Explore–Exploit Decisions under Resource Depletion: Experimental Evidence on Over-Exploitation

Daniel Gómez-Vásquez[*] César Mantilla[†]

June 30, 2026

Abstract

We introduce a novel experimental environment that extends the explore–exploit problem from the canonical bandit framework to resource management. Unlike in bandit models, where the value of exploration is informational, exploration in our setting replenishes a depleting resource: in each round, participants choose between exploiting the current stock—which is halved in the next round—or relocating to reset it to a high or low level. This environment allows us to study systematic deviations from optimal resource management. We show that dynamic programming and the marginal value theorem, despite their different informational and cognitive requirements, converge to the same exploitation threshold, providing a unique benchmark for optimal behavior. In a lab-in-the-field experiment with 384 participants in Bogotá, framed as a street-vending task, participants systematically over-exploit: over-exploitation is two to three times as frequent as under-exploitation. Street vendors over-exploit even more than a reference sample and earn about 13% less. Among vendors, messages discouraging over-exploitation increase earnings, consistent with this pattern deviating from optimality.

Keywords: explore–exploit dilemma; patch foraging; lab-in-the-field experiment; dynamic decision-making; street vending

JEL Classification Codes: C93, D81, D83, O17

Acknowledgments: We thank Eduardo Ferraz for his contribution to this project, Sofía Almeida and Silvia Ortíz for their excellent research assistance, and all the members of the URosario Undergraduate Experimental Economics Group for their support in the field during the execution of the experiment. Comments from Danila Serra and the attendants to the North-American ESA Meeting 2022 were of invaluable help. Financial Support from the program *"Inclusión productiva y social: programas y políticas para la promoción de una economía formal, código 60185, que conforma la Alianza EFI, bajo el Contrato de Recuperación Contingente No. FP44842-220-2018."* is gratefully acknowledged. The authors used Claude (Anthropic) and ChatGPT's Proofreader Pal (OpenAI) for language editing and consistency checking of the manuscript text. All content was reviewed and approved by the authors, who take full responsibility for the integrity of the work.

[*]Department of Economics & Accounting, College of the Holy Cross, Worcester, Massachusetts, USA.

[†]Corresponding author. Department of Economics, Pontificia Universidad Javeriana, Bogotá, Colombia. *e-mail:* mantillacesar@javeriana.edu.co

1 Introduction

The trade-off between exploiting a known opportunity and exploring an uncertain alternative pervades economic life, from a worker weighing whether to stay in her job or look for a better match, to a firm deciding whether to keep charging its current price or experiment with a new one to learn how demand responds (Miller, 1984; Rothschild, 1974). In economics, this trade-off has been studied predominantly through multi-armed bandit problems, in which an agent repeatedly chooses among alternatives with unknown reward distributions to maximize profits under uncertainty (Bergemann & Välimäki, 2018; Robbins, 1952). In this framework, the value of exploration is informational: sampling an option reveals its profitability, while exploitation consists of repeatedly selecting the option believed to yield the highest return. The defining tension is between collecting immediate payoffs and acquiring information.

In many economic environments, however, the explore–exploit trade-off goes beyond pure information acquisition. Consider a street vendor who must decide whether to keep selling at her current location, where the pool of nearby buyers thins out as the day wears on, or to move to a new spot where demand may be higher. For this vendor, the relevant trade-off is between depleting the resource she currently holds, potential buyers, and relocating to replenish it. In this paper, we study an explore–exploit problem of precisely this kind introducing a new task we call the selling-spot game. It departs from the canonical bandit problem in two substantive ways. First, exploitation endogenously depletes future payoffs: extracting the current stock of resources reduces the stock available in subsequent periods. This creates additional incentives to engage in exploration and depart from the current stock. Second, the main benefit of exploration is to replenish the resource rather than generating decision-relevant information. Relocating resets the stock to a new level drawn from a known, binary distribution. Unlike sampling a bandit arm, exploration leaves beliefs about the environment unchanged; its value is the access to new resources it provides.

These two features bring our game close to the patch-foraging paradigm, a sequential decision problem in which an individual must decide between continuing to collect items in a depleting patch or expending energy to move to a new patch where collectible items are replenished (Lloyd et al., 2023). Due to the diminishing returns of staying, the forager faces a trade-off between exploiting the current patch and exploring future ones. Despite its clear economic structure, patch foraging has remained largely outside economics, flourishing instead in biology and ecology, where it is used to model foraging across species from bees and birds to humans and other primates (Krebs et al., 1974; O'Bryan et al., 2020; Stabentheiner & Kovac, 2016), and in the brain and decision sciences, where it illuminates how organisms trade off rewards against an uncertain environment (Harhen & Bornstein, 2023; Hills et al., 2010; Lloyd et al., 2023).

By bringing the patch-foraging problem into economics we also connect two solution concepts characterizing optimal behavior that, in the medium run, provide similar predictions. On the one hand, a backward-induction rule that was derived using dynamic programming, a familiar tool in macroeconomics. On the other hand, a history-based rule grounded in the marginal value theorem (MVT), the standard solution concept for patch-foraging problems in other disciplines (Charnov, 1976; Lloyd et al., 2023). The MVT prescribes leaving the current patch once the next reward falls below the background reward rate (BRR)—the accumulated rewards divided by the total number of actions taken, including both exploitation and exploration—which serves as an endogenous, history-dependent outside option.

Our second adaptation of the patch-foraging problem to economics comes through framing. We replace the usual trees-and-apples instrument with a street-vending problem. In the selling-spot game, the participant operates a vending cart over a 12-hour journey with the goal of making as many sales as possible. Each hour she decides between selling (exploiting the current patch) and moving the cart to a different location (exploring a new patch). The number of buyers in the current location is known, and she knows that after selling this number will be halved the following hour. If she moves, she spends one hour traveling and arrives at a new spot with either 16 or 64 buyers, each with equal probability.

We conducted an incentivized version of this game with 400 participants in Bogotá, including street vendors (57%), undergraduate students (37%), and public officers (6%). Street vendors are our population of main interest: their daily activity is marked by income fluctuations, uncertainty, and a degree of mobility that parallel the exploit–explore trade-off. The vendor's search for a profitable spot is continual, since locations that are sufficiently uncrowded yet busy may vary daily with policing, weather, and public events.[1] Given the novelty of the paradigm, the nearly 150 students serve as a benchmark population. The public officers serve a complementary comparison. They are working adults—numerate professionals such as accountants, lawyers, and economists—with little prior exposure to laboratory-style decision tasks, and for whom, unlike the vendors, the selling-spot framing does not mirror any daily activity. This makes them a useful intermediate case: they share the vendors' profile as non-student adults unaccustomed to such exercises, yet face a framing as unfamiliar as the students do.

A key property of the selling-spot game is that suboptimal behavior can take two forms: over-

[1] Between 2018 and 2022, the number of street vendors in Bogotá rose by roughly 15%, reaching at least 93 thousand (see: https://www.semana.com/nacion/articulo/la-deuda-historica-de-bogota-en-la-formalizacion-de-vendedores-ambulantes/202214/). This growth is attributed in the short run to the pandemic labor situation and in the medium run to a ruling of the Colombian Constitutional Court (T-243/19) protecting the use of public space for these activities.

exploiting (staying too long in a patch) or under-exploiting (leaving it too soon). Over-exploitation is associated with low motivation induced by rewards and difficulty in delaying gratification (seeing exploration as too costly), whereas under-exploitation is associated with low tolerance for uncertainty (Lloyd et al., 2022; Lloyd et al., 2023). Within the vendor sample we added two framing treatments. Participants in the *Encourage* condition were reminded that they have full rights to use public space according to the Constitutional Court; while those in the *Discourage* condition were reminded that local authorities may protect, reassign, and recover public space.

Three findings emerge. First, in the pooled sample participants tend to over-exploit: once exploitation patterns stabilize (rounds 5–11) they exploit with probability 0.61 against the 0.5 predicted by both optimality rules, and their background reward rate settles well below the optimal benchmark. About two-thirds of all decisions deviate from the optimal rules derived from dynamic programming and the MVT, with over-exploitation as the dominant error. Second, street vendors earn roughly 13% less than our (mostly student) reference sample, a gap partly accounted for by the reference sample's greater ability to profit from repeated exposure to rich patches. Third, among vendors, a message *discouraging* exploitation marginally raised earnings, whereas a message encouraging it did not. This extent of how our framing treatments affect behavior differentially is consistent with the prevalence of over-exploitation.

We contribute to the literature in three ways. First, we propose a game that brings closer the patch-foraging problems to the literature on bandit problems.[2] The canonical bandit problem casts the explore–exploit tension as one between selecting the option believed to be most profitable and generating information about unknown payoff distributions. Recent work has extended this paradigm in several directions: constraining the informational gains to exploration, while exogenously setting the entanglement between exploration and exploitation (Lizzeri et al., 2024), varying the horizon and the number of arms to study reinforcement learning and the fit of strategies previously suggested in the literature (Hudja & Woods, 2025), introducing forced experimentation when an option may become temporarily unavailable (Hudja et al., 2026), and characterizing the strategies and factors that govern experimentation (J. Banks et al., 1997; Hudja & Woods, 2024, 2026). A premise unites these environments: exploration is valuable insofar as it produces information. Even when exploration is disentangled from exploitation (Lizzeri et al., 2024), its gain remains in the informational domain.

Our setting departs from the standard bandit framework. Exploitation depletes the stock of resources available in future periods at a known rate, whereas exploration replenishes the stock

[2] Bandit problems have been used to study firm experimentation on pricing (Rothschild, 1974) and technology adoption (J. S. Banks & Sundaram, 1992); consumer experimentation (Bergemann & Välimäki, 2018); job search and matching (Bergemann & Välimäki, 2001); strategic learning by multiple agents (Bolton & Harris, 1999); and resource allocation (Brezzi & Lai, 2002). See Bergemann and Välimäki (2018) for a survey.

by moving to a new state with an uncertain resource level. Consequently, the value of exploration stems from gaining access to new resources rather than from reducing uncertainty about returns. This resource-oriented formulation changes the optimal decision problem: instead of balancing immediate payoffs against the value of information, agents trade off current extraction against the resulting resource depletion. The problem is therefore closer to patch-foraging models and spatial search than to the canonical bandit, while retaining the cyclical pattern of exploitation and relocation. By centering the exploit–explore choice on the remaining stock, we distinguish and quantify two types of errors—over-exploitation and under-exploitation—thereby linking bandit problems to resource management.

Second, our paper relates to the economics of search, where an agent dynamically decides whether to accept a current opportunity or continue searching under uncertainty (McCall, 1970; Mortensen, 1986; Stigler, 1961). Job-search models, in particular, typically treat employment as an absorbing state: once a suitable job is found, search ends. For workers in the informal sector, and street vendors especially, this assumption is too strong, as their incomes fluctuate and their locations are contestable. Our environment connects more closely to recent work incorporating spatial elements, where agents search across locations subject to mobility costs, local competition, and spatial correlation (Malladi et al., 2025; Manning & Petrongolo, 2017; Urgun & Yariv, 2025). We introduce and experimentally study a search environment with endogenous resource depletion and replenishment, framed around the decision environment of street vendors. The documented over-exploitation pattern contrasts with classical search models, in which search frictions typically lead agents to accept too early. In our setting, moving costs and uncertainty about the new draw discourage relocation even when it is optimal. Vendors exhibit this tendency more strongly than our reference sample, suggesting that the deviation is amplified rather than attenuated among those for whom the problem is most familiar.

Finally, our paper adds to the literature on bounded rationality and procedural choice, which studies how individuals simplify complex decision problems and which decision rules they adopt when optimal behavior is cognitively demanding.[3] A growing experimental literature studies complexity directly (de Clippel & Rozen, 2024), measuring its costs and showing that it shapes choice and induces reliance on heuristics (Arrieta & Nielsen, 2024; Oprea, 2020, 2024), alongside work on choices shaped by limited capacity to process information (Salant, 2011) and on agents choosing how much cognitive effort to exert (Alaoui & Penta, 2022). Closely related, Banovetz and Oprea (2023) provide experimental evidence that individuals often adopt simple heuristics in dynamic environments.

[3] Theories of agents relying on simple procedures rather than exhaustive optimization date back to Simon (1955) and were formalized in the automata literature (e.g., Abreu & Rubinstein, 1988; Neyman, 1985; Rubinstein, 1986).

Our contribution to this strand of the literature is twofold. First, we bring together two solution concepts originating in different disciplines—a forward-looking rule derived from dynamic programming, standard in economics, and a history-based rule derived from the marginal value theorem, the standard solution concept in behavioral ecology—and show that, despite resting on very different informational and cognitive requirements, they prescribe nearly identical behavior, converging to a common stock threshold once the agent has experienced a sufficiently varied history of patches. Second, this convergence sharpens our empirical test: because the cognitively demanding forward-looking optimum and the simple history-based rule prescribe the same threshold once the history is sufficiently varied, the deviations we document can be measured against a single benchmark. The observed over-exploitation is a departure from optimality itself, not an artifact of which solution concept we adopt as the benchmark.

The paper is organized as follows. Section 2 presents the experimental task. Section 3 presents the results. Section 4 concludes.

2 Experimental Task

2.1 The selling-spot game

The player, in the role of vendor, plays for $T = 12$ rounds, framed as working hours, to accumulate the largest possible profit. In each round, the player chooses between exploiting its current selling spot, or exploring a new spot. Exploiting the current spot in round t ($x_t = 1$) adds the current stock of buyers $b_t = S_t$ to the total profits, and reduces the number of remaining buyers in this patch by half ($S_{t+1} = S_t/2$). Exploring a new spot in round t ($x_t = 0$) means that no buyers will be accumulated in that round ($b_t = 0$), but the next round's stock S_{t+1} will be reset to 64 or 16 buyers, with equal probability. Total profits $\pi = pB$ are equal to the number of accumulated buyers $B = \sum_{t=1}^{T} b_t$, multiplied by a piece-rate payment p.[4] To lighten the notation, we normalize $p = 1$, meaning that $\pi = \sum_{t=1}^{T} x_t S_t$. Hence, players have an incentive to exploit the selling spot ($x_t = 1$) when the current stock of buyers is high, and to explore a new spot to restart the stock ($x_t = 0$)—at the cost of not accumulating buyers for one round—when this stock is low.

The optimal decision consists of finding the critical stock level that makes the player switch from exploitation to exploration. We will study the backward-induction optimization using the Bellman Equation, and a history-based optimization using the marginal value theorem, the standard solution concept for patch foraging problems in other behavioral sciences (Charnov, 1976).

[4] In our study, we set $p = 50$ Colombian pesos (COP). By the time of the experiment, 50 COP were about 0.0125 USD. Participants accumulated, on average, 242 buyers, equivalent to 3 USD.

2.2 Theoretical predictions

2.2.1 Dynamic programming: a recursive backward-induction optimization

We use dynamic programming as a recursive method for solving sequential decision problems. By writing the problem's Bellman equation, we can characterize the system at any round and guarantee that, regardless of the system's state, the remaining decisions can constitute an optimal policy (Bellman & Dreyfus, 2015). In our case, and with some abuse of notation, we can write the Bellman's value function that depends on the stock of buyers, $V_t(S_t)$, as follows:

$$V_t(S_t) = \max\left(S_t + \beta V_t\left(\frac{S_t}{2}\right), \beta\left(\frac{V_t(16) + V_t(64)}{2}\right)\right) \tag{1}$$

The maximum of the value function has two arguments. The first one is the value of exploitation. Written recursively, this equals the contemporary exploited stock, plus the time-discounted value function in $t+1$. However, we directly substitute S_{t+1} for $S_t/2$, since the round's decay in the stock follows this pattern. The second argument is the value of exploration. Since this decision comes at the cost of no exploitation in the current period, the whole term is multiplied by the discount factor β. The parentheses indicate that the stock will be reset to 16 or 64 units with equal probability. Therefore, the value of exploration is the discounted average of $V_t(16)$ and $V_t(64)$.

Intuitively, the left argument of the maximum function decreases when choosing to exploit. Since the right argument does not depend on S_t because the potential restart values of the stock are known, the solution to this dynamic programming problem yields a threshold stock level S_t^* that makes the left argument smaller, triggering exploration. The solution to this problem, provided in Appendix A.2, is $S_t^* = 24$ for the infinitely repeated game.

This infinitely repeated case is more general and also helpful to gain some intuition on the recursive problem because the value function does not depend on how many periods are left. However, since our game has a known terminal round ($T = 12$), and the time spent playing our game is relatively short (less than 10 minutes), the finitely repeated solution is more informative. We use backward induction to find the solution in this scenario because V_t depends on β and the periods left before the terminal round, $(T - t)$. In T, the remaining stock is exploited. We then find that, for any previous round, a pattern similar to the solution in the infinitely repeated game emerges: when $t < T$, and assuming that $\beta \to 1$, it is optimal to exploit whenever $S_t \geq 32$ and to explore if $S_t \leq 16$. The fact that exploitation will occur in at most two consecutive rounds simplifies the backward induction problem to a rule applying for $(T - t) > 0$. In Appendix A.2, we provide the complete backward induction solution.

The dynamic programming solution for our finitely repeated problem, which we will call the

DyP rule, can be summarized as follows:

If $S_t \geq 32$ or $t = T$, exploit the current patch. Otherwise, explore a new selling spot.

We can also think about the optimal solution in terms of "cycles" defined by the number of exploitation rounds before exploring again. Rich spots ($S = 64$) generate a 3-step cycle (exploit-exploit-explore), while scarce spots ($S = 16$) generate a 1-step cycle (explore). This cycling structure characterizes both the infinite horizon solution and the finite horizon solution for any $k \geq 2$ rounds remaining before the terminal round T.

2.2.2 Marginal Value Theorem: a history-based optimization

The MVT states that the current patch should be exploited as long as its instantaneous reward exceeds the background reward rate (Charnov, 1976; Lloyd et al., 2023), defined as the average reward across *all* past rounds (both exploitation and exploration). Recall that $b_t = S_t$ when exploiting and $b_t = 0$ when exploring. The background reward rate entering round $t+1$ is:

$$BRR_{t+1} = \frac{\sum_{s=0}^{t} b_s}{t+1}, \tag{2}$$

that is, the mean reward up to and including round t. The optimal decision rule, henceforth the *MVT rule*, is:

If $S_t \geq BRR_t$, or $t = T$, exploit the current patch. Otherwise, explore a new patch.

This rule implies exploitation in the first round, since BRR_0 is undefined and no prior earnings exist. In Appendix A.3, we provide four step-by-step examples of how the *MVT* rule operates. These examples show that when the agent encounters a variety of patch types, the background reward rate BRR_t converges toward a value close to the threshold S^* of the *DyP* rule, leading both rules to agree on the optimal decision. Conversely, repeated exposure to the same patch type prevents this convergence.

2.3 Comparative analysis of predictions

To understand whether the observed behavior in our problem is closer to the *DyP* or *MVT* rule, we need to examine how and when their predictions differ. Given the randomness of the initial patch stock and the game's length, we rely on simulations to compare the predictions of each decision rule in our setting. We conducted 5,000 simulations using the *DyP* rule, the *MVT* rule,

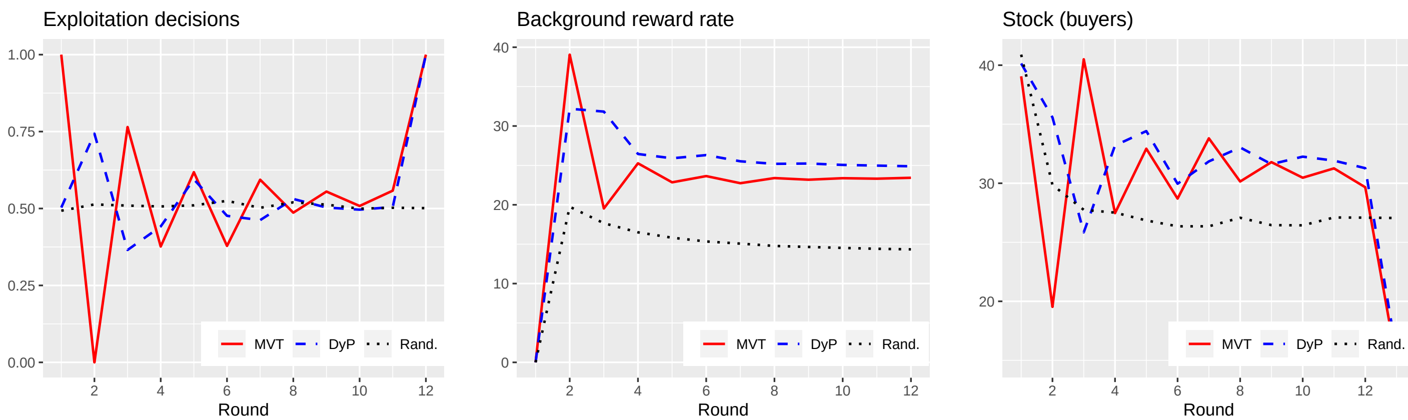

Figure 1: Exploitation decisions, background reward rates, and stock levels for three decision models: marginal value theorem (MVT), dynamic programming (DyP), and random exploit-explore decisions. Average values from 5,000 repetitions of the simulation.

and a *random* decision maker that, in every period, chooses to exploit or explore with the same probability, regardless of the stock level.[5] Figure 1 reports these results for three variables of interest: the decision x_t, the stock of buyers S_t, and the background reward rate BRR_{t+1}.

The leftmost panel depicts the probability of an exploitation decision. The two optimality rules dictate oscillating and different exploitation patterns for the first 4 rounds, and then become more similar. The initial probability of exploitation in the *DyP* rule is 0.5, as it depends on whether the participant starts in a rich spot ($S_1 = 64$) or a scarce spot ($S_1 = 16$). This is followed by a 75% exploitation rate, 50% comes from those previously exploiting the stock that now face ($S_1 = 32$), and the other 25% is half of the half that explored and landed in $S_2 = 64$.

By contrast, under the *MVT* rule, since BRR_0 is undefined, the agent always exploits in round 1, and the high initial background reward rate discourages exploitation in round 2. The larger initial amplitude, and further convergence, of the *MVT* rule comes from initial 2-step cycles (exploit-explore) while the agent gathers environmental information about the selling spots. As the BRR_t accumulates history and environmental information from different restarting spots, it stabilizes in values similar to the S^* in the *DyP* rule. In fact, the central panel reports the dynamics of the background reward rate for the three decision rules. The key insight is that the *MVT* and the *DyP* rules converge, from below and above, respectively, to 24. This is the same decision threshold $S^* = 24$ computed for the infinitely repeated game. Hence, an important result of our theoretical comparison is that the exploitation threshold of the two optimality rules become indistinguishable once

[5] Twelve consecutive exploration rounds yield $2^{12} = 4,096$ possible outcomes, so 5,000 simulations ensures coverage of the full outcome space as a simple rule of thumb.

the *MVT* has acquired sufficient information from the environment. Note also that, under random decision-making, the *BRR* falls below 16, getting farther from the optimality models. This helps us define some natural boundaries for the empirical background reward rate and check whether it approaches optimality.

The rightmost panel completes the depiction of how optimality rules dictate behavior by reporting the average stock level. Note again the more pronounced fluctuations in rounds 1-4 under *MVT* than under the *DyP* rule: the short cycle under *MVT*, regardless of $S_1 \in \{16, 64\}$, resets S_3 in all the simulations, with an average of (64+16)/2=40. Note that, over time, both optimality rules also keep the average stock around 32. This follows from the cycling structure of the optimal policy. In each cycle, the agent observes stocks of 64, 32, and 16 (rich spot) or just 16 (scarce spot), with equal probability. The expected sum of stocks per cycle is therefore $0.5 \times (64 + 32 + 16) + 0.5 \times 16 = 64$, and the expected cycle length is $0.5 \times 3 + 0.5 \times 1 = 2$ rounds, yielding a long-run average of $64/2 = 32$. The average stock under random decision-making is lower, given that $S \leq 8$ could be exploited.

The dynamics of the background reward rate and the stock are helpful to see that, even if the probability of exploitation is around 1/2 with the three decision rules, the underlying patterns are different. The *DyP* and *MVT* rules, in the medium run, converge to the 3-step cycles after landing on a rich spot, or the 1-step cycle after landing on a scarce spot. The expected cycle length is therefore 2 rounds, with an expected number of exploitation rounds of $0.5 \times 2 + 0.5 \times 0 = 1$. The long-run exploitation probability is thus $1/2 = 0.5$, but the optimal rules are far from random exploitation.

2.4 Hypotheses

In this section, we present two types of hypotheses. First, we test some hypotheses that allow us to detect the empirical deviations from following the *DyP* and *MVT* rules. Second, we propose hypotheses to examine differences among the subgroups participating in our study.

2.4.1 Behavior with respect to the optimality rules

Figure 1 is helpful to make some testable predictions regarding optimal behavior, irrespective of the employed rule:

H1. *After some experience, participants behave as if they follow one of the optimization rules. If this is true, then:*

- H1A. In rounds 5-11, they exploit if $S_t \geq 32$ and explore if $S_t \leq 16$.

- H1B. The background reward rate at the end of the game is 24.

In H1 we did not include the exploitation patterns during the first 4 rounds because these periods help us identify the differences between the *DyP* and *MVT* rules. For stating H2, we need first to define a *cycle* as the number of consecutive periods with exploitation, and being closed by an exploration decision. For instance, a vector of four decisions $x_t = (1, 1, 0, 0)$ (i.e., exploit-exploit-explore-explore) constitutes two cycles, one of length 3, and then one of length 1. We thus have H2:

H2. *Early in the game, participants behave more closely to the MVT than to the DyP rules. If this is true, then in rounds 1-4:*

- H2A. Participants explore with positive probability when $S_t = 32$ and exploit with positive probability when $S_t = 16$.

- H2B. The average cycle length is the same, regardless of the starting stock.

2.4.2 Between-sample and between-treatment comparisons

We conducted the game with three populations and collected 391 observations, of which 384 were available for the analysis:[6] a sample of street vendors ($N = 223$); two samples of students, from non-STEM and Economics ($N = 87$ and $N = 52$, respectively), and a small sample of public officers ($N = 22$). Our main interest dwells in the street vendors, since this game is designed to capture elements of the exploit-explore problems they face during their daily activities. Students, on the other hand, are a helpful benchmark given the novelty of the proposed game within the experimental economics domain. The sample of public officers constituted a convenient sample, part of a strategy course for public servants, helpful for drawing comparisons between non-student populations.

In the following hypotheses, we exclude the public officers given our reduced sample size. We focus on the expected differences in performance, translated into earnings, because street-vendors are more likely to over-exploit. We propose two mechanisms explaining why this may happen. First, if they bring into the game their own experience, exploring new locations may encompass an exploration cost that includes novelty but, also, potential disputes with other vendors. Second, attention and numeracy may also put the street-vendors in disadvantage with respect to students, as they are more used to following instructions and retaining new information. To separate the role of attention from numeracy abilities, we included a sample of non-STEM students (including

[6]The decisions failed to record for the remaining 7 participants.

Economics and Finance) and a sample of Economics students who were visiting the university for a disciplinary contest between undergraduate students of Colombia's Economics Departments. We thus have our third hypothesis:

H3. *Street-vendors' earnings are lower than students' earnings.*

Finally, given the potentially higher susceptibility of street-vendors to the framing, we added a variation in the instructions to encourage or discourage exploitation.

In the *Encourage* treatment, we reminded street vendors that they have the right to use public space as part of their economic activity, and that this right was recently ratified by the Colombian Constitutional Court. We expect that this framing increases vendors' over-exploitation because we are reminding them of their right to remain where they currently are. In the *Discourage* treatment, we reminded street vendors that local authorities must protect, recover, and reassign public space (which relates more to how local governments manage public space). We expect that this framing increases vendors' exploration (i.e., under-exploitation), because we are reminding them that local authorities are entitled to relocate them to safeguard public space. We have a *Control* group with no mention of encouragement or discouragement regulations to remain in the public space, which was the same neutral version employed in the data collection with students and public officers. We assume that the control group was closer to an optimal behavior, and that, given that the optimal rules recommend exploration for $S \leq 16$, which could be counterintuitively high considering that this is a starting stock in the poor patch, discouragement to exploit is more beneficial than encouragement. Hence, as fourth hypothesis we have:

H4. *Among street-vendors, the average earnings per treatment follow the order:*

$$\textit{Control} > \textit{Discouraging message} > \textit{Encouraging message.}$$

2.5 Implementation and payments

We programmed the instrument, including the game and a post-experimental survey, in Qualtrics. This protocol was approved by the Ethics Committee at Universidad del Rosario (ID: 609-CS378).

All the participants were paid, at the end of their interaction, in cash. For street-vendors, the average completion time was 7.7 minutes, and they earned on average 16,200 COP (std. dev. 3,974). For students, the average completion time was 9 minutes, and they earned on average 18,000 COP (std. dev. 3,633). These earnings were approximately half the daily minimum wage when the experiment was conducted.

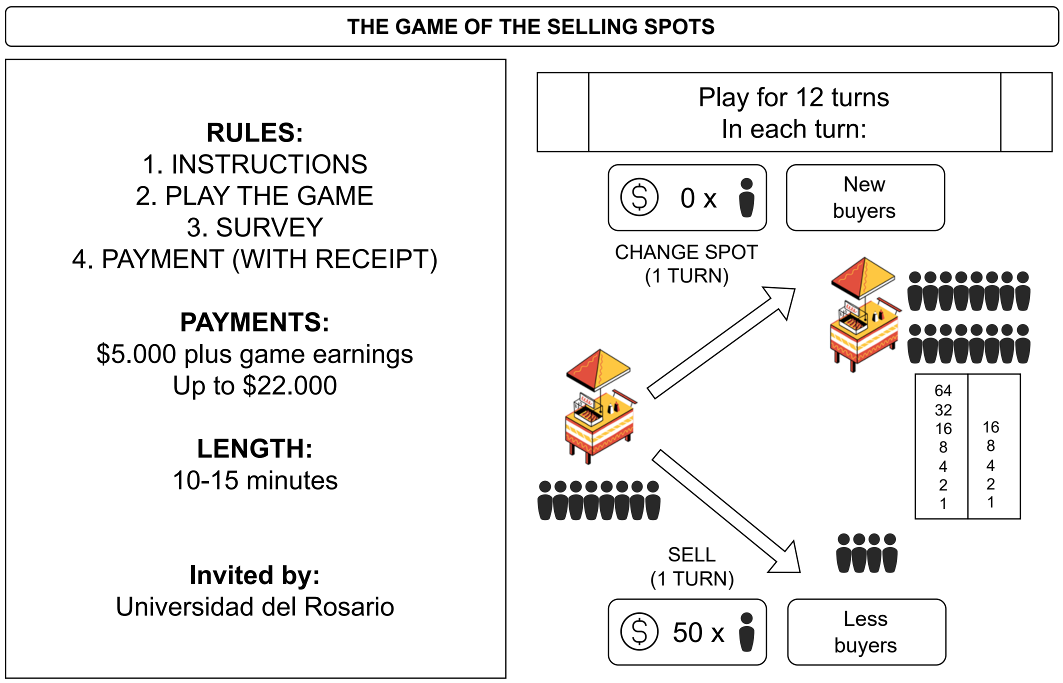

Figure 2: Flyer with the invitation to participate, summarizing the activity and the game rules.

Sample of street vendors

We conducted the study in a lab-in-the-field setting in the streets of Bogotá, Colombia, in May 2022. We invited participants located in Bogotá's Downtown by approaching them during their work journey, briefly explaining the game and the payment ranges and procedures, to be delivered after completing a survey at the end of the study. We handed potential participants a flyer, displayed in Figure 2, that summarized the procedure. An enumerator was in charge of explaining the game rules to potential participants. The left side of the flyer showed the general procedure, length and expected payment. On the right side, the flyer summarized the game incentives: sell and receive 50 COP for each available buyer, knowing that the number of buyers will be halved the next turn (i.e., to exploit); or not sell and spend the turn finding a new selling spot, with a reset in the number of buyers to 64 or 16.

Once a street vendor accepted the invitation, she was paired with an enumerator, who carried a tablet in which she shared the game instructions and decision interphases. The monitor read the instructions out loud, while the participant could follow the same written instructions (see Appendix A.4 for the translated version, and Appendix A.5 for the original instructions in Spanish). The treatment variation was introduced early: in the introduction page, right after the objective of the research study was mentioned, participants in the *Encouragement* and *Discouragement* treatments received the relevant message. Participants provided consent to participate after the instructions were read. Since participants were on their work journey, we allowed them to briefly interrupt the activity if a buyer came during the instructions or the final survey, and then continue with our study. We requested them to ask another vendor to take care of any sale occurring while they were making the choices in the game. Participants were respectful of this rule.

All the game decisions were made directly by the participants on the tablet. Afterwards, the monitor took the tablet to complete the post-experimental survey. The payment was automatically calculated in the tablet, and given to the participant at the end of the interaction.

Samples of students and public officers

We collected data from two samples of students with different nature. The first batch, of 87 students, was collected in May 2022, right after we completed the sample of street vendors. We invited students from Universidad del Rosario who were not enrolled in an Economics, Finance, or any STEM field degree to participate. We did not organize laboratory sessions but rather followed a recruitment procedure as similar as possible to the one with street vendors: we approached them in the campus' hallways, invited them to the study with flyers with the summary of instructions, and those interested were directed to the laboratory. The invitation included a QR code, so they could access to the instrument from their phones or use the laboratory's computers. The second batch, 52 students, was collected in October 2022, as part of an introductory activity to a National student contest among undergraduates from different Colombian economics departments. The day after our data collection, they participated in a disciplinary contest. Hence, this is a high-skilled group of students with preferences for competition that would be intrinsically motivated to make a good performance in the selling-spot game.

We also collected data from a small sample ($N = 22$) of public officers during a visit they made to the experimental economics laboratory, as part of a 8-session course on strategic behavior in their roles as financial inspectors. This sample included accountants, lawyers and economists with intermediate to high work experience.

3 Results

3.1 Comparison with optimality predictions

Figure 3 reports the observed round's average for exploitation, the background reward rate, and the remaining stock (i.e., number of buyers). To ease the comparison, this figure also includes—

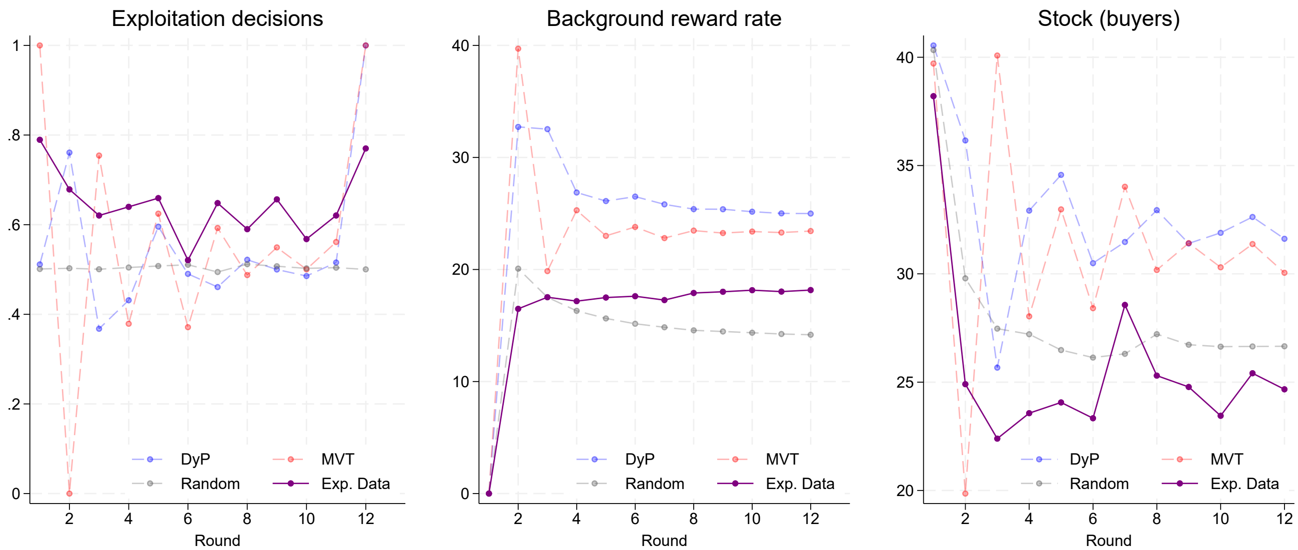

Figure 3: Observed versus predicted behavior

in a transparent layer—the predictions initially reported in Figure 1. Our main finding is the participants' tendency to over-exploit their current patch. In Table 1, we report the exploitation rates for early (1-4) and late (5-11) rounds. In rounds 5-11, after exploitation patterns stabilize, the theoretical probability to exploit fluctuates around 0.5. For the same periods, we observe a probability of 0.61, which is statistically different from 0.5 ($F(1,383) = 87.5$, $p < 0.001$). As additional evidence of over-exploitation, also reported in Table 1, cycle length is 2.7, 0.7 rounds longer than the theoretically expected value ($F(1,383) = 166.62$, $p < 0.001$).

This over-exploitation leads to a background reward rate–a measure of average earnings–that is lower than the level achieved under the optimality rules but higher than that generated by a random decision-maker. Taking again as reference the rounds 5-11, the observed BRR is 17.96. This value is statistically different from 24 ($F(1,383) = 325.8$, $p < 0.001$), the expected BRR under optimal rules; and also different from 16 ($F(1,383) = 34.7$, $p < 0.001$), a reference value that the random-decision making does not exceed after round 5. The gap in average earnings is larger with respect to the optimality rules because over-exploitation entails two costs: a static cost, by having periods with very low earnings; and a dynamic cost, because participants do not reset their stock often enough.

Regarding the stock levels, the rightmost panel shows that the observed stock remains at low levels, even below the simulated predictions for a random decision-maker. The observed average

Table 1: Extraction rate, cycle lenghts, and frequency of suboptimal behavior according to the DyP and MVT rules.

	Rounds				
	All	Early	Late	First	Last
	1-12	1-4	5-11	1	12
Exploitation rate	0.65	0.68	0.61	0.79	0.77
Cycles (cumulative)	4.22	1.27	3.99	.	.
Mean cycle length	2.70	2.44	2.67	.	.
Suboptimal behavior					
DyP: over-exploitation	0.53	0.57	0.49	0.68	.
DyP: under-exploitation	0.15	0.14	0.17	0.08	.
MVT: over-exploitation	0.48	0.53	0.44	.	.
MVT: under-exploitation	0.24	0.30	0.23	.	.

earnings are larger than for a random decision-maker, but the stock remains below their levels, because participants are biased towards one type of error: over-exploitation. By contrast, they rarely forgo the earnings from a rich patch ($S = 64$ or $S = 32$) by leaving it unexploited, as could happen to a random decision-maker.

The frequency of suboptimal decisions according to both models, as well as the exploitation rates and number of cycles and their length, are reported in Table 1. Taking the *DyP* rule as a reference, over- and under-exploitation occur 53% and 15% of the time, respectively. There seems to be some learning that reduces over-exploitation, as late in the game this pattern drops to 49% (diff. -0.08 percentage points–pp, $F(1, 383) = 27.01$, $p < 0.001$). By contrast, the difference in the under-exploitation rate between early and late rounds is not statistically significant (0.029 pp, $F(1, 383) = 2.66$, $p = 0.103$). Using the *MVT* rule, over- and under-exploitation occur 48% and 24% of the time, respectively. For this rule, there is also some learning reducing both over- and under-exploitation (-0.097 pp, $F(1, 383) = 17.20$, $p < 0.001$; and -0.067 pp, $F(1, 383) = 15.69$, $p < 0.001$).

The lower over-exploitation but higher under-exploitation under *MVT* relative to *DyP* in Table 1 follows from the endogeneity of the MVT threshold. Since BRR_t is built from each participant's own decision history, and over-exploitation is the dominant bias in our sample, repeated low-stock exploitation progressively lowers BRR_t below the *DyP* threshold. This downward-biased benchmark has two opposing effects: it reclassifies some genuinely excessive exploitation (e.g., at $S_t = 16$) as *MVT*-consistent, thereby reducing measured over-exploitation; however, it also classifies exploration decisions that are optimal under the fixed *DyP* rule as under-exploitation once BRR_t falls below 16.

Finally, note that exploitation rates are larger in the first (0.79) and last (0.77) rounds. An exploitation rate in round 1 above 0.5 favors the *MVT* over the *DyP* rule: it remarks that, in absence of any information retrieved from experience, the recommended first action is to exploit. On the other hand, an exploitation rate below 1.0 in the last round suggests lack of attention or confusion when facing the game's terminal round.

Coming back to the hypotheses linked to the optimality rules, we reject **H1**, given our evidence of a considerable over-exploitation rate in late rounds (H1A) and that the BRR falls below 24 at the end of the game (H1B).

Result 1: *Participants systematically over-exploit their selling locations, according to the DyP and MVT optimality rules. Over-exploitation is between 2 and 3.5 times more prevalent than under-exploitation.*

On the other hand, **H2** aims at detecting which optimality rules, between the DyP and the MVT frameworks, describe early behavior (i.e., before these rules start converging) more accurately. H2A identifies actions that are considered mistakes under DyP but not under MVT: exploring with $S_t = 32$ and exploiting with $S_t = 16$, given that the game history influences the BRR. We find that, in rounds 1-4 and for $S_t = 32$, exploration occurs in 23% of the cases. Moreover, for the same rounds and $S_t = 16$, the exploitation rate is 63%. Hence, the tendency to over-exploit appears to favor the MVT model.

However, H2B proposes that, if participants follow the MVT rule, regardless of the starting patch, 64 or 16, the first two rounds follow the pattern exploit-explore. If this is true, then the average cycle length should not differ early in the game (rounds 1-4), on average, for those starting in a rich or a poor patch. We find that the cycle length in rounds 1-4 is 0.52 rounds longer for those starting in $S_1 = 64$ ($t = 5.19$, $p < 0.001$, see Table A.1 in the Appendix). Since we reject H2B, and the systematic tendency to over-exploit, we cannot say that participants are better described by the MVT than by the DyP rule, and **H2** is therefore rejected.

Result 2: *The systematic over-exploitation prevents that one optimality rule outperforms the other.*

Heterogeneity in earnings and in deviations from optimality

The left panel in Figure 4 reports the cumulative distribution of earnings for the three prediction models and the observed behavior. As hinted by the BRR, the distribution of observed earnings dwells in-between the distributions under the optimal rules and random behavior. The variation in earnings does not emerge uniquely from the random variation in the landing patches, but also from variation in the decisions characterized as suboptimal according to the DyP and MVT rules. Between rounds 2 and 11 (we omit round 1 because the BRR cannot be computed), the average number of decisions classified as suboptimal, which we will call errors for brevity, is 3.78 (std. dev. 2.13) under the DyP rule, and 3.42 (std. dev. 1.84) under the MVT rule.

To explore the persistence of under- and over-exploitation decisions, the right panel in Figure 4 displays the distributions of a signed error count computed for each participant: each under-exploitation decision adds -1 and each over-exploitation decision adds 1. The upper and bottom panels correspond to signed errors under the DyP and MVT rules, respectively. In each histogram, the lighter filling represents negative counts and indicates participants who under-exploited more often than over-exploited; whereas the darker filling corresponds to the opposite pattern of more over-exploitation. The histograms reveal that, although over-exploitation was already identified as a dominant pattern, there is substantial variation in the participants' tendency to do this, but

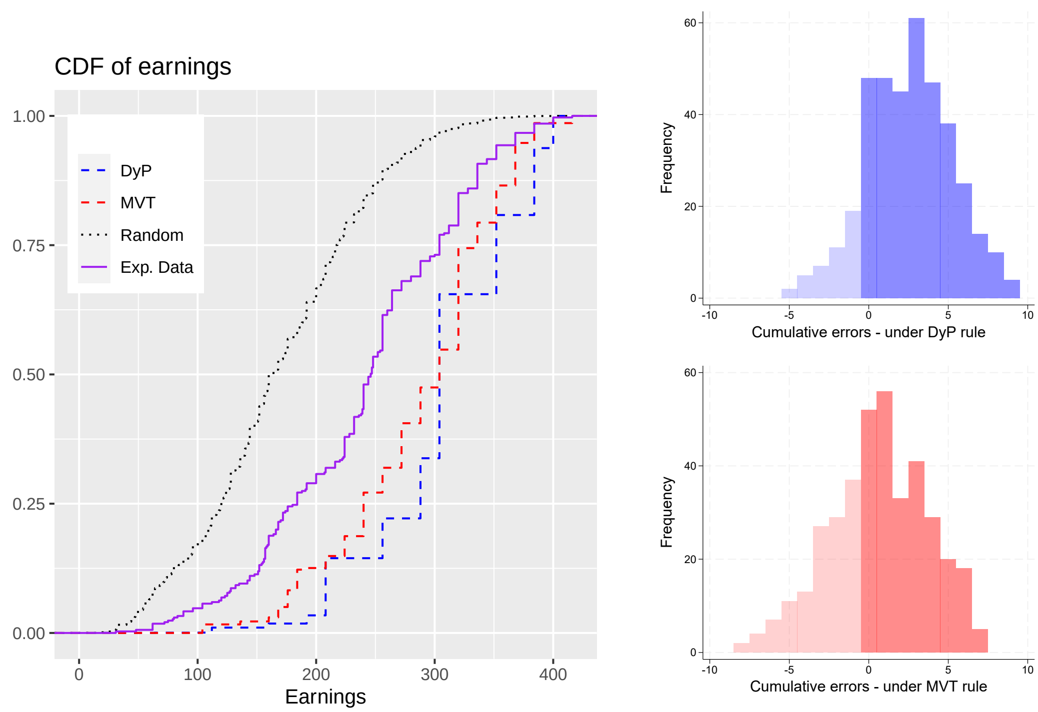

Figure 4: Evidence of participant's heterogeneity. ***Left panel:*** Cumulative distribution of earnings. Comparison of simulated *versus* observed outcomes. ***Right panels:*** Histogram of cumulative deviation from optimality per player. Under-exploitation adds -1, and over-exploitation adds 1. Top-right corresponds to DyP, bottom-right to MVT.

also a non-negligible proportion of participants systematically under-exploiting the patch.

3.2 Comparisons between samples

Figure 5 reports the differences in earnings and errors (according to the two optimality rules) between the samples that participated in our study.

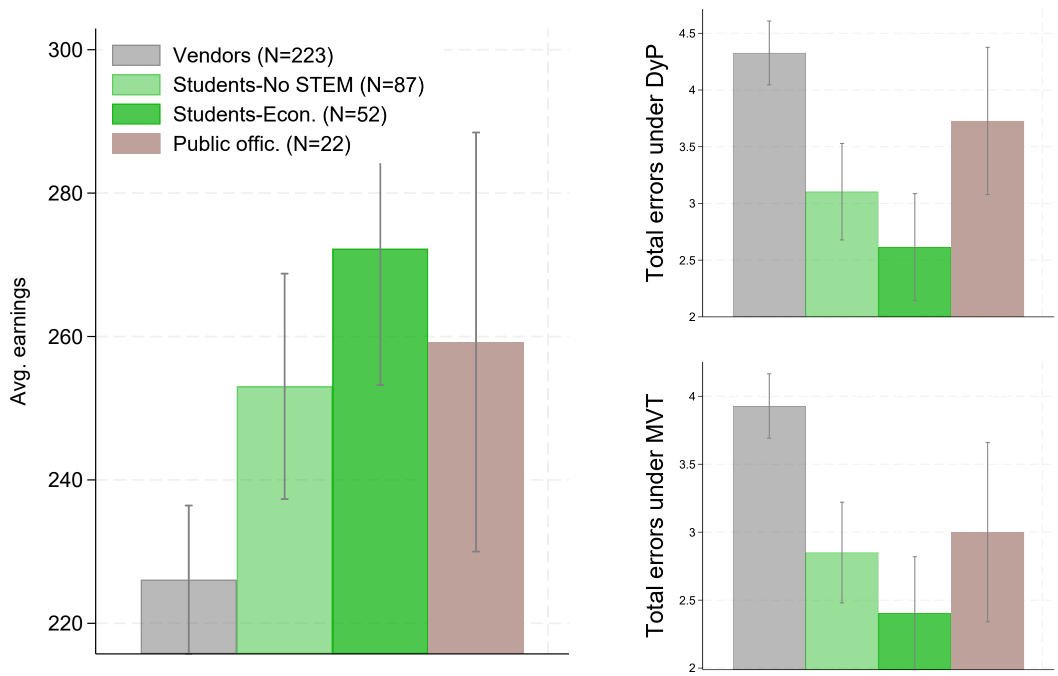

Figure 5: Participants' earnings (left panel) and frequency of total errors by participant (right panels) across samples. Total errors computed as sum of deviations per participant according to the DyP (top right) and MVT (bottom right) optimality rules.

Regarding earnings, street vendors accumulate, on average, 226 units (std. dev. 78.6) throughout the game. Non-STEM students accumulate 253 units (std. dev. 74.6), and this difference is statistically significant ($t-$test, $p = 0.006$). The largest earnings correspond to the selected Economics undergrad students who were attending a contest, with 272 units (std. dev. 69.7), also significantly larger than for vendors ($t-$test, $p < 0.001$). Finally, in the sample of public officers the average earnings are 259 units (std. dev. 69.7). However, given the small sample size ($N = 22$), this difference is also statistically significant with respect to vendors ($t-$test, $p = 0.058$). By contrast, none of the differences between students and public officers is statistically significant.

Regarding errors, classified according to the *DyP* optimal rules, vendors have the largest count (4.33, std. dev. 2.13), followed by public officers (3.72, std. dev. 1.54), No-STEM students (3.10, std. dev. 2.02), and Economics students (2.62, std. dev. 1.72). The difference between vendors and the separate samples of students are statistically significant ($p < 0.001$ in both comparisons), and between Economics students and public officers ($p = 0.011$). The pattern is very similar when errors are defined as deviations from the *MVT*.

To test **H3**, we complement the statistical comparisons above with OLS regressions that pool all non-vendors into a single category and compare them with vendors, using earnings and the total number of exploration decisions as the dependent variables. To investigate additional mechanisms underlying differences in earnings, we include two plausibly exogenous variables capturing variation in the environment: whether the participant started the game in a rich patch ($S_1 = 64$) and the share of rich patches encountered during the game (i.e., the proportion of rounds in which the reset stock was 64, including Round 1). We are particularly interested in the interaction between these variables, which capture exogenous variation in the returns to exploitation, and the vendor indicator, as this may reveal an additional mechanism besides the higher error rates reported above and in Table A.2: vendors may take less advantage of more abundant environments.

Table 2 displays the regression outputs, with the first three columns remarking the lower vendors' earnings. Column 1 shows that vendors' earnings are (34/260=) 13% lower. Column 2 shows, as expected, that landing first in a rich patch increases earnings. The interaction term is negative but non-significant, meaning that vendors are not profiting less (with respect to the rest of the sample) from an initial rich patch. In column 3 we observe that a higher share of rich patches increases earnings. More importantly, its interaction with the vendor categorical variable is negative and significant, and the vendor dummy alone is no longer significant. The lack of differences between samples in the first patch, followed by an increasing gap in the ability to exploit future rich patches, suggests that non-vendors adjust and learn more rapidly to play in rich environments, or that they remain more attentive during the game.

Columns 4 and 5 show that the number of exploration decisions is not statistically different between vendors and non-vendors, neither the interaction term. That is, vendors are equally likely to stay for too long in their initial patch as non-vendors. The apparent tension between this null result and the higher prevalence of over-exploitation documented in Table A.2 is resolved by the direction and magnitude of the coefficients: the negative point estimate (-0.25) is consistent with vendors exploring somewhat less often, but since vendors commit only between 0.7 and 0.9 additional over-exploitation errors relative to non-vendors, this difference does not appear large enough to speak of genuinely different exploration patterns.

Thus, we find support for **H3** and conclude:

Table 2: OLS regressions for earnings and number of cycles across sample types

	Earnings			Exploration decisions	
	(1)	(2)	(3)	(4)	(5)
Vendor	-34.01***	-24.76**	-13.88	-0.25	-0.20
	(7.87)	(10.22)	(12.36)	(0.17)	(0.22)
First patch is rich		60.00***			-0.50**
		(11.31)			(0.25)
Vendor × First patch is rich		-14.97			-0.16
		(14.87)			(0.33)
Share of rich patches			229.50***		
			(17.89)		
Vendor × Share of rich patches			-51.31**		
			(22.21)		
Constant	260.09***	230.65***	148.92***	4.37***	4.61***
	(6.00)	(7.92)	(9.69)	(0.13)	(0.17)
Observations	384	384	384	384	384

Note: Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Result 3: *Street vendors earn about 13% less than a reference sample (with mostly students). These differences are partly explained by some learning about how to play in repeated rich patches, and not by a systematic difference in exploration decisions.*

3.3 Comparisons between treatments

Recall that, in the sample of street vendors, we added two exogenous variations to the initial framing. The *Encourage* treatment aims at inducing over-exploitation, whereas the *Discourage* treatment aims at inducing under-exploitation. There is a third group with neutral framing serving as control. Table 3 reports the results of an OLS regression comparing earnings across treatments. We use a specification similar to the one described in Section 3.2, including the indicator of an initial rich patch and the share of rich patches. We are interested in their interactions with the treatment variables, *Encourage* and *Discourage*.

Column 1 in Table 3 shows that the *Discourage* treatment increases earnings by 23.5 units. This coefficient is marginally significant ($p = 0.072$). By contrast, the coefficient for *Encourage* is positive but statistically not different from zero. One explanation for the benefits on earnings

Table 3: OLS regression for earnings across treatments in street vendors' sample

Dependent variable: Earnings	(1)	(2)	(3)
Encourage	11.86	-7.91	-7.37
	(12.85)	(16.32)	(20.04)
Discourage	23.52*	11.82	-26.75
	(13.02)	(16.84)	(21.13)
First patch is rich		22.34	
		(18.16)	
Encourage × First patch is rich		42.05*	
		(24.96)	
Discourage × First patch is rich		20.48	
		(25.22)	
Share of rich patches			143.61***
			(26.24)
Encourage × Share of rich patches			23.76
			(35.22)
Discourage × Share of rich patches			76.30**
			(36.79)
Constant	214.13***	205.00***	146.23***
	(9.30)	(11.61)	(14.33)
Observations	223	223	223
R-squared	0.01	0.10	0.42

Note: Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

from discouragement is the reported pattern of over-exploitation. Since the most prevalent error is exploiting too much, a contextual message that discourages this behavior may be helpful.

In Column 2, we interact the treatment variables with the indicator of an initial rich patch. Its interaction term with *Encourage* is positive and marginally significant ($p = 0.094$), suggesting that the messages inducing more exploitation could hint participants to do so in the first round. Finally, Column 3 suggests that the *Discourage* treatment helps participants in profiting more from rich patches (coefficient 76.3, $p = 0.039$). The interaction terms in this table should be interpreted cautiously, especially because we do not observe treatment effects interacting with the number of errors using any of the optimality rules (see Table A.3). We attribute this null result to limited statistical power rather than to the absence of an underlying relationship between our treatments and exploitation behavior.[7]

[7] The sample of street vendors comprises only 223 participants split across three treatment arms, and the standard errors on the *Discourage* × *Share of rich patches* interaction across the four error specifications (1.08–1.33) are of comparable magnitude to the point estimates themselves (-1.39 to 0.38).

In **H4**, we hypothesized that the *Encourage* message would yield the lowest earnings, followed by the *Discourage* treatment, and the higher earnings would be obtained in the control condition. This prediction was based on the assumption that treatment messages would deviate participants from a more optimal behavior. However, we find non-conclusive evidence that the *Discourage* treatment yields the largest earnings. We thus reject **H4** and argue that this finding is explained by the participants' over-exploitation pattern: messages preventing this pattern can improve earnings.

Result 4: *Messages discouraging exploitation increase earnings with respect to the control condition and to messages of encouragement.*

4 Concluding discussion

We designed a new experimental game that extends the canonical bandit problem by making the value of exploration resource-based rather than informational. Relocating replenishes the stock of buyers by drawing a new location from a known distribution, leaving beliefs about the environment unchanged, whereas exploitation endogenously depletes future payoffs. This shifts the trade-off from balancing immediate rewards against the value of information to balancing current extraction against resource depletion, allowing us to define and quantify two distinct deviations from optimal behavior: over- and under-exploitation.

By recasting the problem in this way, we connect the patch-foraging paradigm–until now largely confined to biology, ecology, and the decision sciences–to experimental economics. The experimental approach is well suited to this setting because it allows us to evaluate behavior against clear normative benchmarks: two solution concepts with very different informational and cognitive demands nevertheless prescribe nearly identical behavior. Dynamic programming, which yields a forward-looking optimal policy, and the marginal value theorem, a history-based heuristic grounded in realized payoffs. Their convergence to a common exploitation threshold implies that the over-exploitation we document reflects a genuine departure from optimality rather than an artifact of the benchmark used. In the short run, when predictions differ, we do not find conclusive evidence favoring one rule over the other.

We took the game to the field using a framing close to the daily activity of our main subsample, street vendors: a vending cart with the choice between selling at the current spot and relocating. Vendors are more likely than our (mostly student) reference sample to deviate from the optimal rules, and these errors translate into lower earnings. The dominant deviation is over-exploitation: staying too long in a depleting patch is roughly two to three times more frequent than leaving it too soon. Framing messages tailored to the vendor sample also appear to shape earnings. Because over-exploitation is the prevailing error, messages that discourage staying in a selling spot outperform those that encourage it: reminding vendors that authorities may protect and reassign public space marginally raised earnings, whereas reminding vendors of their right to remain did not. The asymmetric effect of our treatments is itself consistent with the prevalence of over-exploitation.

Introducing this paradigm into experimental economics opens at least two avenues for future work. First, the environment connects naturally to environmental and resource economics: this dynamic problem can be reframed to study intertemporal resource management–isolated from the social dilemma layer–to study technology adoption. Second, the simplicity of the game–despite its elaborated solution–makes it a useful vehicle for studying how participants transmit advice or recommendations to future players. This last direction speaks to the growing literature on complexity, by examining which rules of thumb people distill from experience and choose to pass on (Arrieta & Nielsen, 2024).

Data availability statement

The replication material for the study is available at the Open Science Framework: https://osf.io/p5ab4/overview?view_only=530c6dcb1fa547d49531f9a725521599

References

Abreu, D., & Rubinstein, A. (1988). The structure of Nash equilibrium in repeated games with finite automata. *Econometrica: Journal of the Econometric Society*, 1259–1281.

Alaoui, L., & Penta, A. (2022). Cost-benefit analysis in reasoning. *Journal of Political Economy*, *130*(4), 881–925.

Arrieta, G. R., & Nielsen, K. (2024). *Procedural decision-making in the face of complexity* [Working paper].

Banks, J., Olson, M., & Porter, D. (1997). An experimental analysis of the bandit problem. *Economic Theory*, *10*(1), 55–77.

Banks, J. S., & Sundaram, R. K. (1992). Denumerable-armed bandits. *Econometrica: Journal of the Econometric Society*, 1071–1096.

Banovetz, J., & Oprea, R. (2023). Complexity and procedural choice. *American Economic Journal: Microeconomics*, *15*(2), 384–413.

Bellman, R. E., & Dreyfus, S. E. (2015). *Applied dynamic programming* (Vol. 2050). Princeton University Press.

Bergemann, D., & Välimäki, J. (2001). Stationary multi-choice bandit problems. *Journal of Economic Dynamics and Control*, *25*(10), 1585–1594.

Bergemann, D., & Välimäki, J. (2018). Bandit problems. In *The new Palgrave dictionary of economics* (pp. 665–670). Springer.

Bolton, P., & Harris, C. (1999). Strategic experimentation. *Econometrica: Journal of the Econometric Society*, *67*(2), 349–374.

Brezzi, M., & Lai, T. L. (2002). Optimal learning and experimentation in bandit problems. *Journal of Economic Dynamics and Control*, *27*(1), 87–108.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129–136.

de Clippel, G., & Rozen, K. (2024). Bounded rationality in choice theory: A survey. *Journal of Economic Literature*, *62*(3), 995–1039.

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences*, *120*(13), e2216524120.

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2010). The central executive as a search process: Priming exploration and exploitation across domains. *Journal of Experimental Psychology: General*, *139*(4), 590.

Hudja, S., & Woods, D. (2024). Exploration versus exploitation: A laboratory test of the single-agent exponential bandit model. *Economic Inquiry*, *62*(1), 267–286.

Hudja, S., & Woods, D. (2025). Strategies in the multi-armed bandit. *Experimental Economics*, *28*(4), 974–996. https://doi.org/10.1017/eec.2025.10027

Hudja, S., & Woods, D. (2026). Examining experimentation: An elicitation approach. *Available at SSRN 5346974*.

Hudja, S., Woods, D., & Gately, J. B. (2026). Forced experimentation: An analysis of multi-armed bandits with an inconsistent option. *Journal of Economic Behavior & Organization*, *243*, 107434.

Krebs, J. R., Ryan, J. C., & Charnov, E. L. (1974). Hunting by expectation or optimal foraging? a study of patch use by chickadees. *Animal Behaviour*, *22*, 953–IN3.

Lizzeri, A., Shmaya, E., & Yariv, L. (2024). *Disentangling exploration from exploitation* (tech. rep.). National Bureau of Economic Research.

Lloyd, A., McKay, R. T., & Furl, N. (2022). Individuals with adverse childhood experiences explore less and underweight reward feedback. *Proceedings of the National Academy of Sciences*, *119*(4), e2109373119.

Lloyd, A., Viding, E., McKay, R., & Furl, N. (2023). Understanding patch foraging strategies across development. *Trends in Cognitive Sciences*, *27*(11), 1085–1098.

Malladi, S., Martinez Marquina, A., & Morozov, I. (2025). Space exploration. *Available at SSRN 5193781*.

Manning, A., & Petrongolo, B. (2017). How local are labor markets? evidence from a spatial job search model. *American Economic Review*, *107*(10), 2877–2907.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, *84*(1), 113–126.

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy*, *92*(6), 1086–1120.

Mortensen, D. T. (1986). Job search and labor market analysis. *Handbook of Labor Economics*, *2*, 849–919.

Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. *Economics Letters*, *19*(3), 227–229.

O'Bryan, L. R., Lambeth, S. P., Schapiro, S. J., & Wilson, M. L. (2020). Unpacking chimpanzee (Pan troglodytes) patch use: Do individuals respond to food patches as predicted by the marginal value theorem? *American Journal of Primatology*, *82*(12), e23208.

Oprea, R. (2020). What makes a rule complex? *American Economic Review*, *110*(12), 3913–3951.

Oprea, R. (2024). Decisions under risk are decisions under complexity. *American Economic Review*, *114*(12), 3789–3811.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*(5), 527–536. https://doi.org/10.1090/S0002-9904-1952-09620-8

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory*, *9*(2), 185–202.

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory*, *39*(1), 83–96.

Salant, Y. (2011). Procedural analysis of choice rules with applications to bounded rationality. *American Economic Review*, *101*(2), 724–748.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 99–118.

Stabentheiner, A., & Kovac, H. (2016). Honeybee economics: Optimisation of foraging in a variable world. *Scientific Reports*, *6*(1), 28339.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, *69*(3), 213–225.

Urgun, C., & Yariv, L. (2025). Contiguous search: Exploration and ambition on uncharted terrain. *Journal of Political Economy*, *133*(2), 522–567.

A Appendix

A.1 Additional Tables and Figures

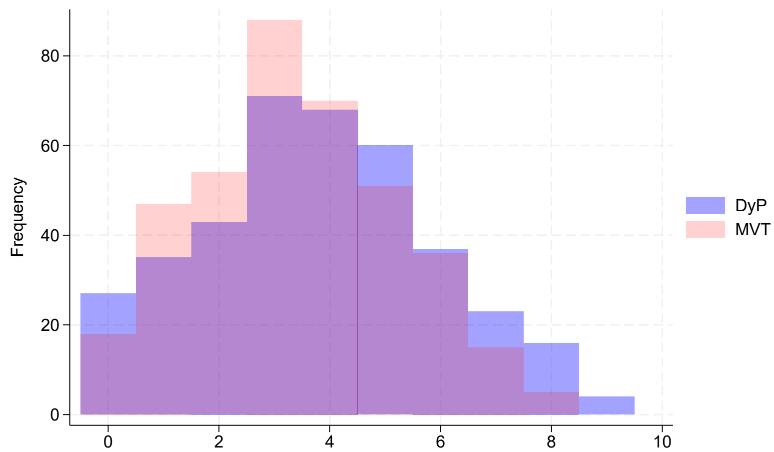

Figure A.1: Distribution of actions in rounds 2-11 classified as suboptimal (errors) under the DyP (in purple) and the MVT (in pink) rules.

Table A.1: Effect of lagged stock on cycle duration (Rounds 1-4).

Dep. Var: Mean cycle length	Coeff.	SE
Stock= 64 in Round 1	0.522^{***}	(0.101)
Constant	2.193^{***}	(0.063)
Observations	384	

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Deviations (errors) from the DyP and MVT optimality rules.

	DyP Rule		MVT Rule	
	(1) Total errors	(2) Over-exploit	(3) Total errors	(4) Over-exploit
Vendor	1.30***	0.90***	1.20***	0.69***
	(0.21)	(0.22)	(0.18)	(0.19)
Constant	3.03***	2.63***	2.73***	1.61***
	(0.16)	(0.17)	(0.14)	(0.15)
Observations	384	384	384	384
R-squared	0.09	0.04	0.10	0.03

Note: Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: OLS regressions for errors across treatments in street vendors' sample.

	DyP: Total errors		DyP: Over-exploit		MVT: Total errors		MVT: Over-exploit	
	(1) First patch	(2) Share rich	(3) First patch	(4) Share rich	(5) First patch	(6) Share rich	(7) First patch	(8) Share rich
Encourage	0.01	-0.17	-0.01	-0.18	0.29	0.25	0.10	-0.21
	(0.46)	(0.67)	(0.51)	(0.72)	(0.39)	(0.59)	(0.42)	(0.68)
Discourage	-0.64	0.21	-0.43	0.52	-0.19	0.11	-0.24	-0.31
	(0.48)	(0.70)	(0.52)	(0.76)	(0.40)	(0.62)	(0.43)	(0.72)
First patch is rich	0.41		0.38		0.86**		1.73***	
	(0.52)		(0.57)		(0.43)		(0.47)	
Encourage ×	-0.79		-0.70		-0.92		-0.78	
First patch is rich	(0.71)		(0.78)		(0.59)		(0.64)	
Discourage ×	0.14		0.02		-0.25		-0.10	
First patch is rich	(0.72)		(0.79)		(0.60)		(0.65)	
Share of rich patches		-2.26**		-2.56***		-1.06		-0.77
		(0.87)		(0.95)		(0.77)		(0.89)
Encourage ×		-0.09		-0.02		-0.56		0.13
Share of rich patches		(1.17)		(1.27)		(1.03)		(1.20)
Discourage ×		-1.11		-1.39		-0.52		0.38
Share of rich patches		(1.22)		(1.33)		(1.08)		(1.25)
Constant	4.45***	5.69***	3.62***	4.99***	3.69***	4.54***	1.71***	2.79***
	(0.33)	(0.48)	(0.36)	(0.52)	(0.28)	(0.42)	(0.30)	(0.49)
Observations	223	223	223	223	223	223	223	223
R-squared	0.02	0.13	0.01	0.14	0.03	0.05	0.13	0.01

Note: Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.2 Solution to the dynamic optimization problem

A.2.1 Solution to the Bellman equation with an infinite horizon

Although our setting has a finite horizon of 12 rounds, the infinite-horizon solution provides useful intuition and anticipates the key result: the threshold stock level S^* converges to 24 as $\beta \to 1$.

We begin by recalling the value function in Eq. 2 (Section 2.2), which takes the form:

$$V_t(S_t) = \max\left(S_t + \beta V_t\left(\frac{S_t}{2}\right), \beta\left(\frac{V_t(16) + V_t(64)}{2}\right)\right)$$

The first argument captures the value of exploitation: current payoffs (equal to the stock of buyers S_t) plus the discounted continuation value from the remaining stock $S_t/2$. The second argument captures the value of exploration. It is multiplied by β because exploring forgoes current payoffs in exchange for a stochastic reset of the stock to either 64 or 16 buyers with equal probability. We can write this equation in terms of the stock level S^* that makes the agent indifferent between exploitation and exploration: any $S_t \geq S^*$ leads to exploitation, while any $S_t < S^*$ leads to exploration. This condition yields:

$$S^* + \beta V\left(\frac{S^*}{2}\right) = \frac{\beta V(16) + \beta V(64)}{2} \tag{3}$$

By construction, the stock follows the recursion $S_{t+1} = S_t/2$ under exploitation, so $S_t \in \{64, 32, 16, 8, ...\}$. Hence, we define rules on the minimum stock level required for exploitation, starting from the maximum value ($S_t = 64$), and verify whether each rule is internally consistent (i.e., if given the value functions it implies, no agent would deviate from the prescribed action at any stock level).

We start with $S^* > 32$, meaning that only $S_t = 64$ leads to exploitation, and any $S_t \leq 32$ triggers exploration. Hence, $S_t = 32$ and $S_t = 16$ share the same continuation value:

$$V(32) = V(16) = \beta\left(\frac{V(16) + V(64)}{2}\right),$$

and therefore $V(64) = 64 + \beta V(32) = 64 + \beta V(16)$. Substituting $V(64)$ into the expression for $V(16)$ and solving:

$$V(16) = \frac{64\beta}{2 - \beta - \beta^2} = \frac{64\beta}{(2+\beta)(1-\beta)}$$

For this case to be consistent, exploiting at $S_t = 32$ must be suboptimal. That is, $32 + \beta V(16) < V(16)$, which simplifies to:

$$32 < (1-\beta)V(16) = \frac{64\beta}{2+\beta}$$

Since $(64\beta)/(2+\beta) < 32$ for all $\beta \in (0,1)$, the consistency condition is never satisfied. Hence, exploiting at 32 is profitable and $S^* < 32$. The next scenario to check is whether $S^* > 16$, meaning that $S_t = 64$ and $S_t = 32$ lead to exploitation, while $S_t = 16$ triggers exploration.

If we now depart from $S^* > 16$, $V(64)$ unfolds recursively as:

$$V(64) = 64 + \beta V(32) = 64 + 32\beta + \beta^2 V(16) \tag{4}$$

Since $S_t = 16$ triggers exploration, $V(16)$ satisfies the fixed point condition:

$$V(16) = \beta\left(\frac{V(16)+V(64)}{2}\right)$$

Substituting Eq. 4 and rearranging $V(16)(2-\beta-\beta^3) = \beta(64+32\beta)$ we obtain:

$$V(16) = \frac{\beta(64+32\beta)}{2-\beta(1+\beta^2)} \tag{5}$$

For this case to be consistent, exploring at $S_t = 16$ must be optimal. This means that $V(16) > 16 + \beta V(8)$. Since the exploration value $V(16) = \beta[V(16)+V(64)]/2$ is independent of the current stock, exploration is optimal for all $S_t \leq 16$, and hence $V(8) = V(16)$. The consistency condition then becomes $V(16) > 16 + \beta V(16)$, which can be rearranged as $(1-\beta)V(16) > 16$. Note that, from the indifference condition Eq. 3 and the fact that $V(S^*/2) = V(16)$ for any $S^* \in (16, 32)$ because both S^* and $S_t = 16$ trigger exploration, we have:

$$S^* + \beta V(16) = \beta\left(\frac{V(16)+V(64)}{2}\right) = V(16)$$

Hence $S^* = (1-\beta)V(16)$, and the consistency condition reduces to $S^* > 16$. Substituting Eq. 5 in this expression:

$$S^* = (1-\beta)V(16) = \frac{\beta(1-\beta)(64+32\beta)}{2-\beta(1+\beta^2)} \tag{6}$$

This case is therefore consistent whenever $S^* > 16$, which holds for all $\beta > \beta^*$ where $\beta^* \approx 0.56$. Since our experimental setting features patient agents given the short duration of the whole game (i.e., $\beta \to 1$), then we can work with $S^* \in (16, 32)$.

Finally, for a more precise solution as $\beta \to 1$, note that in Eq. 6, both the numerator and denominator vanish as $\beta \to 1$, so we apply l'Hôpital's rule:

$$\lim_{\beta\to 1} S^* = \lim_{\beta\to 1} \frac{\frac{d}{d\beta}\left[\beta(1-\beta)(64+32\beta)\right]}{\frac{d}{d\beta}\left[2-\beta(1+\beta^2)\right]} = \lim_{\beta\to 1} \frac{64(1-2\beta)+32(2\beta-3\beta^2)}{-(1+3\beta^2)} = \frac{-64-32}{-4} = 24$$

Hence, $S^* \to 24$ as $\beta \to 1$. In the infinitely repeated game, agents exploit whenever their stock exceeds 24, and explore otherwise.

A.2.2 Solution with a finite horizon

When the problem has a terminal round T, we find its solution by backward induction. As we show below, the optimal policy stabilizes after just two periods: for any $k \geq 1$, the agent exploits $S \geq 32$ and explores $S \leq 16$, consistent with the infinite horizon solution.

Let us start with the terminal round. Since no further rounds follow, exploration yields neither immediate profit nor future benefit. The agent simply extracts the current stock, so the value function is $V_T(S) = S$, with threshold $S^* = 0$.

One period earlier ($k = 1$), the Bellman equation is:

$$V_{T-1}(S) = \max\left\{S + \beta\left(\frac{S}{2}\right), 40\beta\right\}$$

The value of exploitation corresponds to the earnings from repeating this strategy over the last two periods, $S + \beta S/2$, while the value of exploration, 40β, is the discounted expected value of the reset patch in the terminal round. As in the infinite horizon case, we find S^*_{T-k} by equating the two arguments of $V_{T-k}(S)$. This yields:

$$S^*_{T-1} = \frac{80\beta}{2+\beta}$$

Note that, for $\beta > 1/2$, then $S^*_{T-1} > 16$ and exploitation remains optimal for stock levels of 32 or above. Given our design, it is plausible to assume that $\beta \to 1$. Hence, $S^*_{T-1} = 80/3 \approx 26.7$, close to the infinite horizon solution.

Moving to $k = 2$, the Bellman equation is:

$$V_{T-2}(S) = \max\left\{S + \beta V_{T-1}\left(\frac{S}{2}\right), \frac{\beta\left[V_{T-1}(16) + V_{T-1}(64)\right]}{2}\right\}$$

For large β, we substitute $V_{T-1}(16) = 40\beta$ (exploration) and $V_{T-1}(64) = 64 + 32\beta$ (exploitation), which simplifies the exploration value to $\beta(32 + 36\beta)$. Evaluating the first argument at each relevant stock level:

$$
\begin{aligned}
V_{T-2}(64) &= \max\left\{64 + \beta(32 + 16\beta),\ \beta(32 + 36\beta)\right\} \\
V_{T-2}(32) &= \max\left\{32 + 40\beta^2,\ \beta(32 + 36\beta)\right\} \\
V_{T-2}(S \leq 16) &= \max\left\{S + 40\beta^2,\ \beta(32 + 36\beta)\right\}
\end{aligned}
$$

As $\beta \to 1$, the optimal decision is to exploit for $S \geq 32$, and to explore for $S \leq 16$. Going back two periods with respect to the terminal round is sufficient to characterize the cyclic structure of the solution: an agent will, at most, exploit two consecutive times. This will occur when landing on $S = 64$, otherwise the agent will immediately explore.

This cyclic structure implies that the optimal policy is stationary for $k \geq 1$: the threshold S^*_{T-k} satisfies $S^*_{T-k} \geq 16$ for all $k \geq 1$ as $\beta \to 1$, so the agent always exploits $S \geq 32$ and explores $S \leq 16$. This can be verified recursively: if the optimal policy at step k is to exploit $S \geq 32$ and explore $S \leq 16$, then the exploration value at step $k+1$ is $\beta[V_{T-k}(16) + V_{T-k}(64)]/2$, and the exploitation values $V_{T-k}(32)$ and $V_{T-k}(64)$ always exceed it, preserving the same policy at step $k+1$.

A.3 Examples of the MVT rule

We present four examples of the MVT rule in operation over the first four rounds, covering different sequences of patch types. The examples illustrate when the *MVT* rule agrees with the *DyP* rule after some rounds, and when it does not. As we show below, disagreements arise when exploration repeatedly yields the same patch type, preventing the agent from acquiring new environmental information.

Example 1. *Starting in a poor patch followed by a rich one.*

The agent begins with no prior earnings, so she exploits in round 1 despite the low stock, yielding $S_2 = 8$ and $BRR_2 = 16$. Since $S_2 < BRR_2$, she explores in round 2, earning nothing. The background reward rate is updated to $BRR_3 = 16/2 = 8$. If exploration leads to $S_3 = 64$, she exploits ($S_3 \geq BRR_3$), bringing cumulative earnings to 80 and updating $BRR_4 = 80/3 \approx 26.7$. Since $S_4 = 32 \geq BRR_4$, she exploits again in round 4. In this case, the *MVT* rule coincides with the *DyP* rule throughout: the poor patch is exploited once, followed by two consecutive exploitations of the rich patch.

Example 2. *Starting in a rich patch followed by a poor one.*

The agent begins with no prior earnings, so she exploits in round 1. This yields $S_2 = 32$ and updates the background reward rate to $BRR_2 = 64$. Since $S_2 < BRR_2$, she explores in round 2, earning nothing. The background reward rate is updated to $BRR_3 = 64/2 = 32$. If exploration leads to $S_3 = 16$, she does not exploit ($S_3 < BRR_3$) and explores again, earning nothing. The background reward rate is updated to $BRR_4 = 64/3 \approx 21.3$. At this point, the background reward rate lies in between 64 and 16. Hence, the *MVT* rule will coincide with the *DyP* rule, regardless of the restarting stock: the agent correctly skips the poor patch and waits for a richer one, or immediately exploits the rich patch.

Example 3. *Start and restart in a rich patch.*

The agent begins with no prior earnings, so she exploits in round 1. This yields $S_2 = 32$ and updates the background reward rate to $BRR_2 = 64$. Since $S_2 < BRR_2$, she explores in round 2, earning nothing. The background reward rate is updated to $BRR_3 = 64/2 = 32$. If exploration leads to $S_3 = 64$, she exploits again ($S_3 \geq BRR_3$), bringing cumulative earnings to 128 and updating $BRR_4 = 128/3 \approx 42.7$. Since $S_4 = 32 < BRR_4$, she explores once more. Note that the agent never exploits a stock of 32, which is suboptimal according to the *DyP* rule.

Example 4. *Start and restart in a poor patch.*

With no prior earnings, the agent exploits in round 1 despite the low stock, yielding $S_2 = 8$ and $BRR_2 = 16$. Since $S_2 < BRR_2$, she explores in round 2. If exploration leads to $S_3 = 16$, she exploits ($S_3 \geq BRR_3 = 16/2 = 8$), bringing cumulative earnings to 24 and updating $BRR_4 = 24/3 = 8$. Since $S_4 = 8 \leq BRR_4$, she explores again. This example illustrates that landing on two consecutive poor patches— an event that occurs with probability 1/4 —leads the MVT agent to exploit a stock of 16 twice, which is suboptimal according to the *DyP* rule.

A.4 Instructions translated to English

Welcome. We want to thank you for accepting the invitation to participate in this activity with an approximate length of 10 minutes. This length includes an explanation of the activity, participating in the activity, and a short survey in which we will ask questions about you. After completing the survey, we will give you your game earnings. You will receive 5,000 COP for participating. Depending on your decisions in the game, you could earn up to an additional 17,000 COP. The funds to cover these expenses have been provided by Universidad del Rosario and MinCiencias.

This activity generates earnings such that the decisions you make have economic consequences, and are more similar to the decisions you make in your daily life. We hope you participate in future activities from other researchers, even if there is no associated payment.

All the information we collect will be used anonymously, so other participants will not know about your earnings or answers to the survey during or after the experiment.

The Selling Game

The selling game consists of getting the highest number of buyers within a determined time lapse. The higher the number of sales you have, the higher your earnings at the end of the game. You will be located in a commercial area with several selling locations.

In this game, you have 12 hours of labor. In each hour, you must make a decision:

- **Sell:** Add the remaining number of buyers at the selling location to your sales.
- **Move:** Find another selling location.

Why move?

In this game, after each hour at a selling location, the number of remaining buyers reduces in half. For instance, if at a selling location there are 16 buyers at 8 am, the number of buyers at 9 am reduces to 8, then goes down to 4 buyers at 10 am, and so on.

If you decide to move, you will spend one hour of the day, but you will arrive at a new selling location that could have more buyers.

There are two types of selling locations:

- In the high-selling locations, there are 64 initial buyers.
- In the low-selling locations, there are 16 initial buyers.

When moving, you will not know if the new selling location is high or low. Imagine that you are flipping a coin while moving. If it lands on "heads", you will move to a high-selling location and to a low-selling location if it lands on "tails."

At the end of the game, you will receive 50 COP for each buyer accumulated during the entire game. Your payment will be rounded to the nearest 1,000 COP. If, for instance, your earnings are 12,300 COP, you will receive 12,000 COP. If your earnings are 13,700 COP, you will earn 14,000 COP. Remember, you have 12 movements in total, representing a workday. Click “Start” and you will begin at the first selling location.

A.5 Instructions in Spanish

Bienvenido(a). Queremos agradecerle por haber aceptado la invitación a participar en esta actividad que tendrá una duración aproximada de 10 minutos. Este tiempo incluye la explicación del ejercicio, la participación en el ejercicio y una corta encuesta al final en la que le haremos unas preguntas sobre usted. Tras finalizar la encuesta, le entregaremos sus ganancias del juego. Usted recibirá $5.000 por participar. Dependiendo de sus decisiones, en el juego podrá ganar hasta $17.000 adicionales. Los fondos para cubrir estos gastos han sido proporcionados por la Universidad del Rosario y MinCiencias.

Esta actividad tiene unas ganancias para que su toma de decisiones tenga consecuencias económicas y, así sea más parecida a las decisiones que toma en su vida diaria. Esperamos que participe en futuras actividades de otros investigadores, aún si no hay un pago de por medio.

Toda la información recogida será utilizada anónimamente, así que los otros participantes no sabrán durante o después del experimento nada sobre sus ganancias o sus respuestas en la encuesta.

El juego de las ventas

El juego de las ventas consiste en obtener el mayor número de compradores en un lapso de tiempo determinado. Entre mayor sea el número de ventas que tenga, mayor será su ganancia al final del juego. Para esto, usted estará ubicado en una zona comercial con varios puntos de venta.

En este juego, usted dispone de 12 horas de trabajo. Cada hora usted deberá tomar una decisión:

- **Vender:** Sumar a sus ventas el número de compradores que quedan en el punto deventa.
- **Moverse:** Buscar otro punto de venta.

¿ Por qué moverse?

En este juego, luego de cada hora en un punto de venta, el número de compradores que quedan se reduce a la mitad. Por ejemplo, si en un punto de venta hay 16 compradores alas 8am, el número de compradores a las 9am bajará a 8, luego bajará a 4 compradores a las 10am, y así sucesivamente.

Si decide moverse, consumirá una hora del día, pero llegará a un nuevo punto de venta que podría tener más compradores.

Hay dos tipos de puntos de venta:

- En los puntos de venta altos, hay 64 compradores iniciales.
- En los puntos de venta bajos, hay 16 compradores iniciales.

Al moverse, usted no sabrá si el nuevo punto de venta es alto o bajo. Imagínese que, al moverse, lanzará una moneda. Si cae "cara" se mueve a un punto alto, y si cae "sello" a un punto bajo.

Al final del juego, usted recibirá \$50 pesos por cada cliente acumulado durante todo eljuego. Redondearemos su pago al múltiplo de \$1.000 más cercano. Si, por ejemplo, sus ganancias son \$12.300, usted recibirá \$12.000. Mientras que, si sus ganancias son \$13.700, usted recibirá \$14.000. Recuerde que tiene 12 movimientos en total, que representan una jornada de 12 horas de trabajo. Presione el botón "Comenzar" y usted arrancará en el primer punto de venta.