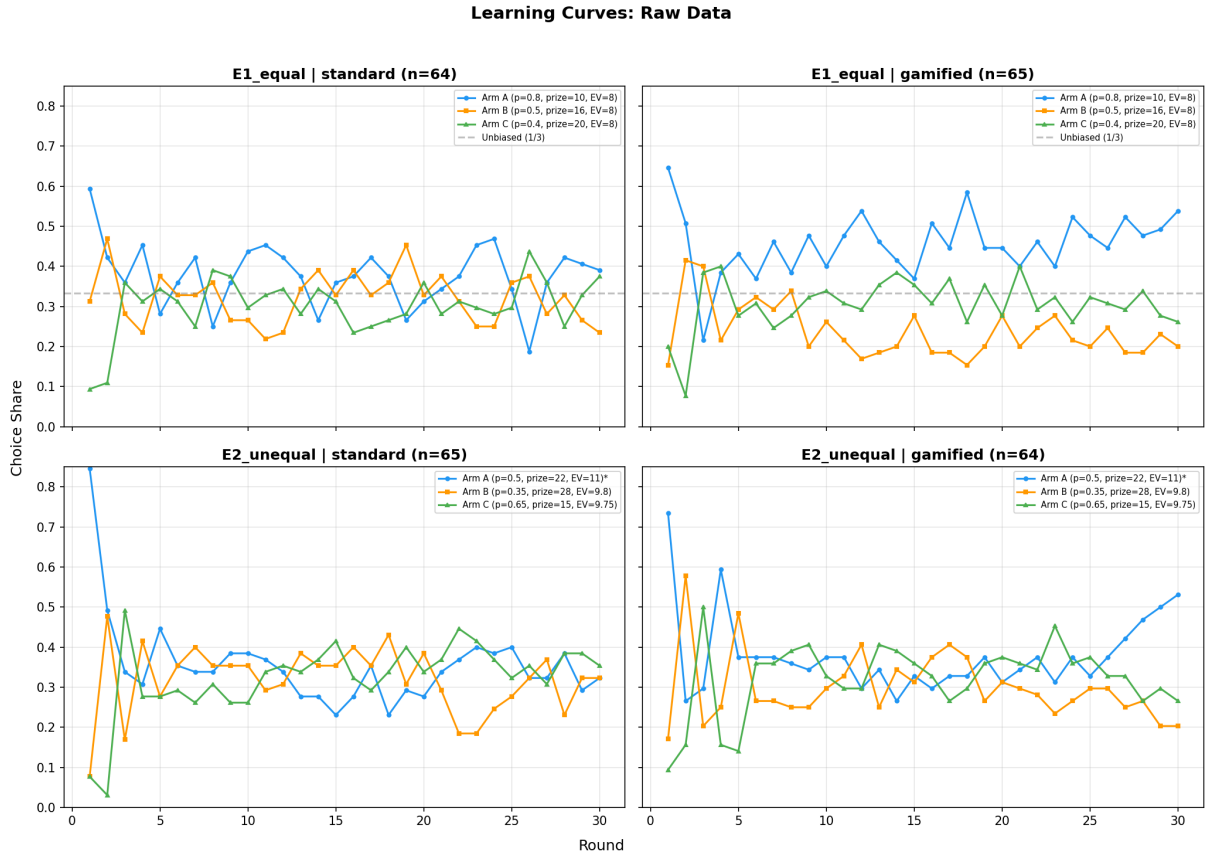


Figure 3: Choice shares by round

In the unequal-payoff environment, the initial position bias is again present but dissipates more quickly. By around round 10, choice shares begin to stabilize. In the standard framing, the three arms remain closely intertwined between roughly 0.3 and 0.4 for most of the task, with Arm A ($EV = 11$) only intermittently and marginally ahead. In the gamified framing, a clearer separation emerges in the final rounds: the share of Arm A rises toward 0.5 while Arms B and C decline. These descriptive patterns suggest that the aggregate dynamics are not pure random fluctuation, and that participants move away from the initial position bias. Although the formal tests reported later do not provide robust statistical evidence of expediency or monotonicity, the curves—particularly under the gamified framing—are weakly consistent with gradual concentration on the highest-EV arm. Throughout the empirical analysis, we drop the first four rounds and use rounds 5–30, since the earliest rounds reflect initial position bias

and high choice noise before learning begins.

Figure 4 shows that median response times decline sharply from approximately 4,000 ms in the first few rounds to below 1,000 ms by round 10. This pattern is consistent across all four conditions and likely reflects a combination of task familiarization and a shift from exploration to exploitation. Notably, response times in the gamified conditions start at a similar level but remain slightly elevated in mid-rounds compared to the standard conditions, suggesting that the game framing may sustain engagement and deliberation for longer.

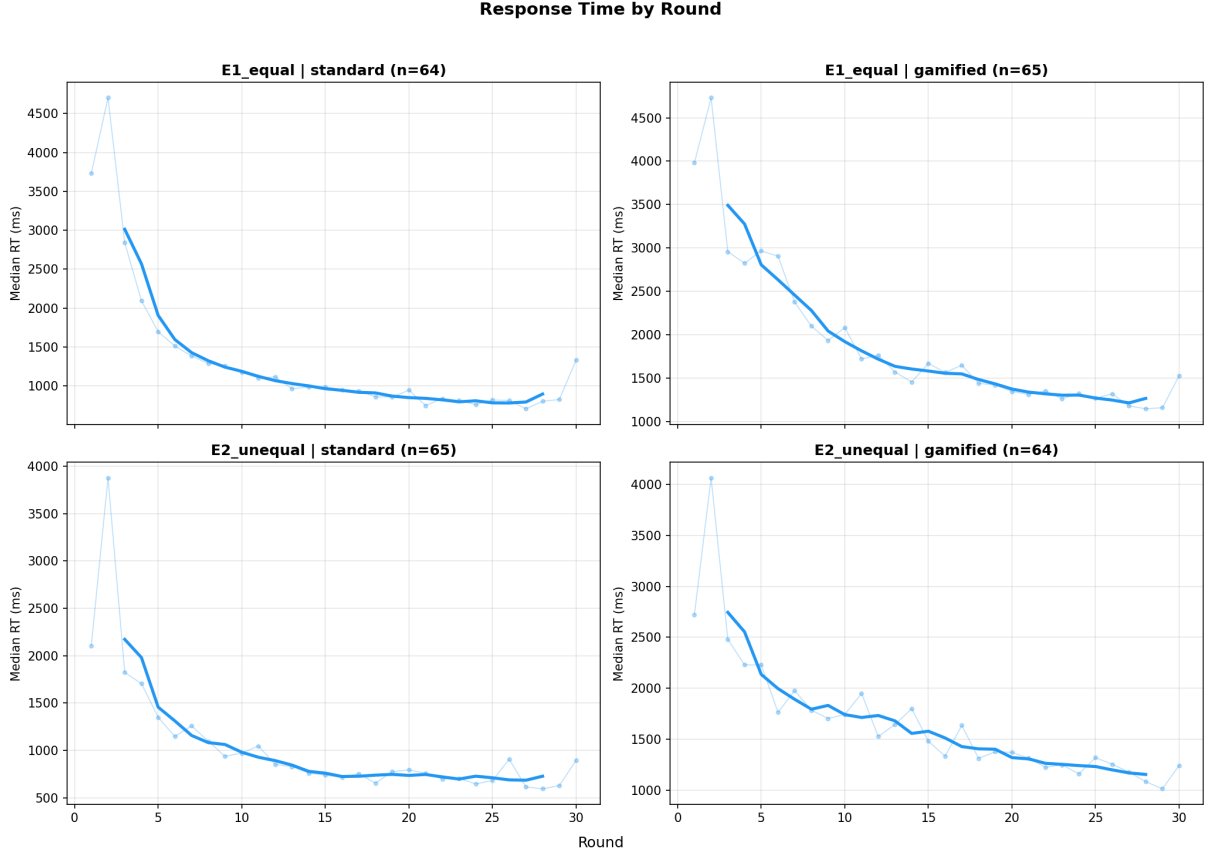


Figure 4: Median response time by round

4.2 Learning Properties

We now apply the learning-property tests to the experimental data. Each test is computed separately for the standard and gamified framings to assess whether behavior within each framing is consistent with the corresponding learning-property benchmark.

The H1 unbiasedness test is implemented as the same early-late equality test described in Section 3, separately within each framing. We report per-arm mean changes and 95% confidence intervals, while the joint inference comes from the vector F test. The uniform-level test remains an auxiliary diagnostic and asks whether average choice shares differ from one third.

For H2 and H3, the experimental implementation differs from the aggregate simulation tests by using participant fixed-effects log-round regressions with standard errors clustered by participant. For H2, the outcome is the expected value of the chosen arm, $v_{it} = \mu_{a_{it}}$. For H3, the outcome is the best-arm indicator, $b_{it} = \mathbf{1}\{a_{it} = A\}$. In both cases, the tested prediction is a

positive log-round coefficient.

As a supportive non-parametric check, we also use a two-sided Wilcoxon signed-rank test on participant-level early-late differences. For each participant, the early value is the average outcome in rounds 5–10 and the late value is the average outcome in rounds 25–30. The null hypothesis is identity of the paired early- and late-window distributions. We therefore interpret a significant result as evidence that the early- and late-window distributions differ.

Tables 3 and 4 report the H1 unbiasedness tests in the equal-payoff environment separately by framing. In the standard framing, the early-late equality test does not reject temporal stability of arm-choice shares ($p = 0.9736$). The early-late changes are all close to zero, and the widest 95% confidence interval has a half-width of about 5.3 percentage points. The auxiliary uniform-level test also does not reject equality of average choice shares to one third ($p = 0.4285$); the widest confidence interval for the average choice shares has a half-width of about 4.8 percentage points. Thus, the standard framing provides no statistically significant evidence against either temporal stability or equal average shares, although modest deviations cannot be ruled out.

Table 3: Temporal invariance and auxiliary diagnostics: standard framing

Item / pair	Mean	95% CI	Statistic / p
Temporal invariance: early-late equality			
Arm A	-0.006	[-0.059, +0.047]	$F(2, 62) = 0.027, p = 0.9736$
Arm B	+0.001	[-0.047, +0.049]	
Arm C	+0.005	[-0.048, +0.057]	
Uniform level: shares equal to one third			
Arm A	0.365	[0.317, 0.412]	$F(2, 62) = 0.859, p = 0.4285$
Arm B	0.320	[0.282, 0.359]	
Arm C	0.315	[0.273, 0.357]	

Notes: $n = 64$. The analysis drops the first four rounds and uses rounds 5–30.

As shown in Table 4, the gamified framing also shows no statistically significant evidence against temporal stability. The early-late equality test does not reject stability over time ($p = 0.2660$), and all early-late confidence intervals include zero. The widest confidence interval has a half-width of about 5.2 percentage points, so moderate early-late changes cannot be ruled out. By contrast, the auxiliary uniform-level test rejects equality of average choice shares to one third ($p < 0.001$). Arm A is chosen substantially more often than one third (0.460), while Arm B is chosen substantially less often (0.228). Thus, the gamified framing is consistent with temporal stability, but it shows a clear stable level difference across arms.

Table 4: Temporal invariance and auxiliary diagnostics: gamified framing

Item / pair	Mean	95% CI	Statistic / p
Temporal invariance: early-late equality			
Arm A	+0.037	[−0.010, +0.084]	$F(2, 63) = 1.352, p = 0.2660$
Arm B	−0.024	[−0.068, +0.021]	
Arm C	−0.013	[−0.065, +0.039]	
Uniform level: shares equal to one third			
Arm A	0.460	[0.390, 0.530]	$F(2, 63) = 13.570, p < 0.001^{***}$
Arm B	0.228	[0.187, 0.270]	
Arm C	0.312	[0.249, 0.374]	

Notes: $n = 65$. The analysis drops the first four rounds and uses rounds 5–30.

Table 5 reports the H2 results. In the standard framing, neither the FE coefficient nor the Wilcoxon comparison provides evidence of the predicted learning pattern. In the gamified framing, the FE coefficient is positive but not significant, and the two-sided Wilcoxon check does not reject equality of the early- and late-window distributions. The result is therefore not treated as evidence for expediency.

Table 5: Expediency tests in the unequal-payoff environment

Test	Standard	Gamified
FE log-round regression, $\hat{\beta}_E$	-0.0313 (0.7484)	+0.0435 (0.1164)
Wilcoxon signed-rank	$W = 791.0, p = 0.4819$	$W = 806.0, p = 0.1184$

Notes: FE entries report one-sided p -values. Wilcoxon p -values are two-sided. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 6 reports the H3 results. In the standard framing, neither test supports monotonicity. In the gamified framing, the FE coefficient remains insignificant, while the two-sided Wilcoxon comparison provides weak auxiliary evidence that the early- and late-window distributions differ. The result is therefore suggestive rather than conclusive.

Table 6: Monotonicity tests in the unequal-payoff environment

Test	Standard	Gamified
FE log-round regression, $\hat{\beta}_M$	-0.0235 (0.7342)	+0.0368 (0.1068)
Wilcoxon signed-rank	$W = 575.0, p = 0.7122$	$W = 516.5, p = 0.0777^*$

Notes: FE entries report one-sided p -values. Wilcoxon p -values are two-sided. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Because the FE log-round tests do not provide robust evidence of expediency and monotonicity, we next examine the sensitivity of these null results. This analysis asks how large a positive log-round slope would have to be for the current design to detect it reliably. Holding the estimated clustered standard errors fixed, we compute the minimum detectable effect (MDE) for each FE trend test at 80 percent power and a one-sided 10 percent significance level. We then translate the slope MDE into the implied outcome change over rounds 5–30 by multiplying it by $\log(30) - \log(5)$.

Table 7: Sensitivity analysis for FE log-round tests

Property	Framing	Obs. β	MDE β	MDE window change
Expediency	Standard	-0.0313	+0.0996	+0.1784 EV points
Expediency	Gamified	+0.0435	+0.0774	+0.1387 EV points
Monotonicity	Standard	-0.0235	+0.0801	+0.1435 best-arm prob.
Monotonicity	Gamified	+0.0368	+0.0627	+0.1124 best-arm prob.

Table 7 shows that the FE tests are imprecise. In the standard framing, the observed slopes are negative. In the gamified framing, the observed slopes are positive but smaller than the corresponding MDEs. The current design therefore has limited ability to detect directional learning effects of this size.

Overall, H1 is not rejected in either framing. The auxiliary uniform-level diagnostic, however, indicates a stable choice-level difference in the gamified framing. In the unequal-payoff environment, neither H2 nor H3 is robustly detected by the main FE log-round tests. The Wilcoxon checks are used only as auxiliary paired-window tests of whether the early- and late-window distributions differ. They do not provide evidence of such a difference for H2, while the gamified H3 comparison is weakly significant at the 10 percent level. Because these checks compare only two windows and do not model the full round-by-round learning path, they do not overturn the main FE results.

4.3 Framing Difference

The property tests in Section 4.2 describe the learning properties separately by framing. This section tests H4 directly using panel regressions with framing indicators and framing-by-round interactions.

For the pooled outcomes, we estimate variants of the following specification:

$$y_{it} = \alpha + \beta G_{it} + \gamma r_t + \delta(G_{it} \times r_t) + \eta U_{it} + \kappa(G_{it} \times U_{it}) + X_i' \theta + \varepsilon_{it},$$

where $G_{it} = 1$ for observations in the gamified framing and 0 for the standard framing, r_t is the round number, and $U_{it} = 1$ for the unequal-payoff environment. Standard errors are clustered at the participant level.

The exact set of regressors differs by outcome. For choosing the optimal arm, the regression is estimated only in the unequal-payoff environment, so no environment dummy or framing-by-environment interaction is included. For expected payoff, the pooled regression includes both the unequal-environment dummy and the gamified-by-unequal interaction. For log response time, the pooled regression includes the unequal-environment dummy and a switching indicator, in addition to the gamified-by-round interaction. In all specifications, X_i includes age, economics coursework, and previous economics experiment experience.

Table 8 reports the framing-related coefficients for three outcomes: choosing the optimal arm in the unequal-payoff environment, the decision maker’s expected payoff, and log response time. To focus on post-initial behavior, all regressions exclude the first four rounds.

Table 8: Framing effects in panel regressions

Variable	Choosing optimal arm	Expected payoff	Log response time
Gamified	-0.0493 (0.3994)	-0.0491 (0.1594)	0.6576 (0.0000)***
Gamified $\times$ Round	0.0045 (0.1597)	0.0028 (0.1632)	-0.0033 (0.2544)
Gamified $\times$ Unequal	—	0.0363 (0.3864)	—

Notes: Entries report coefficients with p -values in parentheses. The table reports only framing-related coefficients; control variables are included but not displayed.

For both choosing the optimal arm and expected payoff, the gamified main effect is negative and the gamified-by-round interaction is positive, but none of these coefficients is statistically significant. The panel regressions therefore provide no robust evidence that gamification changes choice performance or learning trends relative to the standard framing.

The strongest framing effect appears in response time. Participants take substantially longer in the gamified task, and this effect is large and statistically significant. The gamified-by-round interaction is not significant, so the longer response time persists across rounds rather than being driven by a different time trend. This sustained elevation is itself informative for data quality. A recurring concern in online and unsupervised experiments is that participants disengage and respond mechanically once the task becomes repetitive, collapsing into early, low-effort exploitation. The persistently higher response times under the gamified framing suggest the opposite: the mini-game appears to sustain deliberation deeper into the task, consistent with an interruption effect in which the periodic mini-game repeatedly re-engages attention. Gamification thus changes the decision process primarily by sustaining engagement and deliberation, even though this does not translate into robustly improved payoff performance.

To examine this engagement more directly, we turn to participants’ behavior within the mini-game itself. Because an upgrade costs tickets and thereby slightly lowers the chance of winning the monetary prize, purchasing upgrades is not payoff-maximizing; the decision to buy therefore reflects how much a participant engages with the game for its own sake rather than any monetary incentive calculation. We use the upgrade purchase rate as a behavioral measure of this engagement and classify participants into four buyer types accordingly. Buy rate is defined as the conditional share of upgrade opportunities in which the participant bought a weapon upgrade, where upgrade opportunities are rounds in which the participant won at least one ticket. Never-buyers did not purchase any upgrades. Occasional buyers purchased upgrades in more than 0 but at most 50 percent of opportunities, moderate buyers in more than 50 but at most 80 percent, and frequent buyers in more than 80 percent.

As shown in Table 9, the upgrade mechanism generated substantial behavioral variation rather than being ignored. A majority of participants purchased at least one upgrade, and almost one fifth of the sample are classified as frequent buyers who upgraded in more than 80 percent of available opportunities. Even among occasional buyers, participants still interacted with the upgrade option selectively rather than avoiding it entirely. This pattern suggests that the upgrade feature successfully created an active decision margin within the gamified task: participants noticed the option, weighed it against their available tickets, and differed meaningfully in how often they chose to use it.

Table 9: Buyer type classification

Buyer type	Definition	Participants
Never buyer	Buy rate = 0	24
Occasional buyer	$0 < \text{Buy rate} \leq 50\%$	57
Moderate buyer	$50\% < \text{Buy rate} \leq 80\%$	22
Frequent buyer	Buy rate $> 80\%$	26

The upgrade option also has an incentive implication. Participants spent tickets to buy upgrades, and these tickets would otherwise have raised their chance of winning the prize, creating a voluntary trade-off inside the gamified task between keeping tickets for the lottery and spending them on in-game improvements. Importantly, this again leaves the underlying bandit payoff structure unchanged. Table 10 reports the expected monetary value of tickets spent on upgrades. The 5,352 tickets spent correspond to an expected payout reduction of EUR 5.35, which means a 15.0 percent reduction relative to the expected gross lottery cost of EUR 35.74—the payout that would have been incurred if no tickets had been spent on upgrades.

Table 10: Expected incentive offset from weapon upgrades

Group	Environment	Gross tickets	Spent tickets	Offset (EUR)	Reduction
Standard-first	Unequal	20,240	2,299	2.299	11.4%
Game-first	Equal	15,504	3,053	3.053	19.7%
Total	–	35,744	5,352	5.352	15.0%

Notes: Each ticket has an expected value of EUR 0.001, based on a EUR 5 lottery prize with winning probability equal to tickets divided by 5,000.

4.4 Belief Elicitation

We now turn from choice dynamics to the relationship between participants’ stated beliefs and their actions. This serves two purposes: it provides a convergent check on whether choices reflect subjective expected-value maximization, and it offers a further dimension on which the two framings can be compared. The subjective expected value of each arm is the participant’s elicited belief about that arm multiplied by its payoff. An observation is classified as belief-consistent if the arm with the highest subjective expected value is also the modal choice over rounds 5–30. We then test whether belief-consistent observations are associated with better performance.

Table 11 reports belief-action consistency. Overall, 58.5 percent of the 258 observations are classified as consistent, indicating that stated beliefs and observed choices are only moderately aligned. Consistency is somewhat higher under the standard framing (62.0 percent) than the gamified framing (55.0 percent).

Table 11: Belief-action consistency (rounds 5–30)

	N	Consistent	Inconsistent	Consistency rate
Overall	258	151	107	58.5%
Standard framing	129	80	49	62.0%
Gamified framing	129	71	58	55.0%

Table 12 examines whether belief-action consistency is associated with better performance, using OLS regressions of each performance outcome on a belief-consistency dummy in the unequal-payoff environment only, controlling for framing, task order, and previous coursework and experiment experience. The coefficient on the consistency dummy is positive for both outcomes: consistent observations have a higher expected payoff by 0.0623 points and a higher best-arm choice rate by 4.77 percentage points. In both cases the positive coefficient falls short of conventional significance ($p = 0.127$ and $p = 0.152$), so the association between belief-action consistency and performance is directionally consistent across the two outcomes but not statistically reliable.

Table 12: Performance and belief-action consistency

	(1) Expected payoff	(2) Best-arm chosen rate
Belief-action consistency	0.0623 (0.127)	0.0477 (0.152)
Observations	129	129

Notes: Two-sided p -values in brackets, based on standard errors clustered at the participant level.

4.5 Types and Strategies

In this section, we classify participants’ behavioral patterns into different categories. Following the likelihood-based strategy classification approach used in the bandit-strategy literature, we classify each participant-task observation by comparing the likelihood of the observed choice sequence under several candidate behavioral models. This approach is similar to Hudja and Woods (2025), who estimate the likelihood of deterministic strategies with a tremble parameter and use BIC for model comparison. In this thesis, we adapt the same framework to a mixture of probabilistic learning rules and deterministic rules with noise.

Table 13: List of strategy types used in the MLE classification

Name	Description	Parameters
Cross rule	Chooses arms according to a probability vector that starts uniform. After each round, the probability of the chosen arm is increased toward one in proportion to its normalized realized payoff, scaled by a learning rate λ , while the probabilities of the unchosen arms are shrunk proportionally.	Learning rate λ
Random	Chooses each arm with equal probability in every round, independent of previous choices and payoffs.	None
Naive-payoff	Chooses the arm that has produced the highest single realized payoff observed so far; before any positive payoff has been observed, chooses uniformly. With probability ε chooses uniformly at random, and ties are split evenly.	Tremble ε
Naive-frequency	Chooses the arm with the highest observed success count (number of positive payoffs) so far; before any success, chooses uniformly. With probability ε chooses uniformly at random, and ties are split evenly.	Tremble ε
Stubborn	Repeats the arm chosen in the first round in every subsequent round. With probability ε chooses uniformly at random instead.	Tremble ε
win-stay-lose-shift	After a positive payoff, stays with the same arm with probability p_{stay} and otherwise splits across the other two arms. After a zero payoff, switches with probability p_{shift} , split across the other two arms.	$p_{\text{stay}}, p_{\text{shift}}$

Notes: In all models the first-round choice probability is uniform (1/3). Parameters are estimated by maximum likelihood for each participant–task observation.

For participant-task observation i and model m , the log-likelihood is

$$LL_{im}(\theta_m) = \sum_{t=1}^T \log P_m(c_{it} \mid c_{i,1:t-1}, x_{i,1:t-1}; \theta_m), \quad (3)$$

where c_{it} is the observed choice in round t , $x_{i,1:t-1}$ is the previous payoff history, and θ_m is the parameter vector of model m . We estimate θ_m by maximizing the log-likelihood, or equivalently minimizing the negative log-likelihood. The Bayesian information criterion is then calculated as

$$BIC_{im} = 2NLL_{im}(\hat{\theta}_{im}) + k_m \log(T), \quad (4)$$

where k_m is the number of estimated parameters and $T = 30$. Each participant-task observation is assigned to the model with the lowest BIC.

The classification results are reported in Table 14. The most common type is win-stay-lose-shift, followed by Random and the Cross rule. This broadly matches Hudja and Woods (2025), who also find that the large majority of subjects are best fit by either a probabilistic win-stay-lose-shift rule or reinforcement learning, with win-stay-lose-shift fitting the most subjects. Our results diverge in that Random is classified more often than Cross (the reinforcement-type rule).

The distribution also differs across framings. Win-stay-lose-shift dominates the standard task, whereas the gamified task shows higher shares of Random, Cross, and especially Naive-

frequency. The last of these is the most telling: because upgrades could only be purchased with the payoff earned in the current round, the mechanic made positive-payoff events more salient, leading some participants to track how often each arm paid out rather than its full expected value. The increased weight on payoff frequency thus surfaces not only in behavior within the task but also in the distribution of estimated strategy types, reinforcing the interpretation that gamification shifts what participants value.

Table 14: MLE strategy classification by framing

Strategy type	Overall	Standard	Gamified
Win-stay-lose-shift	94	64	30
Random	69	30	39
Cross	55	20	35
Naive-payoff	18	10	8
Naive-frequency	13	3	10
Stubborn	9	2	7
Total	258	129	129

5 Discussion

A central finding of this study is that the directional learning properties proposed by Börgers et al. (2004), expediency and monotonicity, are hard to observe under realistic experimental conditions, whereas the data are consistent with unbiasedness under both framings. Several factors contribute to this gap between the directional theory and the evidence.

First and most fundamentally, the directional properties have limited statistical detectability under the sample size and horizon used here, so their weak appearance in the data is anticipated rather than evidence against the theory. The simulations make this explicit: even when all agents follow the Cross rule, which Börgers et al. (2004) show satisfies expediency and monotonicity, the log-round test recovers these properties in only about 58% of replications at the 30-round horizon, and in only about two thirds even at 1,000 rounds, which is far beyond any feasible experimental horizon. With 64–65 participants per framing and $T = 30$, the present design remains in a low-power region. The failure to find robust expediency or monotonicity is therefore largely a power phenomenon: the test cannot reliably distinguish an expedient population from a non-expedient one at this scale.

Second, the short horizon amplifies this power problem. Most participants settle into a stable, low-deliberation pattern after approximately five rounds, as the response-time data show, leaving only a narrow window in which any directional trend could accumulate enough signal to be detected. Simply extending the number of rounds offers no clear remedy, since the simulations show the detection rate rising only modestly even at 1,000 rounds; the binding constraint is the small size of the sample and the horizon relative to how slowly these properties express themselves.

Third, the gap between the theory’s primitives and what participants observe adds further noise that the limited sample cannot average out. The theory is formulated in expectation, yet participants observe only realized payoffs, and this stochastic discrepancy slows the convergence

of choice propensities. In a larger sample this noise would wash out; however, at the present scale it further diminishes statistical power.

A behavioral channel may also be present, but it is secondary to these statistical limitations. The large proportion of participants classified as win-stay-lose-shift, and the roughly 41 percent of cases in which stated beliefs and observed choices diverge, suggest that some participants rely on simple reinforcement heuristics rather than the gradual propensity-updating the theory assumes. Whether this divergence reflects deliberate exploration or simple inattention, it would add to the noise already discussed rather than constitute an independent verdict on the theory. With the current sample we cannot cleanly separate a genuine behavioral departure from the detection failure that the power analysis predicts in any case.

Taken together, these considerations suggest that expediency and monotonicity, while theoretically valid, function in finite decision environments more as asymptotic benchmarks than as empirically testable predictions. This is not because the data contradict them, but because an experiment of the current size cannot recover them reliably. Unbiasedness, by contrast, is within empirical reach.

A second finding concerns how the gamified framing reshapes behavior in the equal-payoff environment. The standard framing is consistent with both temporal invariance and the uniform-level diagnostic. The gamified framing remains temporally stable but changes where choices concentrate: participants choose the high-probability arm more often and the middle arm less often. The framing effect is therefore located in choice levels, not in early-late movement.

A plausible mechanism links this level shift to the upgrade system and to the way it alters the participant’s objective. As established earlier, the gamified framing introduces an additional argument into the utility function: because an upgrade can be financed only from a winning round’s payoff, and because its monetary opportunity cost is small at the individual choice margin, participants who value the game derive utility from winning beyond the monetary lottery alone. Winning frequency therefore becomes instrumentally valuable under the gamified framing, and Arm A, with the highest winning probability ($p = 0.8$), becomes the most reliable source of tickets for upgrades. This gives the high-probability arm an appeal it does not have when tickets matter only through the lottery.

This incentive channel is reinforced by a salience channel. Following salience theory (Bordalo et al., 2012), decision-makers overweight the attribute of an option that stands out relative to the available alternatives. Arm A is the standout on the probability dimension, Arm C on the payoff dimension (prize of 20), while Arm B (probability 0.50, prize 16) is salient on neither. By making winning frequency important, the gamified mechanic both raises the relevance of the probability dimension and amplifies attention to the probability-salient Arm A; Arm B, distinctive on no dimension, loses share, whereas the prize-salient Arm C, which is unaffected by a mechanic tied to winning frequency, holds its position.

The same mechanism may also explain why the gamified condition shows more favorable descriptive learning patterns. By giving tickets an immediate in-game use, the game layer may make outcomes feel more consequential and sustain learning effort.

This matters for experimental design: equal expected value does not guarantee neutral behavior once a framing layer changes the subjective value of arm attributes. A frame that makes one attribute especially useful or salient can generate stable choice concentration even

when the arms are equal in expected value.

A third finding concerns the effect of gamification on the decision process. The gamified task significantly increases response time, and participants actively interact with the upgrade system, yet these process changes do not translate into improved payoff performance. This pattern is consistent with a growing body of evidence suggesting that gamification reliably increases motivation and time on task but does not necessarily improve the quality of task performance.

Hawkins et al. (2013) found that gamified features added to cognitive decision tasks produced no measurable improvement in data quality, despite participants reporting that the gamified versions were more enjoyable. Similarly, Lumsden et al. (2016) conclude in their systematic review that while gamification can improve engagement and reduce attrition in cognitive tasks, the evidence for performance gains is mixed and often null. Gundry and Deterding (2019) offer a theoretical explanation: game elements introduce additional cognitive demands, such as processing narrative context, evaluating upgrade options, and managing in-game resources. In the present design, the upgrade decision and mini-game create a structural interruption between consecutive bandit choices, which may explain the elevated response times. This interruption appears to sustain engagement and deliberation deeper into the task, but it does not induce deeper reflection about arm payoffs.

This thesis also has several limitations. First, with 64–65 participants per framing, the design has limited power to detect modest framing effects. Second, the monetary stakes were far below standard experimental levels, which may have shaped motivation and attrition in ways that interact with the framing comparison. A related limitation is selection into completion. Because the experiment was voluntary and distributed online, participants who remained until the end may have been more tolerant of, or more attracted to, the gamified design than those who dropped out or chose not to participate. The estimated effects of gamification may therefore partly reflect the preferences of participants who selected into completing this type of task, rather than the response of a fully representative subject pool. And the gamified task is one specific implementation, so the process effects we document need not extend to other forms of gamification. Finally, the three arms in the unequal-payoff environment have very close expected values (11, 9.8, and 9.75), leaving little room for a payoff-maximizing tendency to express itself, which may itself contribute to the weak expediency and monotonicity signal, independently of statistical power.

6 Conclusion

This thesis tests three theoretical learning properties—unbiasedness, expediency, and monotonicity—in a three-armed bandit experiment and examines whether gamified framing affects learning behavior and experimental costs.

The main finding is that the three properties differ sharply in how readily they can be detected. Unbiasedness, defined as temporal invariance of arm-choice frequencies, is confirmed under both framings: the early–late equality test does not reject in either framing. The directional properties, expediency and monotonicity, are not robustly detected. Human behavior is also more heterogeneous than any single learning rule predicts: participants use a mixture of strategies, and stated beliefs are only moderately aligned with observed choices.

Gamification does not robustly improve learning outcomes, but it changes the decision process: it increases response time, generates active interaction with the task, reduces expected incentive costs, and creates a stable choice-level shift in the equal-payoff environment.

These results suggest two directions for future work. First, longer horizons, larger samples, and richer payoff distributions are needed to test whether the directional properties emerge more clearly when participants have more experience. Second, gamification deserves further investigation as a cost-reduction and engagement tool in online experiments, provided that researchers account for its potential to introduce framing effects.

References

- Arechar, A. A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.
- Armantier, O. and Treich, N. (2013). Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review*, 62:17–40.
- Banks, J., Olson, M., and Porter, D. (1997). An experimental analysis of the bandit problem. *Economic Theory*, 10(1):55–77.
- Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.
- Börgers, T., Morales, A. J., and Sarin, R. (2004). Expedient and monotone learning rules. *Econometrica*, 72(2):383–405.
- Charness, G., Gneezy, U., and Rasocha, V. (2021). Experimental methods: Eliciting beliefs. *Journal of Economic Behavior & Organization*, 189:234–256.
- Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.
- Cross, J. G. (1973). A stochastic learning model of economic behavior. *The Quarterly Journal of Economics*, 87(2):239–266.
- Gundry, D. and Deterding, S. (2019). Validity threats in quantitative data collection with games: A narrative survey. *Simulation & Gaming*, 50(3):302–328.
- Hawkins, G. E., Rae, B., Nesbitt, K. V., and Brown, S. D. (2013). Gamelike features might not improve data. *Behavior Research Methods*, 45(2):301–318.
- Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.
- Hudja, S. and Woods, D. (2025). Strategies in the multi-armed bandit. *Experimental Economics*, 28:974–996.
- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., and Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games*, 4(2):e11.
- Lumsden, J., Skinner, A., Coyle, D., Lawrence, N., and Munafò, M. (2017). Attrition from web-based cognitive testing: A repeated measures comparison of gamification techniques. *Journal of Medical Internet Research*, 19(11):e395.
- Paolacci, G. and Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3):184–188.
- Roth, A. E. and Malouf, M. W. K. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review*, 86(6):574–594.

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B*, 23(1):1–25.

A Experimental Instructions

The instructions below are reproduced verbatim from the oTree interface. Each participant completed the equal-payoff task first and the unequal-payoff task second, with the framing (standard or gamified) counterbalanced across the two tasks. The two framings shared an identical bandit structure and lottery rule; they differed only in their visual and narrative presentation and, in the gamified framing, in the addition of the mini-game and upgrade decision. Bold type reproduces the emphasis shown to participants. Correct comprehension-check responses are marked with ✓; participants could not proceed until all answers were correct.

A.1 Standard Framing

Task instructions. Welcome! You will participate in a task that lasts for **30 rounds**.

How It Works. In each round, you will choose one of **three options**: A, B, or C. Each option works like a slot machine with two possible outcomes:

- With some probability, you **win** a fixed number of points.
- Otherwise, you **win nothing** (0 points).

Each option has its own fixed probability and its own fixed payoff. These values **stay the same** throughout all 30 rounds, but they are **hidden from you**. Your job is to figure out which option is best by trying them out.

Example (for illustration only — not the actual values). Suppose one of the options had these hidden values:

Win Probability	Payoff When You Win
60%	15 points

This means each time you choose this option, you have a 60% chance of winning 15 points and a 40% chance of winning nothing (0 points). Each of the three options in your task works this way, but with its own hidden probability and payoff.

€5 Personal Lottery. Every point you save in this task becomes a **lottery ticket**. At the end of the task, the computer will draw a random number between 1 and 5,000. **If the random number is equal to or less than your total number of tickets, you win €5.** If the random number is higher, you win nothing. The more points you save, the more tickets you collect — and the better your chance of winning. This is your **individual** lottery — your result depends only on your own performance, not on other participants.

Comprehension check. To ensure you understand the rules of the task, please answer the following questions. If any answer is incorrect, you will need to correct it before continuing.

1. In this task, every point you save:
 - Becomes a lottery ticket ✓
 - Is converted into euros directly
 - Has no effect on the lottery

- I am not sure
2. Within this task, the probabilities associated with the options are:
- Fixed throughout the task ✓
 - Changing every round
 - Chosen by the participant
 - I am not sure
3. You win €5 if:
- The random number is equal to or less than my tickets ✓
 - The random number is higher than my tickets
 - Everyone wins regardless of their tickets
 - I am not sure

A.2 Gamified Framing (Space Defender)

Mission briefing. Welcome to **Space Defender**. You will complete **30 waves** of combat across three sectors.

Notice. Your point total has been reset to 0. This task has a **separate €5 personal lottery**. The sector reward probabilities and amounts are **completely different** from the previous task. Treat this as a brand new learning task.

Core Task. Before each wave, choose **one of three sectors**: Alpha, Beta, or Gamma. Each sector works like a slot machine with two possible outcomes:

- With some probability, you **win** a fixed number of coins.
- Otherwise, you **win nothing** (0 coins).

Each sector has its own fixed probability and its own fixed reward. These values **stay the same** throughout all 30 waves, but they are **hidden from you**. Your job is to figure out which sector is best by trying them out.

Example (for illustration only — not the actual values). Suppose one of the sectors had these hidden values:

Win Probability	Reward When You Win
60%	15 coins

This means each time you choose this sector, you have a 60% chance of winning 15 coins and a 40% chance of winning nothing (0 coins). Each of the three sectors in your task works this way, but with its own hidden probability and reward.

€5 Personal Lottery. Every coin saved in your **Bank** becomes a **lottery ticket**. At the end of this task, the computer will draw a random number between 1 and 5,000. **If the random number is equal to or less than your total tickets, you win €5.** If the random number is higher, you win nothing. The more coins you save, the more tickets you collect — and the

better your chance of winning. This is your **individual** lottery — your result depends only on your own performance, not on other participants.

The Mini-Game & Your Choice. After choosing a sector and seeing your coins, those coins become your **Pocket Money** for that wave. Before the mini-game starts, you must decide:

- **Option A — Save your coins.** Your Pocket Money goes straight into your Bank and becomes lottery tickets.
- **Option B — Buy an upgrade.** Spend some Pocket Money on a temporary weapon upgrade. Spent coins do not go to your Bank and cannot become lottery tickets.

Important Rule.

- **Mini-game performance does NOT affect your reward.** Losing HP or failing the mini-game will **not** reduce the coins already saved in your Bank.

Comprehension check. Answer all questions correctly to begin Space Defender. If any answer is incorrect, the question will be highlighted and you will need to correct it before continuing.

1. Buying a weapon upgrade means:

- Fewer coins saved in the Bank and fewer lottery tickets ✓
- More coins saved in the Bank and more lottery tickets
- No effect on coins saved or lottery tickets
- I am not sure

2. If you lose HP or fail the mini-game, this means:

- Your Bank balance is reduced
- Your Bank balance is unchanged ✓
- Your sector reward probabilities will change
- I am not sure

3. In this task, the sector reward patterns are:

- The same as in the previous task
- Different from the previous task ✓
- Chosen by the participant
- I am not sure

4. You win €5 if:

- The random number is equal to or less than my tickets ✓
- The random number is higher than my tickets
- Everyone wins regardless of their tickets
- I am not sure

B Agent-Based Model: Specification

B.1 Environments

Each arm is a two-outcome lottery paying a fixed prize with a fixed probability and zero otherwise. Two environments are simulated, matching the experimental designs.

Table 15: Simulated arm parameters

Environment	Arm	p	Prize	EV
E1 (equal EV)	A	0.80	10	8.00
	B	0.50	16	8.00
	C	0.40	20	8.00
E2 (unequal EV)	A	0.50	22	11.00
	B	0.35	28	9.80
	C	0.65	15	9.75

Unbiasedness is tested in E1, where all three arms share $EV = 8$. Expediency and monotonicity are tested in E2, where Arm A is the unique highest-EV (“best”) arm ($EV = 11$).

B.2 Agent population

The simulation framework implements the same family of behavioral rules used in the MLE classification: Cross, Naive-payoff, Naive-frequency, Random, Stubborn (three position variants), and win-stay-lose-shift. Implementing the full family within one framework ensures that the decision rules assumed in the classification and those used in the simulations are defined identically.

The simulations reported in Section 3.2 use a single population:

- **Pure-Cross** ($n = 60$). Sixty Cross agents. The Cross rule is the only rule in the family that Börgers et al. (2004) prove satisfies unbiasedness, expediency, and monotonicity; a homogeneous Cross population therefore provides the appropriate ground-truth benchmark. In E1 these runs serve as a size check for the unbiasedness tests (the null holds by construction), and in E2 as a power check for expediency and monotonicity (the alternative holds by construction). The remaining rule types are not suitable for this purpose, since they carry no theoretical guarantee of satisfying the properties.

B.3 Agent decision rules

All agents start each task with no information and choose among the three arms $k \in \{\text{Left}, \text{Mid}, \text{Right}\}$, indexed $\{0, 1, 2\}$. After choosing arm k in round t , an agent receives payoff $\pi_t = \text{prize}_k$ with probability p_k and $\pi_t = 0$ otherwise. The eight types differ only in how the choice is made.

Cross. The Cross agent maintains a choice-probability vector $\sigma_t = (\sigma_{t,0}, \sigma_{t,1}, \sigma_{t,2})$, initialized to $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and chooses arm k with probability $\sigma_{t,k}$. Payoffs are normalized by the largest prize in the environment, $x_t = \pi_t / \max_j \text{prize}_j \in [0, 1]$. After choosing arm k and observing x_t , the

chosen arm’s probability is reinforced and the others are reduced proportionally:

$$\sigma_{t+1,k} = \sigma_{t,k} + (1 - \sigma_{t,k}) x_t, \quad \sigma_{t+1,j} = \sigma_{t,j} - \sigma_{t,j} x_t \quad (j \neq k).$$

This is the linear reward–inaction rule of Cross (1973); it has no free parameters and is the benchmark whose theoretical properties the simulation verifies.

Random. Chooses uniformly among the three arms every round, ignoring all feedback. No free parameters.

Naive-payoff. Explores for the first $E_p = 6$ rounds and then exploits the arm with the highest *prize observed when winning*. In the exploration phase it samples arms it has not yet seen pay out; once all three have revealed a winning payoff it switches to the highest-prize arm. After exploration it fixes on the best arm seen so far (chosen at random if none has paid out). Parameter: exploration length $E_p = 6$.

Naive-frequency. Explores for the first $E_f = 10$ rounds, sampling uniformly, then fixes on the arm with the highest empirical *win frequency* $\hat{p}_k = \text{wins}_k / \text{tries}_k$ and chooses it thereafter. Parameter: exploration length $E_f = 10$. The contrast with Naive-payoff—frequency of winning versus size of the prize—is what lets the classification separate frequency-driven from payoff-driven behavior.

Stubborn (three variants). Chooses the same fixed arm every round regardless of feedback. The three variants are Stubborn-A, Stubborn-B, and Stubborn-C, fixed on arms 0, 1, and 2 respectively. No free parameters.

Win-stay-lose-shift (WSLS). Following Hudja and Woods (2025), the agent chooses randomly in round 1. Thereafter, if the previous round won ($\pi_{t-1} > 0$) it repeats the previous arm with probability p_{stay} and otherwise switches to one of the other two arms (chosen uniformly); if the previous round lost ($\pi_{t-1} = 0$) it switches with probability p_{shift} and otherwise repeats. Parameters: $p_{\text{stay}} = p_{\text{shift}} = 0.9$.

Execution order. Within each round all agents act once, in a random order that is reshuffled every round. Since agents do not interact (each faces its own independent draws), this ordering does not affect behavior; it is a property of the `mesa` scheduler.

B.4 Simulation protocol

Each simulation initializes a fresh population and runs for T rounds; every agent chooses one arm per round and receives the corresponding stochastic payoff. We run $n_{\text{sim}} = 1,000$ independent simulations per condition. Simulation s uses random seed s ($s = 0, \dots, 999$), so all results are exactly reproducible. The main horizon is $T = 30$ (matching the experiment); the counterfactual analysis additionally uses $T = 1,000$.

B.5 Detection procedure

Within each simulation, the tests are computed across the n agents. For unbiasedness, the reported consistency rate is one minus the fraction of simulations in which the corresponding null is rejected. For expediency and monotonicity, the reported detection rate is the fraction of simulations in which the one-sided log-round trend test detects a positive trend.

Unbiasedness (E1). *Share level.* For the mean share vector, we test $H_0 : (\bar{s}_{\cdot,0}, \bar{s}_{\cdot,1}) = (\frac{1}{3}, \frac{1}{3})$ with a one-sample Hotelling T^2 test on the first two arms only, since the three shares sum to one. With $p = 2$,

$$T^2 = n (\bar{\mathbf{s}} - \boldsymbol{\mu}_0)^\top S^{-1} (\bar{\mathbf{s}} - \boldsymbol{\mu}_0), \quad F = \frac{n-p}{p(n-1)} T^2 \sim F_{p, n-p}.$$

Early-late change. With $H = \lfloor T/2 \rfloor$, let $\delta_{i,k}$ be the change in agent i 's share of arm k between the first H and last H rounds. The three changes sum to zero, so we test H_0 : change vector $= \mathbf{0}$ with the same Hotelling $T^2 \rightarrow F$ construction on the first two arms.

Expediency (E2). Let $v_{i,t} = \text{EV}_{c_{i,t}}$ be the expected value of agent i 's chosen arm, using $\text{EV} = (11, 9.8, 9.75)$. In each simulation, we compute the aggregate expected payoff

$$\bar{v}_t = \frac{1}{n} \sum_i v_{i,t}.$$

We then estimate

$$\bar{v}_t = a_E + b_E \log(t) + u_t$$

and count the property as detected if a one-sided test rejects $H_0 : b_E \leq 0$ in favor of $H_1 : b_E > 0$.

Monotonicity (E2). Let $m_{i,t} = \mathbb{1}\{c_{i,t} = A\}$, where Arm A is the unique best arm. In each simulation, we compute the aggregate best-arm choice share

$$\bar{m}_t = \frac{1}{n} \sum_i m_{i,t}.$$

We then estimate

$$\bar{m}_t = a_M + b_M \log(t) + u_t$$

and count the property as detected if a one-sided test rejects $H_0 : b_M \leq 0$ in favor of $H_1 : b_M > 0$.