Submitted to *Management Science*

Dynamic Decision-making in Operations Management

Evgeny Kagan
Carey Business School, Johns Hopkins University, ekagan@jhu.edu

Stephen Leider
Ross School of Business, University of Michigan, leider@umich.edu

Ozge Sahin
Carey Business School, Johns Hopkins University, ozge.sahin@jhu.edu

Many Operations Management (OM) models assume that people act as forward-looking optimizers in dynamic environments. We experimentally examine this assumption. To cover a wide range of settings we look at several common classes of dynamic decision problems, and characterize behavior as either optimal, or as consistent with one of several non-optimal policies. At a high level, our results suggest that behaviors are not uniform and depend on the features of the problem. Specifically we find that: (1) Decisions are generally forward-looking (though not always optimal) in Technology Adoption and Capacity Allocation problems, but not in Search/Stopping problems. (2) The optimal policy is a good, but not the best, representation of behavior in both Technology Adoption and in Capacity Allocation; in both tasks simpler forward-looking heuristics achieve a better fit. (3) Performance (payoff) is correlated within-subject for different dynamic tasks but the specific policy usage may vary even within-subject. Together, these results provide new micro-foundations for researchers interested in building more descriptive models of dynamic behavior.

Key words: dynamic decision-making, behavioral operations management, experiments
History: **October 6, 2021**

1. Introduction

Many problems in Operations Management (OM) involve decisions that are "dynamic" or "sequential". The common approach to solving these problems is dynamic programming (Bertsekas 1995, Puterman 2014, Sutton and Barto 2018), which produces decision rules that can often be executed by software algorithms with no human involvement. Some sequential decisions, however, are less likely to be automated, for example a supplier choosing among retailers when allocating capacity (Cachon and Lariviere 1999), a consumer making periodic purchase decisions (Levin et al. 2009), or a shopper searching for the best alternative (Stahl 1989, Cachon et al. 2008). The use of dynamic programming in these settings presumes that human decision-makers are capable and willing to act as forward-looking optimizers. The goal of this study is to evaluate this premise.

Failure (or unwillingness) to perform forward-looking optimization is recognized in a number of studies that distinguish between "myopic" and "strategic" consumers (Cachon and Swinney

2009, Levin et al. 2009, Li et al. 2014). Indeed, some work in the revenue management and pricing literature has began incorporating richer behavioral phenomena related to sequential consumer choices (Özer and Zheng 2012, 2016). However, there has not been a systematic investigation into when (i.e. in what types of dynamic settings) we can expect decision-makers to be myopic or forward-looking, and whether a binary classification into one of these two categories is an accurate depiction of reality. Some dynamic decisions, such as search problems (Hey 1982, 1987, Seale and Rapoport 1997, 2000), or patient admissions (Kim et al. 2020, Kim and Tong 2021) have been explored from a behavioral perspective. However, the focus is typically on identifying heuristics specific to the focal environment, rather than on more generalizable behavioral rules that would apply to a range of settings.

The lack of behaviorally plausible alternatives to the standard model of forward-looking optimization is surprising given the emergence and broad acceptance of "behavioral", i.e. boundedly rational models in other decision domains, such as risk and uncertainty ("Prospect Theory", Tversky and Kahneman 1974, 1992), time delays ("Hyperbolic Discounting", Laibson 1997) and social comparisons ("Inequity Aversion", Fehr and Schmidt 1999, Bolton and Ockenfels 2000). These models posit that by focusing on (expected) utility maximization alone the standard economic model may oversimplify behaviors. Similarly, in the domain of dynamic decision-making, the classic forward-looking optimization paradigm may be inconsistent with how people approach the problem. Finding the optimal solution may be beyond an individual's cognitive abilities – or the cognitive effort may not be worth the returns.[1]

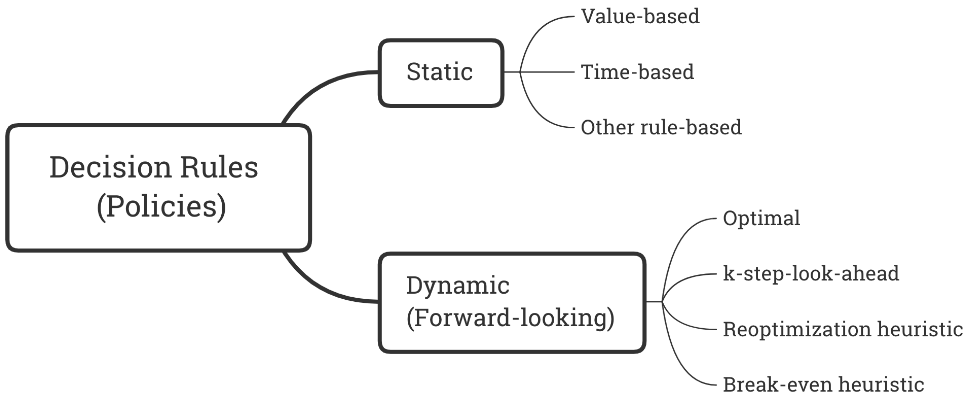
The goal of this study is to take a broad look across different dynamic settings to identify the major behavioral "landmarks" as a guide to future efforts to develop a general behavioral model of dynamic decision-making. To do so, we will explore multiple distinct decision problems, and will be focusing on decision rules that apply across problems (rather than policies tied to one specific problem). Our research questions are:

- What types of dynamic environments prompt a forward-looking vs. a more static (not forward-looking) solution approach?
- Among the forward-looking solution approaches, which are the most prevalent ones?
- At the individual level, do people use similar approaches across tasks? Are there individual markers that predict the solution approach?

To cover a wider range of plausible behaviors we begin by developing a general taxonomy of possible approaches to solving dynamic problems (Figure 1). This taxonomy is not exhaustive, but

[1] It is important to note that our focus is not on time discounting, i.e. the utility weights placed on future payoffs, which have been studied extensively in the literature, e.g. Laibson (1997), Tirole (2010). None of our experiments involve discounted payoffs. Rather, we focus on the cognitive complexity of considering the range of future possibilities when making a dynamic decision.

Figure 1 Taxonomy of Decision Rules



it synthesizes the types of heuristics proposed in the literature (Section 2), with the addition of several decision rules reported by our experimental participants in post-study questionnaires.

At a high level, decision rules can be either static or dynamic (forward-looking). Static rules presume that the decision-maker uses a fixed decision rule throughout the task horizon (e.g. to always purchase or always stop if the available product has a value above some fixed threshold). More sophisticated, forward-looking policies include the optimal policy and several non-optimal heuristics. These heuristics simplify the problem by shortening the decision horizon, by repeatedly solving a simplified problem, or by using a break-even rule for accruing value/spending resources. Together, these rules capture a range of potential behaviors, are applicable across a range of problems and are identifiable from subjects' choices.

We mathematically formulate each decision rule (policy) for three classes of dynamic decision problems: (1) Technology Adoption, (2) Capacity Allocation, and (3) Search/Stopping problems. We choose these problem classes because they represent a large share of dynamic decision problems in the Operations Management literature and apply to a wide range of decision contexts, both for businesses and consumers. Each problem class has its own unique structure, and so it is not a priori clear whether the same strategies will be used across problems, or whether each category prompts decision-makers to use a different strategy. To explore a variety of decision environments, we choose versions of these problems that span a range of complexity, e.g. our Technology Adoption problem is deterministic with a simple decision and state space, while the Capacity Allocation is stochastic and has a richer state and action space. We then conduct laboratory experiments with a naturalistic framing (e.g. purchasing a new laptop, or searching for an apartment) to examine the prevalence of each type of decision rules in each problem class.

To be able to identify which policies are common in each task, a key design feature of our experiments is the use of the strategy method. That is, rather than directly asking participants to make a decision (e.g. stop or continue), we elicit their acceptance thresholds (e.g. "What value would make you stop in the next period?"). To ensure that different policies are well identifiable, we choose versions and parametrizations of tasks that produce a noticeable difference in thresholds implied by each policy. The resulting data set is a richer representation of behavior that captures a large number of counterfactuals and allows us to reliably match decision rules to subjects.

Our first experimental result is that the degree of strategy sophistication varies by problem. For the Capacity Allocation and Technology Adoption tasks, most subjects use a forward-looking strategy, whereas in Search and Stopping tasks the majority use simple static rules. This result is not fully explained by the differences in incentive strength, in payoff accumulation rate, or in decision elicitation procedures. Rather, what prompts the different types of approaches appears to be a more fundamental difference between Search/Stopping problems and other dynamic problems.

Second, specific policies also differ by problem class. Static strategies in Search/Stopping problems tend to reflect salient characteristics of the problem. Depending on the problem, subjects may either target a certain, satisfactory payoff value or, a certain probability of a random event. Forward-looking policies also differ by task. The most common policy in Technology Adoption is the limited look-ahead policy, while the most common policy for the Capacity Allocation task is the reoptimization heuristic. This aligns with the increased complexity of our Capacity Allocation task relative to the Technology Adoption task, in part because the former is stochastic, while the latter is deterministic. The complexity and stochasticity of the Capacity Allocation task seem to require the kind of substantial simplification offered by the reoptimization heuristic.

Third, performance (payoff) is correlated within subject for different dynamic tasks. However, the strategies used in different tasks often differ within-subject. Even top performers do not consistently follow optimal or even forward-looking policies across all tasks, but rather use a mix of forward-looking, and well-calibrated static policies. Additionally, subjects' performance on the Cognitive Reflection Task (CRT) correlates with their strategies and performance in (the simplest) Technology Adoption task but is a weaker predictor in other tasks. These results point to an underlying ability to solve dynamic problems that is distinct from standard measures of analytic intelligence, but also suggest that this ability may translate into different decision rules, depending on the features of the problem.

Taken together, our results suggest that treating people as either fully myopic or as perfect forward-looking optimizers may be unrealistic. Rather, behavioral patterns are sometimes sophisticated (though not always in ways optimal policy would imply) and other times not, where sophistication is determined by the person's abilities, problem features, and the cognitive costs and monetary benefits of using a more sophisticated approach. Understanding these patterns can help researchers build better, more descriptive models of human decision-making, and can help managers appropriately respond to consumers exhibiting such behaviors.

2. Literature

While some dynamic settings have been explored in the literature, there are no studies that we are aware of that study behaviors across multiple, distinctly different dynamic settings. Rather, most existing work examines a particular dynamic setting, e.g. job search or life-cycle consumption-savings, with a natural focus on the models and behaviors tied to that setting. The narrow focus helps better understand and inform the relevant setting; however behaviors identified in one setting may not easily export to another. For example, it is not clear how to take the intuition of an explore-exploit heuristic from the search domain and apply it to a technology adoption problem. Nonetheless, our study is informed by, and helps integrate some of these streams of literature.

Search and Stopping Early studies of search and stopping behavior found subjects to perform quite close to optimal, using sophisticated threshold rules (Schotter and Braunstein 1981, Cox and Oaxaca 1989). However, much more commonly subjects end the search too early (Seale and Rapoport 1997, 2000, Bearden et al. 2006, Eriksson and Strimling 2010). Some work has found that structural factors, such as search cost (Zwick et al. 2003) or completeness of information (Palley and Kremer 2014), affect individuals' propensity to search too little or too much. Common heuristics include trend-based rules (Hey 1982, 1987, Moon and Martin 1990), searching for a fixed number of periods (Moon and Martin 1990), explore-exploit patterns (Seale and Rapoport 1997) and satisficing (Phipps and Meyer 1985, Sonnemans 1998). Schunk and Winter (2009) find that most people use some combination of a fixed stopping rule, a satisficing stopping heuristic, or a finite-horizon search rule. Importantly, the existing studies focus on a specific economic setting and choose the variant of the search problem that represents that setting. In contrast, our objective is to characterize the policy used by each subject (as opposed to documenting *when* they stop); we therefore choose structural features of the search problem that best reveal the policy in use.

Intertemporal Saving and Spending A second well-studied type of dynamic problems are lifecycle savings. In these problems an individual earns income in each period and chooses how much to consume versus save/invest (see Hey and Dardanoni 1988, for an early example, and Duffy and Li 2019, Cornand and Heinemann 2019, for excellent recent surveys). The classic finding is that subjects do too little income smoothing, with current consumption too closely tied to current income which leads to too little wealth accrual (Carbone and Hey 2004) . Noussair and Matheny (2000) show that either over- or under-consumption can occur depending on the investment technology, but that in either case suboptimal consumption "binges" are common. Carbone (2006) finds that subjects are more forward-looking at the start of a problem but become more myopic over time. Brown et al. (2009) examine two potential explanations for these behaviors: a time preference explanation (preference for immediacy) and a cognitive limit explanation (bounded rationality).

Meissner (2016) shows that debt aversion significantly drives subject choices. Ballinger et al. (2003, 2011) explore individual correlates of task performance and find that pattern completion and working memory tasks predict performance; that is, to perform well one requires both, attention to the problem's complexity and the cognitive ability to manage complexity.

Similar to the search literature, the studies in the lifecycle savings literature are geared towards a specific setting – personal finance decisions. The unique features of this setting, such as the negative associations people may have with debt, make it difficult to generalize the results to other settings. Nonetheless, these studies are related to ours: one of our tasks, the Capacity Allocation task, also involves trade-offs between immediate vs. delayed (but potentially higher) earnings. To focus on the dynamic allocation trade-off we remove potentially aversive concepts such as debt and technical concepts such as interest rates, and instead focus on the operational decision to allocate a fixed capacity among sequentially arriving consumption alternatives.

Capacity and Revenue Management Applications This group of studies examines individual ability to balance over and underusage of the available units, as subjects decide how to allocate their inventory among sequentially arriving candidates. Some studies have found overly conservative behaviors at the beginning of the sale cycle (Bearden et al. 2008, Bendoly 2011), while others find overly accepting behaviors (Bendoly 2013). Diehl and Sterman (1995) document an overall poor performance in a more complex inventory management setting. They attribute this to the difficulty of taking a system level view and being (un-)able to correctly understand how one's actions interact with environment to influence outcomes (Similar to the typical behaviors in the Beer Game, see for example Sterman 1989). Kim et al. (2020) and Kim and Tong (2021) study capacity management in a health care setting, where inventory (hospital bed availability) changes over time and the task is to admit (or not) sequentially arriving patients.

Leider and Şahin (2014) examine behavior in a capacity allocation setting where the task is to accept or reject sequentially arriving consumption opportunities with different values (See also Ascarza et al. 2012, Gopalakrishnan et al. 2015). They find consumption decisions to be sophisticated, but not optimal, with features best described by a forward-looking reoptimization heuristic. Our Capacity Allocation problem (Section 3) builds on their design, with some modifications to allow better separation between heuristics. More importantly, to produce more generalizable insights our study compares behaviors in the Capacity Allocation task with other dynamic settings.

Other Dynamic Decision Settings Long et al. (2020) study sequential abandonment decisions for projects with uncertain value and find that abandonment decisions are highly path dependent, being influenced by loss aversion, status quo bias, and the sunk cost fallacy. Kremer and de Vericourt (2020) examine the dynamic decision to continue/stop collecting tests before making a medical diagnose. They find that decisions are linked to system congestion, with subjects diagnose prematurely at low congestion, and not stopping fast enough at high congestion. Hathaway et al. (2021) look at similar issues in a customer service context and find that behaviors correctly respond to system congestion. Anderhub et al. (2000) look at expenditure decisions from a limited monetary endowment over a stochastically varying number of periods and find that subjects are able to correctly incorporate information updates into their decisions.

Gabaix et al. (2006) examine costly information acquisition and find that a partially myopic policy fits the data better than the fully rational model. Lastly, Oprea (2020) study what makes a dynamic decision policy attractive and find the number of decision states, the number of potential transitions, and the presence of absorbing states to matter. These studies are related because they explore the interplay between problem/policy complexity and policy adoption. Rather than studying the features of a specific problem, we study a range of problem classes and explore the differences and similarities of approaches used to solve these problems.

3. Dynamic Decision-making Models

We begin by introducing a general discrete-time finite-horizon dynamic problem and its main ingredients. We then present the three problem classes studied in our experiments and state the relevant decision rules.

3.1. A General Dynamic Decision Problem

To make the models amenable to experimentation with untrained participants, we restrict our attention to systems in which rewards (costs) are additive over time. The state transitions and the rewards accrued to the decision-maker are illustrated in Figure 2. The system runs from $t=1$ to T in discrete time and with discrete and finite states, with appropriately chosen initial and terminal conditions. If we denote by a_t the action taken at time t and by w_t any stochastic disturbance, then the system's states s_t evolve as follows:

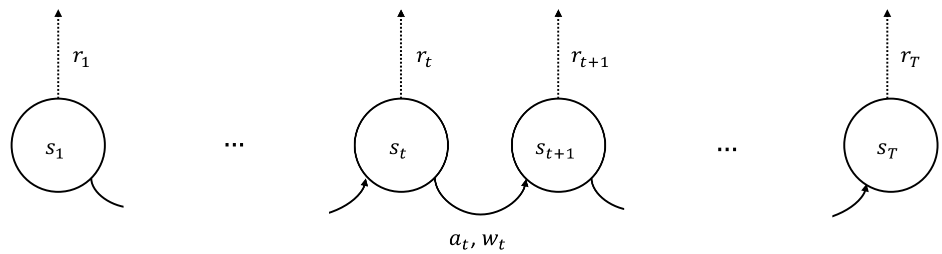
$$s_{t+1} = f_t(s_t, a_t, w_t), \qquad t = \{1, 2, .., T-1\}$$

where f_t is the transition function at time t that can depend on the previous state, on the action taken at time t and on the stochastic disturbance. If we denote the reward (or cost) incurred at time t by $r_t(s_t, a_t, w_t)$ then the total expected reward (or cost) is as follows:

$$E\left[\sum_{t=1}^{T} r_t(s_t, a_t, w_t)\right].$$

We restrict our attention to processes with no time discounting, and note that rewards may be zero for some of the stages (for example, in stopping problems).

Figure 2 A General Dynamic Decision Process



3.2. Problem Formulations

We next introduce the three problem classes studied in our experiments. Each class falls under the general framework of finite-horizon dynamic decision processes (Figure 2); at the same time each class has some unique features that may lead to a more/less widespread adoption of certain types of decision policies. For each problem we use a naturalistic framing with a context that is familiar to a typical experimental participant (college student).

Technology Adoption Problem Sequential technology adoption is a classic problem in many firm and consumer settings (See, for example Balcer and Lippman 1984, Ho et al. 2002). Below we present the consumer version of the problem to mirror the decision situation faced by our experimental participants; however variants of this sequential adoption are quite common in manufacturing settings, with the focal decision being when to invest in additional capacity or in a new technology (See, for example Porteus 1985, Van Mieghem 1999).

Consider a college student who currently owns a laptop and derives u_0 units of value per period from using that laptop. New versions of the laptop are scheduled to be released in years $t = 1, 2, ..., T$ and are valued by the student at u_t units per period ($u_t \geq u_0$ for $t = 1, 2, ..., T$ and at least weakly increasing in t). The release dates of new versions and specifications are common knowledge. Each new model purchase costs p units. In this problem, the state (s_t) is the value of the laptop currently owned by the student in period t, the control (a_t) is whether to upgrade or not in period t and there is no disturbance since the problem is deterministic. If the student upgrades to a new model ($a_t = 1$) then the new state is the value of the new model, $s_{t+1} = u_t$. If not ($a_t = 0$) , then the state remains unchanged, $s_{t+1} = s_t$, i.e. the student keeps the same model as in the previous period. The net reward earned in each period (r_t) is the value of the laptop owned in that period minus the price if a purchase was made. The objective is to maximize the sum of the rewards over T periods net the total cost of acquisition(s).

Capacity Allocation Problem Finite-horizon capacity allocation is another classic dynamic problem in the OM literature (Lee and Hersh 1993, McGill and Van Ryzin 1999). As before, we use a naturalistic context and frame the problem as a series of consumption decisions, where a college student with a meal plan consisting of C free meal tokens needs to consume them over T days with $C < T$. Each day, meal values X are drawn independently from the distribution F known to the student. The goal is to maximize the sum of meal values by allocating C free meal tokens to high value meals over T days. Additional meals can be purchased at price p once the student has run out of free meal tokens. In this problem, the state (s_t) is the number of free tokens left at the start of day t, the control (a_t) is whether to accept or reject the meal offered that day and the disturbance (w_t) is the stochastic value of the meal offered that day. If, on day t the student rejects the meal, then the state remains unchanged, $s_{t+1} = s_t$. If, alternatively, the student accepts the meal, then the state transition depends on the availability of free tokens in the student's account: if free tokens are available ($s_t > 0$) then the student pays nothing and $s_{t+1} = s_t - 1$; If not ($s_t = 0$) then the state remains unchanged, $s_{t+1} = s_t = 0$. Accepted meals increase the accumulated rewards by x units (with x denoting the realization of X). Thus, if a meal is accepted the total net reward increases by $r_t = x$ if $s_t > 0$, and by $r_t = x - p$ if $s_t = 0$.

Product Search Problem The Product Search problem is a variant of the classic secretary problem with full recall. A college student wants to rent an apartment. There are T apartments available in the market, and at each period the student can visit (search) an apartment at cost c. The student wants to maximize their net profit, i.e. the value of the rented apartment net the total cost of apartment search. The value of the apartment visited in period t is a random variable $X_t \sim F_t$ with the realization fully revealed after that apartment is visited. The search is with full recall; i.e. if the student stops searching in period t, then the value received is $X_t^{max} = \max\{X_1, X_2, ..., X_{t-1}\}$. In this problem the state (s_t) is defined by x_t^{max} (realization of the random variable X_t^{max}), the control (a_t) is to stop searching or to continue, and the disturbance (w_t) is the randomness in apartment values. We focus on value distributions $F_t \sim U[a_t, b_t]$ where mean $\frac{a_t+b_t}{2}$ is (weakly) increasing in t. We choose uniform distributions with an increasing mean because they allow better identification of decision policies, while being relatively straightforward to explain to experimental subjects.

3.3. Taxonomy of Decision Rules (Policies)

In Figure 1 we have briefly introduced the classes of decision policies examined in our study. These policies synthesize the heuristics proposed in the previous literature with the type of rules described by our experimental participants in exit surveys. Further, we restrict these policies to the ones that can be adapted to different problem types, i.e. are not uniquely tied to any specific problem. Finally, recognizing that our taxonomy may not be exhaustive, our data analysis will allow classifying a participant as "Unidentified" if no policy achieves a good fit for that participant (See Section 5.1 for the classification procedure).

At a high level, the examined policies fall into two categories: static and dynamic (Figure 1). The static policies generally focus on one prominent feature of the problem, for example the length of the time horizon, and use an ad-hoc rule based on that feature. Dynamic policies include the optimal policy and several non-optimal policies: k-step-look-ahead (LKH), reoptimization heuristic and break-even heuristic. These policies simplify some aspect of the problem, or use some logic that responds to changes in the state and time, but fails to capture some important aspect of optimal decision-making.

In the remainder of this section we present the mathematical formulations for each non-optimal policy. We use one of the problem types to illustrate each policy. The optimal policies are straightforward to obtain via backward induction, have standard (threshold) properties and are provided in Appendix A1. The remaining non-optimal policies are in Appendices A2 and A3.

3.4. Static Policies

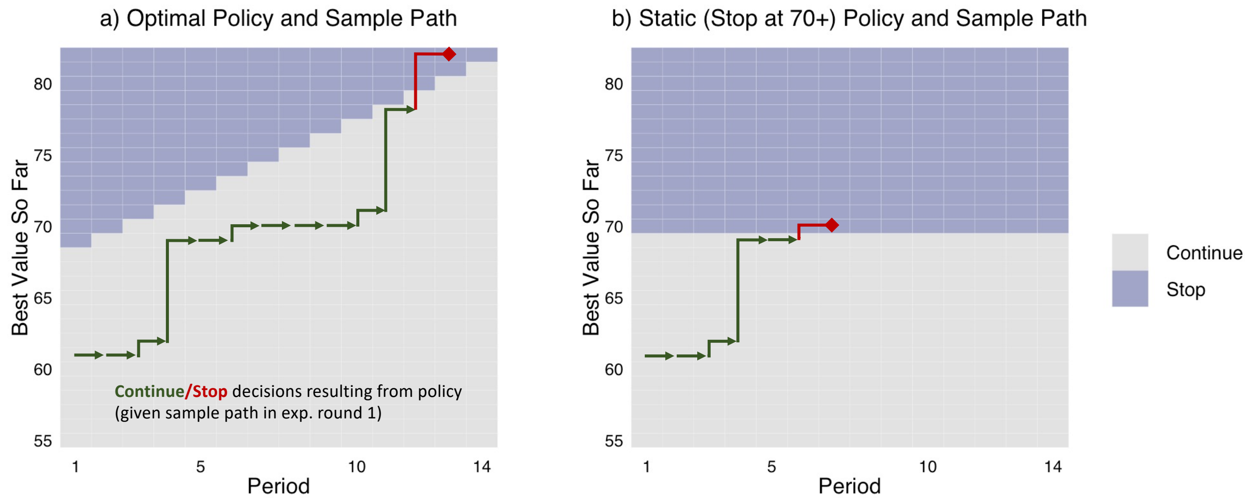
We begin with the class of static policies. These are decision heuristics that, depending on the specific rule, do not require the decision-maker to keep track of the state, or time period, or both. Instead, they focus on a single prominent feature of the problem, and use an intuitive ad hoc rule based on that feature. The feature can be value-based (for example, "Stop if reach value y"), time-based (for example, "Always stop in period t"), or mixed. The categorization into value and time-based decision rules originates in consumer search models (see Feinberg and Huber 1996, and references therein). We adopt this categorization and examine the use of such rules across the different problems in our study.

3.4.1. Static value-based heuristic Static value-based heuristics reduce the multi-period optimization problem to a single period problem using a control on the value (reward) received in each period. The idea is to approximate the main trade-offs of the dynamic problem by choosing a time invariant action that balances expected total cost and benefit over the entire decision horizon. This simplification dramatically reduces the cognitive burden because it only requires comparing the observed values with a single number ("reservation value") throughout the horizon.[2]

Example: Product Search The static value-based rule is to choose a target apartment value, y, at the beginning of the horizon. The decision-maker stops searching as soon as the best apartment value among the apartments searched is greater than or equal to y.

[2] Static value-based approaches have been proposed by Caplin et al. (2011) to represent satisficing behavior when evaluating each alternative is costly, by Borenstein (2009) to represent electricity consumption decisions and by Grubb and Osborne (2015) to represent cellular service usage. In our data, subjects' free-form descriptions of their decision approaches reinforce static value-based policies as being both intuitive and common. For example many subjects describe having a single target value they want to achieve, such as a subject in the Product Search problem that said they "chose 69 or 70 every time, since they are in the middle of the values".

Figure 3 Static vs Optimal Policies: Product Search



The decision-maker can either choose an arbitrary target value, for example the midpoint of the price distribution, or take a more sophisticated approach (while staying within the class of static policies) and choose the "optimal" static value. If we define $y_{(t)} = E[\max(X_1, X_2, \cdots, X_t)]$ then there are N candidate target values $y_{(t)},\ t = 1, \cdots, N$. The optimal target value can be found by solving the following problem

$$\max_{t \leq T} \ E[\max(X_1, X_2, \cdots, X_t) - ct]$$

with the solution being given by $t \times y_{(t)} = \max\{t : t \leq T, E[\max(X_1, X_2, \cdots, X_t)] - E[\min(X_1, X_2, \cdots, X_{t-1})] \geq c\}$.[3]

Figure 3 demonstrates the differences between a static value-based heuristic (stop at value 70) and the optimal policy for the parameter set used in the Product Search task. We also plot the sample path resulting from the random variable realizations in one of the experimental rounds (with green arrows indicating continuation and red arrows indicating stopping). As shown in Figure 3, a subject following any static policy would be easily identifiable because they would use the same threshold throughout the task.

3.4.2. Static time-based heuristic Similar to value-based control, time-based control ignores one aspect of the problem – the values observed by the decision-maker – and uses only the time (period) information to make decisions. We illustrate the intuition of this heuristic again using the Product Search problem. As before, the stopping period can be chosen ad hoc, for example the

[3] One can find the corresponding target value by inverting the optimal time to stop to the target value $y_{(t^*)}E[\max(X_1, X_2, \cdots, X_t^*)]$. Then the individual continues to search as long as $\max(X_1, X_2, \cdots, X_t) < y_{(t^*)}$ and stops otherwise.

middle of the time horizon, or in a more sophisticated way. The best static time-based rule can be found by solving the following problem:

$$\max_{t \leq T} \; E[\max(X_1, X_2, \cdots, X_t) - ct]$$

The solution is given by

$$t^* = \max\{t : t \leq T, E[\max(X_1, X_2, \cdots, X_t)] - E[\min(X_1, X_2, \cdots, X_{t-1})] \geq c\}.$$

Note that unlike the value-based heuristic, t^* is determined before observing any of the apartment values, so that the decision to stop is independent of the actual values of the apartments observed at $t \leq t^*$. As we will see later, this helps identification since a person who uses a value-based vs. a time-based static heuristic would exhibit very different sequences of stopping thresholds.

The versions of static policies for the remaining two problem classes (Capacity Allocation and Technology Adoption) are straightforward to derive and are presented in Appendix A. Additionally, we also examine several other, interval-based static rules that are not defined for Search, for example buying a laptop every s periods in Technology Adoption, or consuming a unit every s periods in Capacity Allocation.

3.5. Dynamic Policies

3.5.1. Limited look-ahead policies (LKH) Under this class of policies the decision-maker assumes that the game ends after the next k periods and makes optimal choices under that assumption. If $k = 0$ the decision-maker is fully myopic and considers only current rewards and costs, ignoring any continuation values. Similarly, if $k = 1, 2, ..., T - 1$, the decision-maker determines the best strategy by considering current and next period costs and rewards by re-solving a k-period counterpart of the dynamic program at each period $t < T$.[4]

Example: Technology Adoption: When $k = 0$, this problem reduces to a single period deterministic decision where we compare the net utility from adopting the currently offered technology, $u(t) - p$ with utility of current technology, $u(n)$: Adopt if $u(n) < u(t) - p$, wait otherwise. Analogously, 1-step-look-ahead policy at time t with laptop version n is the solution to the following optimization problem:

$$V_t^1(n) = \max\{u(n) + V_{t+1}^1(n), u(t) - p + V_{t+1}^1(t)\}, \quad V_{t+1}^1(n) = \max\{u(n), u(t+1) - p\}.$$

[4] Several versions of look-ahead policies, especially with $k = 0$, have been proposed as a myopic counterpart to forward-looking optimization in supply chain settings (Lin et al. 2018), in revenue management and pricing (Aviv and Pazgal 2008, Gallego et al. 2008), and in queuing control (Hanukov et al. 2020). In our data, subjects would often specifically mention how many periods they were considering. For example, a subject in the Technology Adoption problem said they would "only purchase a laptop where the value increased significantly and where it would pay off its purchase price in the next 3 periods."

Figure 4 LKH2 vs Optimal Policies: Technology Adoption

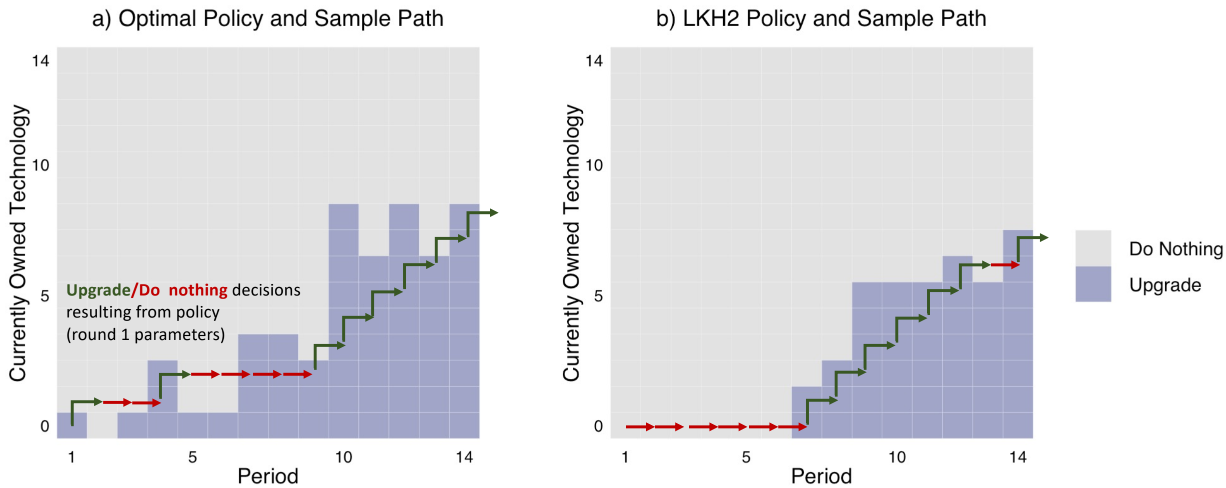


Figure 4 demonstrates the differences between a look-ahead policy with $k=2$ (LKH2) and the optimal policy for the parameter set used in round 1 of our experiments. Note the strong contrast between the policy predictions: the LKH2 policy is quite myopic and predicts no purchases in the early periods. This is because the relatively low value increases in the early periods are not sufficient to offset the purchasing price if one believes that the game would already end in two periods. In this example, a subject following LKH2 policy would deviate from the optimal policy in 6 out of 14 periods.

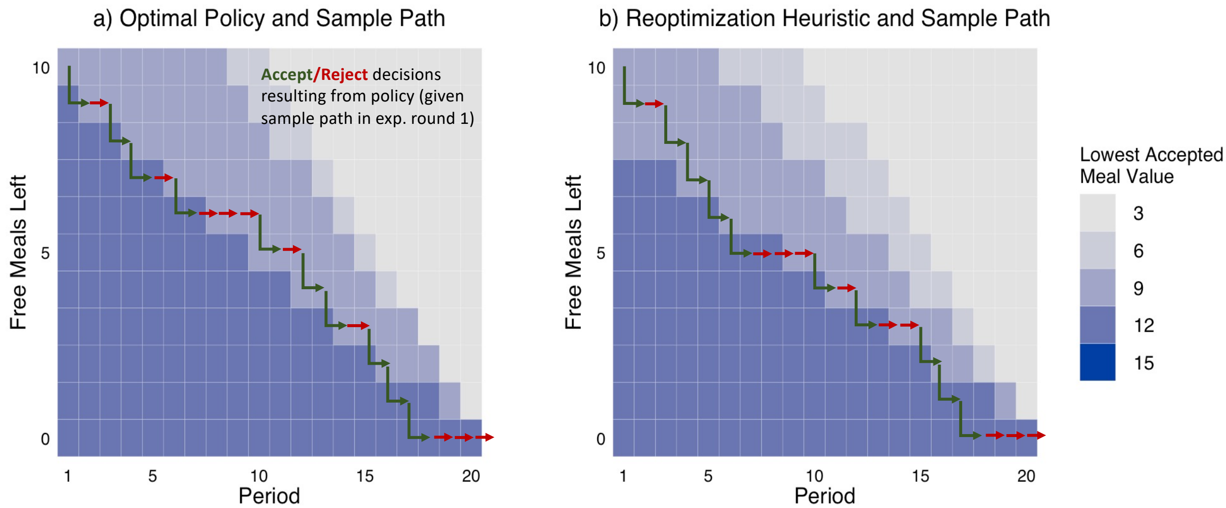
3.5.2. Reoptimization Heuristic The reoptimization heuristic transforms the optimization problem to a single period problem at time t and state s using a control on the reward. Specifically, the decision-maker updates the static value-based control at each period $t=1,2,...,T$ by re-solving the single period optimization problem of finding the best static value-based rule (introduced in 3.4.1). In other words, the policy reoptimizes the static threshold at each period taking into account the current state of the system. Note that although the action is determined assuming a static control for the rest of the horizon, the acceptance thresholds change over time, since the action is updated in each period.[5]

Example: Capacity Allocation Suppose that there are $T-t$ periods left until the end of the horizon, and the decision-maker has k free meals remaining. To calculate the best static value-based policy, at each $t \in T$ the decision-maker chooses the threshold q as follows:

$$\max_{q \le p} \; E[X|X>q]\min((T-t)\bar{F}(q),k) + E(X-p)^{+}\frac{((T-t)\bar{F}(q)-k)^{+}}{\bar{F}(q)}.$$

[5] In our data subjects often describe having an initial simple rule that they would adjust in future periods depending on the outcomes. For example, a subject in the Capacity Allocation problem said that they “started with [accepting] 9,12,15” and then ”saw if I had 5 free meals by the tenth period, and calibrated accordingly.”

Figure 5 Reoptimization vs Optimal Policies: Capacity Allocation



In other words, this approach replaces random meal values with their expectations, and transforms the problem to a repeated single period optimization problem, where all meals arrive at once and the decision-maker chooses the ones that are above a certain value. The decision-maker then updates ("reoptimizes") the threshold q in each period depending on remaining capacity. Figure 5 illustrates this approach and the resulting decision path relative to the optimal policy (under the parametrization used in our experiments). In this example a subject following the reoptimization heuristic would deviate from the optimal policy in 4 out of 20 periods.

3.5.3. Break-even Heuristic Another common simplifying approach is to identify a cost for taking an action (e.g. purchasing in the Technology Adoption problem, consuming a meal in the Capacity Allocation problem, or choosing to continue or stop in the Search problem). The decision-maker then adjusts the threshold so that the value generated at least meets or exceeds that cost. This is a forward-looking heuristic, as the threshold will naturally adjust over time based on the state (e.g. the cost of consuming a meal in the Capacity Allocation problem will vary based on how many periods there are left to go). Setting the break-even threshold is similar to accepting a certainty equivalent in exchange for a risky alternative, and can also capture having a desired value accrual rate (e.g. in the Technology Adoption example the increment in value earned each period has to offset the purchase cost).[6]

Example: Technology Adoption If the decision-maker adopts technology $u(t)$, their total net utility from usage is $(T-t)u(t)-p$. If the decision-maker does not and instead keeps n^{th} technology, the total utility is $(T-t)u(n)$. The break-even heuristic targets a non-negative net utility accrual

[6] In our data, some subjects would describe their strategy as having a cost-based benchmark, such as a subject in Technology Adoption saying they considered the "net cost in upgrading per turn remaining".

per period, i.e., a certain per period utility improvement that offsets its cost if used until the end of the horizon. If the target accrual rate is any value greater than zero for the remaining time, the policy can be stated as follows: Adopt the new technology if and only if the per period additional benefit is larger than per period additional cost, $u(t) - u(n) \geq \frac{p}{(T-t)}$.

4. Experiment Design

To examine the decision rules used by subjects in different types of dynamic environments we conducted two laboratory studies. The treatments and tasks are summarized in Table 1.

In Study 1 we administered three tasks, one for each class of problems. Consistent with the problem description in Section 3.2, each task was framed as a series of consumption decisions familiar to college students. To identify the strategy used in each task we used the strategy elicitation method, i.e. asked subjects to indicate an acceptance threshold (i.e. minimum meal value in Meal Plan, and lowest accepted apartment value in Apartment Search) in each decision period.

In Study 2 we modified the original tasks to examine some of the mechanisms driving the behaviors observed in Study 1. Specifically, Study 2 examines whether policy adoption is affected by the following factors: the way in which the problem (action space) is presented, the sensitivity of payoffs to the strength of the incentives, as well as the way in which payoffs accumulate. We next describe the design of Study 1. Experimental instructions for each task, and screen shots of each decision environment are in Appendix B. The design of Study 2 is described in more detail at the beginning of Section 6.

Table 1 Summary of Treatments and Tasks

Study objectives	Treatment	Task 1 (4 rounds)	Task 2 (4 rounds)	# subjects
Study 1: Investigate decision rules used in different dynamic problems	1A	Technology Adoption	Product Search	77
	1B	Technology Adoption	Capacity Allocation	79
Study 2: Investigate the drivers of differences between stopping/search and other dynamic problems	2A	Machine Replacement	Product Search with continuous action space	43
	2B	Machine Replacement	Product Search with stronger incentives	42
	2C	Machine Replacement	Capacity Allocation with discrete action space	41

Note. Tasks were administered in reverse order for half of the sessions.

4.1. Study Setup and Parameters

Study 1 included two between-subject treatments with two tasks per treatment (presented in random order to control for order effects). In the first treatment 79 subjects worked on the Technology Adoption task and the Product Search task. In the second treatment 77 participants worked on the Technology Adoption task and on the Capacity Allocation task. We chose to expose each subject to two tasks to examine the commonalities as well as potential differences in how human decision-makers approach these dynamic problems.[7]

Participants were undergraduate and graduate students at a large, public Midwestern university. Experiments were programmed in z-Tree (Fischbacher 2007). After the subjects completed the dynamic tasks we elicited their risk attitudes using the multiple price list method (Holt and Laury 2002) and administered the Cognitive Reflection Test (CRT). We also asked subjects to verbally describe the strategies they had used in each dynamic problem. Average earnings was $22, including the show up fee of $7. Sessions lasted approximately 70 minutes and were conducted in person.[8]

All subjects completed four rounds of each task. We refer to “rounds” as complete iterations of a task, not to be confused with “periods” in each dynamic game within a round. To be able to identify general (as opposed to parameter-specific) patterns of behavior in dynamic systems we used several sets of price and value parameters in the deterministic (Technology Adoption) task, and several sets of realizations of the random variables in the stochastic (Capacity Allocation and Product Search) tasks. Within each session, we fixed both the parameters and the realizations of random quantities and controlled for those in our analyses, to reduce noise and be able to better identify relevant behaviors (See section 5.1 for the details of our policy identification approach). The specific parameters used in each task are reported in Appendix B.[9]

5. Experiment 1: Results

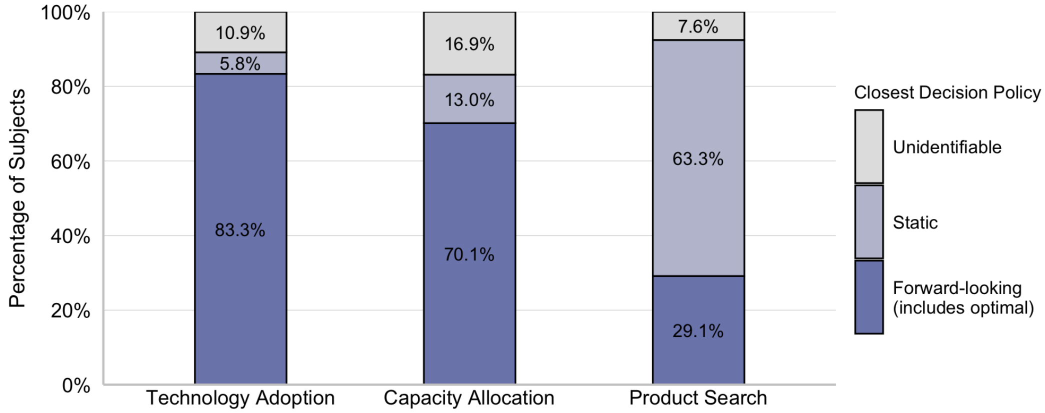
Below we present three sets of results. We begin by analyzing the sophistication of subjects' decision-making approaches in each task. To do so, we characterize subjects based on whether they are using a static or a forward-looking decision rule. Within these two categories we then explore which specific rules and strategies are common in each task. Lastly, we examine the extent to which decision rules generalize/are consistent within-subject (i.e. for the same decision-maker) across different types of dynamic decision-making environments, and whether they are related to individual measures of cognitive performance and to risk preferences.

[7] We chose two rather than three tasks per subject to reduce the effects of fatigue on decision-making, while still being able to observe within-subject behavior under multiple parametrizations/random draws.

[8] Upon receiving the instructions, all subjects were required to pass a mandatory comprehension quiz. The majority of the subjects (93%) passed the quiz on the first attempt. The remaining subjects received additional instructions, after which they were able to pass the quiz.

[9] Before conducting the experiments we performed Monte-Carlo simulations to choose parametrizations that would lead to minimal overlap between different policies. This allows us to separately identify most policies. Details on the simulation approach and analysis are available from the authors upon request.

Figure 6 Study 1: General Sophistication by Task



Note. Figure shows the share of subjects in each task whose decisions are consistent with a static or forward-looking decision approach. Subjects are classified as unidentifiable if they are neither static, nor sufficiently consistent with any of the forward-looking policies (using 70^{th} percentile of all consistency scores within a task as a cutoff for identifiability). Assigning the unidentifiable subjects to the closest policy results in 94% (73%, 37%) of the subjects being forward-looking in the Technology Adoption (Capacity Allocation, Product Search) tasks.

5.1. Sophistication of decision rules

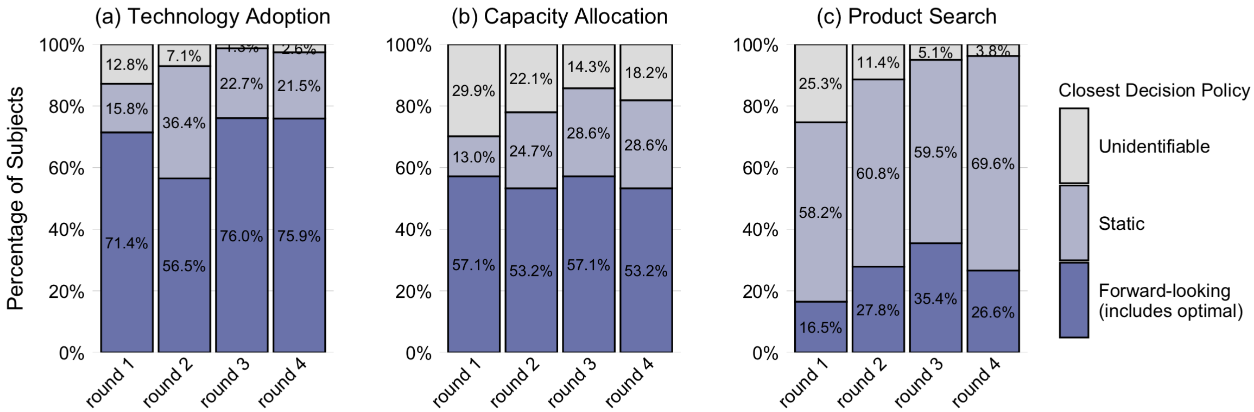
We begin by examining general sophistication of decision rules in each task environment, i.e. the extent to which subjects adopt static vs forward-looking rules. To do so, we first identify whether each decision sequence is consistent with a static decision rule. A decision sequence is considered static if it falls under one of the static rules described in Section 3.4. We classify a decision sequence as static as long as there is no more than one deviation from a static rule. We then label a subject as static if the subject used a static rule in at least three out of four of the experimental rounds.[10]

For subjects that do not appear to be static, we distinguish between subjects that match one of our forward-looking policies, versus subjects that we don't have a good descriptor for. To that end, we compute for each subject, round and policy a "consistency score", normalized to be between 0 and 1, to make the scores interpretable and comparable across tasks.[11] Specifically, for each policy p we calculate the total number of deviations in the observed sequence by subject i in round t as d_{itp}. We also compute the maximum empirically observed deviation d_{max} across all policies, subjects and rounds. Then, the consistency score for each decision sequence is $\frac{d_{max}-d_{itp}}{d_{max}}$, where 1 denotes maximal coincidence with the policy, and 0 denotes the largest possible deviation. To

[10] Our main results are qualitatively similar with alternative cutoffs, e.g. zero or two (instead of one) deviations at the decision sequence level, or four (instead of three) out of four rounds at subject level.

[11] In our computations of consistency scores we exclude (i) decisions where more than two policies coincide, and (ii) decision sequences in which subjects are observed at least once to do the opposite of the policy prediction, for example if the policy is to increase the threshold, and the subject decreases the threshold instead. Neither of these exclusion criteria results in substantive changes to the matching process. However, by leaving these decisions out the consistency scores have a wider range and hence are more interpretable.

Figure 7 Study 1: Learning in Each Task



determine whether a policy is a good descriptor we then use a cutoff equal to the 70^{th} percentile of all scores for that task. If no policy achieves a fit that is above that cutoff, we do not have a good description of the decision sequence and classify it as "Unidentified".[12]

Figure 6 shows the static/forward-looking/unidentifiable breakdown for each task. For all three tasks, most subjects are taking a systematic, identifiable approach - with only between 7.6% and 16.9% of subjects being unidentifiable. We also see a clear split in each task between static and forward-looking approaches to the problem. In the Technology Adoption task subjects are overwhelmingly using forward-looking policies: approximately 83% of subjects use a forward-looking strategy, with less than 6% using a static rule. We see a similar pattern in the Capacity Allocation problem. Approximately 70% of subjects are forward-looking, versus 13% are static. While there are significantly fewer forward-looking subjects in the Capacity Allocation task relative to Technology Adoption (test of proportions: two sided $p = 0.020$), the difference is relatively small in magnitude. By contrast, subjects in the Search task show a very different pattern - approximately 63% of them use a static rule, and only 29% use a forward-looking strategy ($p \ll 0.01$ versus both other tasks). This provides initial evidence that stopping problems may be structurally different from other dynamic problems in the strategies they prompt, and that other task characteristics such as stochastic versus deterministic do not seem to lead to such a sharp divide in behavior.

Figure 7 further breaks down each task by round, allowing us to consider two additional questions related to learning. First, the share of unidentifiable decision sequences decreases in all three tasks (Non-parametric trend tests: $p = 0.000$, 0.040, 0.000). The strong drop after the first round suggests some initial experimentation and a convergence to approaches that are either static or forward-looking. Second, there is some learning towards forward-looking strategies in Technology Adoption

[12] The specific value of the cutoff was chosen in an iterative way, ensuring that (a) the number of unidentified sequences was not too large in any task, and (b) subjects classified as unidentified earned significantly less than subjects identified as forward-looking. Alternative values of the cutoff do not change any of our main results, since the ranking of the policies would be preserved with a lower/higher cutoff.

(Non-parametric trend test: $p = 0.034$), and towards static policies in Capacity Allocation (Non-parametric trend test: $p = 0.020$). Overall however, the categorization as static or forward-looking is remarkably consistent across rounds, with 64% to 88% of subjects remaining in the same category in at least 3 of the 4 rounds. This suggests that when a subject approaches a problem they largely stick with using a static or a forward-looking approach even as they gain more experience.

Before proceeding to more detailed analysis we briefly discuss whether the usage of static/forward-looking approaches is related to the subjects' earnings. As one may expect, deviations from the optimal policy are negatively correlated with task earnings in all three tasks. We find the strongest correlation in the deterministic Technology Adoption task (Technology Adoption: $\rho = -0.940, p \ll 0.01$; Capacity Allocation: $\rho = -0.571, p \ll 0.01$; Product Search: $\rho = -0.789, p \ll 0.01$). Indeed, subjects classified as optimal in Fig. 6 earned 2(5%) more than static subjects and 11(16%) more than unidentified subjects in the Technology Adoption (Capacity Allocation) tasks. In contrast, subjects classified as optimal earned only about 1% more than static and 5% more than unidentified in the Search task. To ensure that our results are not driven by the flatness of the payoff function in the Search problem, one of the treatments in Study 2 ensures that subjects would earn significantly lower payoffs if they deviate from the optimal policy.

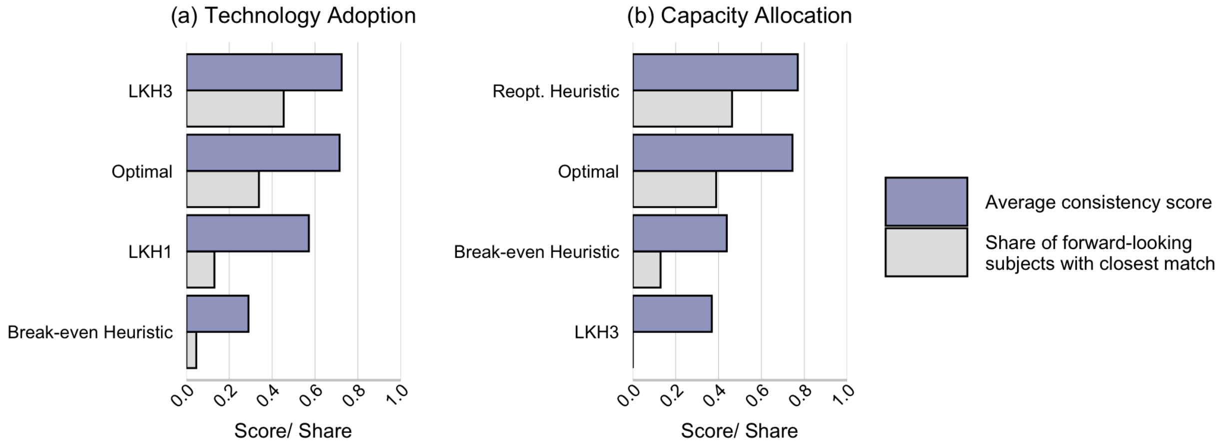
5.2. Decision Rules Used in Each Task

Technology Adoption and Capacity Allocation For the Technology Adoption and Capacity Allocation task, we have identified that subjects are primarily forward-looking. Our next step is to further unpack behaviors in these two tasks to uncover which particular policies subjects are using. Figure 8 examines the four policies that best describe the choices of forward-looking subjects. The dark blue bars measure the descriptive accuracy of the top four policies in each task by looking at the average consistency score of each policy. The light gray bars show the share of forward-looking subjects that have each policy as their best match (i.e. highest consistency score).[13]

For both problems, the optimal policy is the second-best descriptor policy. For the Technology Adoption problem, the optimal policy and look-ahead-3 (LKH3) policy both have a consistency score near 0.72, with the two policies accounting for about three quarters of forward-looking subjects. This is quite intuitive: the Technology Adoption problem is the simplest problem, making it feasible for subjects to have decisions close to the optimum. Further, a large number of subjects

[13] A substantial number of subjects (32% in Technology Adoption and 52% in Capacity Allocation) could be classified as "mixed types" because more than one policy matches their strategy well (less than 5 percentage point difference in consistency scores between the top two policies). For the majority of these subjects the two top policies show substantive overlaps on the sample path chosen by the subject. For example, for subjects who are mixed optimal and LKH3 in Technology Adoption, there is a 70% overlap in predictions on the sample path. For subjects who are mixed optimal and reoptimization in Capacity Allocation, there is a 75% overlap in predictions. Hence, mixed types are subjects that are well described by two similar policies, rather than splitting the difference between two quite different policies.

Figure 8 Study 1: Forward-looking Policies



Note. Four best descriptor policies for each task are displayed. Dark bars show the average consistency score for each policy. Light bars show % subjects with each policy as their best match (i.e. highest consistency score).

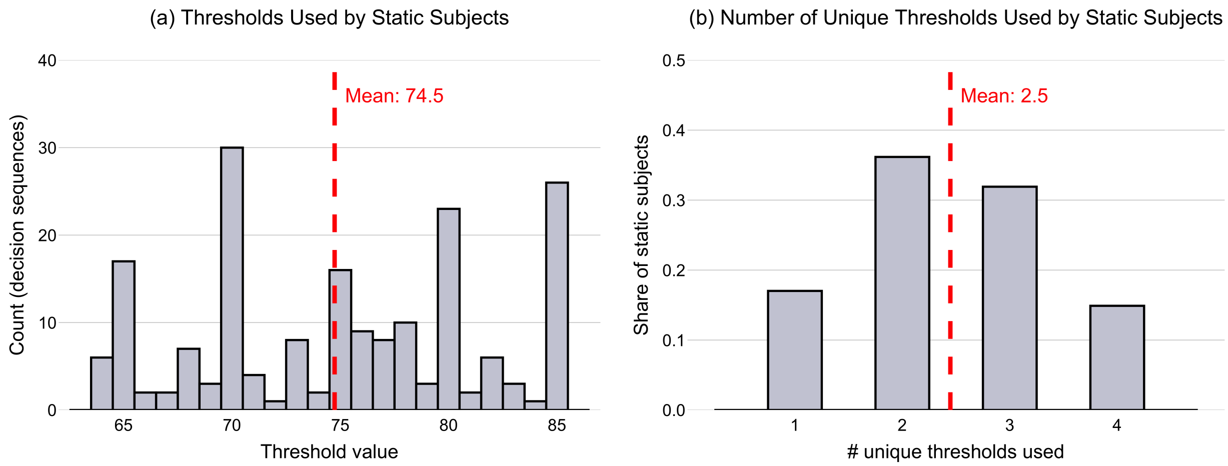
adopt look-ahead policies. The Technology Adoption is a deterministic problem where subjects know in advance the future path of values, making a look-ahead policy straightforward to compute and useful. The last displayed policy, break-even heuristic does substantially worse, with a consistency score well below 0.50, covering less than 5% of forward-looking subjects.[14]

For the Capacity Allocation problem we find that the two policies that best describe subjects are the reoptimization heuristic and the optimal policy. The consistency scores of these two policies are similar; however a larger portion of subjects have the reoptimization heuristic as their best match. This matches the results of Leider and Şahin (2014), who also found that the reoptimization heuristic and optimal policy did a good job of describing subject choices, but that between the two, subjects were more likely to be closer to reoptimization. Further, there is a substantial drop off in explanatory power after the top two policies, with very few subjects being best described by the break-even heuristics or the LKH3 policy.

Together, these results show both the strengths and the limitations of modeling decision-makers as using the standard optimal policy. On one hand, the optimal policy accurately describes a substantial number of subjects for both tasks. On the other hand, the optimal policy is the best descriptor for less than half of the forward-looking subjects (and hence even fewer of the whole subject pool). For both tasks there is another policy that is capturing something about subject behavior that the optimal policy is missing. Hence, a richer model that captures a more diverse set of decision strategies may be a more accurate representation of decision-making. In Section 7 we provide an illustrative example of how accounting for this heterogeneity can increase firm revenues.

[14] As one may expect, LKH2 has a consistency score in between LKH1 and LKH3. However, LKH2 is the *closest* match for only 2% of the subjects.

Figure 9 Study 1: Static Decision Rules



Note. Frequency of static rules in the Search task. Left panel shows the frequency of each rule across rounds. Right panel shows the frequency of subjects using one to four unique thresholds.

Robustness Check: Maximum Likelihood Estimation Our analysis so far has been based on matching each subject to the policy with the highest consistency score (method described in Section 5.1). In Appendix C we present the results from an alternative method based on policy assignments that jointly maximize the likelihood of the observed choices (i.e. a Maximum Likelihood Estimation approach). The MLE analysis suggests that the split among the policies is quite robust, i.e., the ranking among the four top policies remains almost unchanged. Notably, the MLE results confirm that in both tasks the optimal policy is the second best predictor and matches approximately a quarter to a third of non-static subjects. The policy with the largest share continues to be LKH3 for Technology Adoption and the reoptimization heuristic for Capacity Allocation.[15]

Product Search In Section 5.1 we saw that the majority of subjects are using a static decision rule in the Product Search task. Indeed, almost all of these subjects are using a fixed stopping value (i.e. a value-based static rule, as opposed to a time-based static rule). Since the static thresholds can be different for each subject and round, we are interested in the specific threshold values. The left panel of Figure 9 shows the distribution of the thresholds for the subset of subjects identified as static. The right panel shows how many unique thresholds each static subject used across rounds. Several observations are in order. First, most subjects use value thresholds that are multiples of five. Second, most subjects use more than one value threshold throughout the experiment. This suggests that they are experimenting with different thresholds across rounds. Lastly, it is not the case that static Subjects are primarily choosing the value threshold that will yield the highest

[15] We do not use MLE to estimate the proportion of policies in Product Search because most subjects are static in that task, and the number of possible static policies is quite large relative to the number of subjects. Many subjects essentially use a unique static threshold, as illustrated in Figure 9.

Table 2 Correlations Between Optimality and Individual Variables and Across Tasks

	Average deviation from optimal			Task earnings		
	Technology Adoption	Capacity Allocation	Product Search	Technology Adoption	Capacity Allocation	Product Search
Demographics						
Age	-0.063	0.061	-0.084	-0.067	0.037	-0.084
Gender (f=1)	0.044	0.194*	-0.086	-0.072	-0.111	0.048
Engineering Major	-0.233***	-0.246**	-0.082	0.260***	0.142	0.145
CRT score	-0.307***	-0.096	-0.183	0.302***	0.329**	0.112
Risk-aversion	0.001	0.072	-0.016	-0.060	0.064	-0.077
Cross-task correlations		Capacity Allocation	Product Search		Capacity Allocation	Product Search
	Technology Adoption	0.099	0.280**	Technology Adoption	0.306***	0.233**

Note. Pearson coefficients are displayed. *, **, *** denote significance level of 0.1, 0.05, 0.01, respectively. CRT (Cognitive Reflection Test) score is the number of correct responses (out of three) on the standard CRT test. Risk aversion is measured using the Multiple Price List method (Holt and Laury 2002).

average payoff. For the parameter values used in the experiment, given that a subject is using a static value threshold a higher threshold will yield a higher average payoff. While we do see a large number of subjects use the highest threshold of 85, the most common thresholds is 70. There is not an overall monotonic trend of larger thresholds being more common, nor is there a systematic trend for thresholds to either increase or decrease across rounds ($p = 0.523$). This suggests that static subjects are not systematically identifying and using the “best” static threshold (nor are they learning to). Rather, they appear to be using a simpler (and cognitively cheaper) approach of choosing focal, round threshold values throughout the task.[16]

5.3. Individual Markers of Performance and Cross-task Comparisons

Our last research question is what makes someone more or less forward-looking, and whether people’s strategies are similar across tasks. Our analysis is summarized in Table 2. The top panel of Table 2 shows Pearson correlation coefficients and their significance level for the relationship between individual marker variables and subject-level adherence to optimal policies and subject earnings. The bottom panel shows cross-task comparisons.

Several observations are in order. First, both policy adoption and performance in the Technology Adoption task are significantly related to college major and to the Cognitive Reflection Test (CRT) score (at $p < 0.01$). This suggests that performance on that task can be predicted reasonably well

[16] It is plausible that the size of the threshold value is aligned with the subject’s risk appetite, i.e. more risk-averse subjects may choose to stop sooner than more risk-loving ones. To test this conjecture we examined the correlation between thresholds and our measure of risk preferences (elicited after the dynamic task was completed), but found the correlation to be not significant.

by the analytical capabilities and by the analytical training of the participant. Similarly, policy choice in the Capacity Allocation task is directionally related to both major and CRT score, but not all relationships are significant at $p < 0.05$. In contrast, policy adoption and earnings in the Product Search task appear to be unrelated to the demographic variables.[17] Lastly, the bottom panel of Table 2 suggests that adherence to optimal policies is significantly correlated for Technology Adoption and Search ($p = 0.012$) but not for the other pair of tasks; however, earnings are correlated for each pair of tasks ($p < 0.01$, $p = 0.038$).

How do these results inform our understanding of behavior in dynamic decision problems? First, policy adoption may be a function of one's analytic capabilities for simpler tasks, in which the optimum can be computed with some effort (as is the case in our deterministic Technology Adoption task). However, policies are less strongly related to simple measures of analytic fluency in more difficult tasks that involve trade-offs under uncertainty (as is the case in our Capacity Allocation and Product Search task). Second, the extent to which people adopt similar strategies across tasks varies. When multiple well-performing policies are available, people may adopt different policies in different tasks. For example, reoptimization heuristic is both popular and quite profitable in Capacity Allocation, and is used by many people who use an optimal approach in the Technology Adoption task. That is, someone may be using two different forward-looking policies in two different tasks. Lastly, while policies may differ, earnings are correlated across tasks; that is, performance on one dynamic task may be indicative of performance on another one.

6. Study 2

The purpose of Study 2 was to further unpack behaviors in dynamic environments, specifically the differences between the (predominantly static) decision rules used in the Product Search problem and the (predominantly forward-looking) rules in other dynamic problems.

To determine what drives this result we explore several potential explanations. One possibility are the differences in reward functions (payoff accumulation). Product Search is a stopping problem, and thus implies a single large payout once the decision-maker stops; in contrast, the other dynamic problems involve a cumulative payoff function that collects many small payoffs from multiple decisions. The differences in the accumulation of payoffs may prompt different approaches to solving the problem. Another possibility is that the differences are caused by the elicitation process/action space of the task (See Figures B.1-B.3 in the Appendix). In the Search problem we asked subjects to enter an integer number, while the remaining tasks involved either yes/no decisions (Technology Adoption task), or check boxes (Capacity Allocation task). Finally, expected payoffs in stopping

[17] We also examine correlations between performance and other (non-Engineering) college majors and find no significant relationships with either of the three tasks.

problems can be relatively insensitive to deviations from the optimal policy (Seale and Rapoport 1997). Thus, a further possibility is that the simpler decision rules in Search are a result of a lack of a strong financial incentive.

These mechanisms are investigated in Study 2 in three between-subject treatments. As in Study 1, in all three treatments subjects work on two tasks (see Table 1). In all treatments participants play four rounds of a new stopping problem, the "Machine Replacement" problem. As in our previous tasks, in the experiment we use a framing more familiar to subjects - deciding how long to drive a used car before selling it. In this problem subjects iteratively choose between two alternatives: to continue driving a used car (earning a small payment but risking the car breaking), or to stop and sell the car (earning a large payment and ending the task). The basic structure of this problem is similar to the classic finite horizon version of the machine or equipment replacement problem, see for example, Meyer (1971), Sethi and Chand (1979).[18] The per-period payoff for driving and the payoff for selling are fixed throughout the horizon. The probability of the car breaking increases over the horizon, and if the car breaks the opportunity to sell the car is lost. Across the four rounds, the decision horizon varies with either $T = 15$ or $T = 20$.[19]

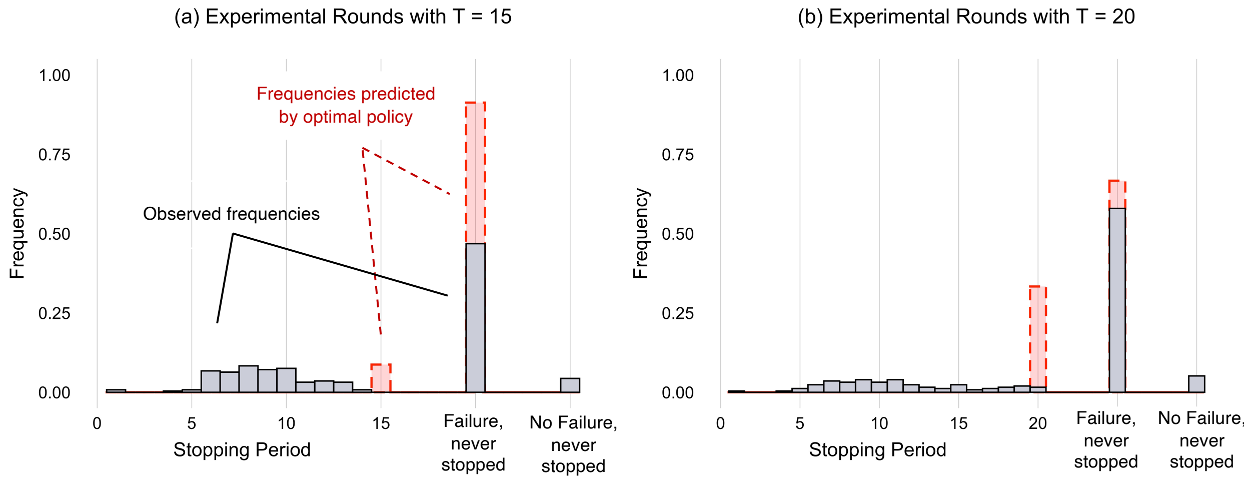
This Machine Replacement task shares the stopping problem structure with the Product Search task from Study 1, but has incremental payoffs, similar to the Technology Adoption and Capacity Allocation tasks. If participants use static decision rules in the Machine Replacement problem, then the primary driver of policy sophistication is the nature of stopping problems versus other dynamic decision problems. If not, then the primary driver is the accumulation of payoffs (single payoff vs. many small payoffs).

In addition to the Machine Replacement problem, subjects played four rounds of a variation of one of the games used in Study 1, designed to test possible alternate mechanisms. As before, the task sequence was reversed in half of the sessions to control for order effects. In the first treatment subjects worked on a version of the Product Search problem, in which their threshold was elicited using yes/no check marks, instead of having to enter the exact threshold for acceptance. In the second treatment, subjects worked on the original Product Search problem, parametrized to yield

[18] Basten and Tan (2021) use a similar experimental task with a machine maintenance framing, with varying costs for machine failure and rates of machine degradation.

[19] In designing the Machine Replacement problem we focused on having a clean separation between sophisticated and non-sophisticated heuristics (to evaluate the robustness of our results from Study 1). However, this means that we are not able to distinguish between many of the forward-looking policies. Specifically, all three limited look-ahead policies coincide with the optimal policy, as does the reoptimization heuristic. This is because there is no changing state variable other than the current time period, hence the myopia of the limited look-ahead and the simplification of reoptimization do not lead to different policies.

Figure 10 Study 2: Frequency of Stopping in Machine Replacement



Note. Histograms show the share of subject-rounds with each possible outcome: having a failure before selling, selling in a given period, or not selling. The optimal strategy is to only sell in the last period; the dotted red boxes denote the expected frequencies under the optimal policy given the realizations of random quantities in each session.

a larger gap between the payoff resulting from the optimal and the next best solution.[20] In the third treatment subjects worked on a revised version of the Capacity Allocation task, in which the accept/reject decision was elicited via a continuous threshold number entered by the subject, as opposed to a list of yes/no check boxes used in Study 1. The parameters and realizations of random quantities were kept similar with the corresponding treatment in Study 1.

A total of 126 subjects were recruited to participate in Study 2 (See Table 1 for the exact subject numbers). We recruited subjects from the same student subject pool who did not participate in Study 1. No significant differences in age, gender or college major were found between Study 1 and Study 2. The setup, instructions and experimental procedures in Study 2 were similar to Study 1 and are omitted for brevity. The tasks, objectives and exact participant numbers are summarized in Table 1. The parameters used in Study 2, as well as screen shots of the decision screens are reproduced in Appendix D.

6.1. Experimental results

Machine Replacement Figure 10 shows the distribution of outcomes in the Machine Replacement problem. For the parameters used, the optimal policy is to continue driving until the last

[20] Specifically, if a decision-maker followed the optimal policy for this new parametrization of the problem, their expected earnings would be 2.09% higher than the ex ante best performing static policy, and 21.25% higher than an average across the five most common static policies used in Study 1. This is a substantially larger difference in payoffs between more versus less sophisticated policies than in Study 1. For the original parametrization, the optimal policy gives an expected payoff 0.01% higher than the ex ante best static policy, and 3.57% higher than the average across the five most common static policies. Additionally, to further increase incentive strength we also increased the point-to-dollar conversion rate by a factor of five.

period, so participants following the optimal policy would either experience a failure, or sell in the last period. The red shaded dashed bars denote the expected number of failures if all participants followed the optimal policy, and the expected number of sales in the last period. We see that instead approximately half of the subjects are choosing to stop sooner than optimal - selling in earlier periods, with the mean sale period occurring near the mid-point of the task.

We now turn to identifying the decision policies in the Machine Replacement problem. The identification is quite straightforward because the optimal actions are unambiguous: subjects should sell only in the final period; however, the data are censored because we often observe a failure before the last period and hence may not know what a subject would have done in the later stages of the problem. We therefore identify strategies by looking at behavior *across* all four rounds of the task, such that at least some rounds yield a (near-)complete sequence of the subjects' decisions.

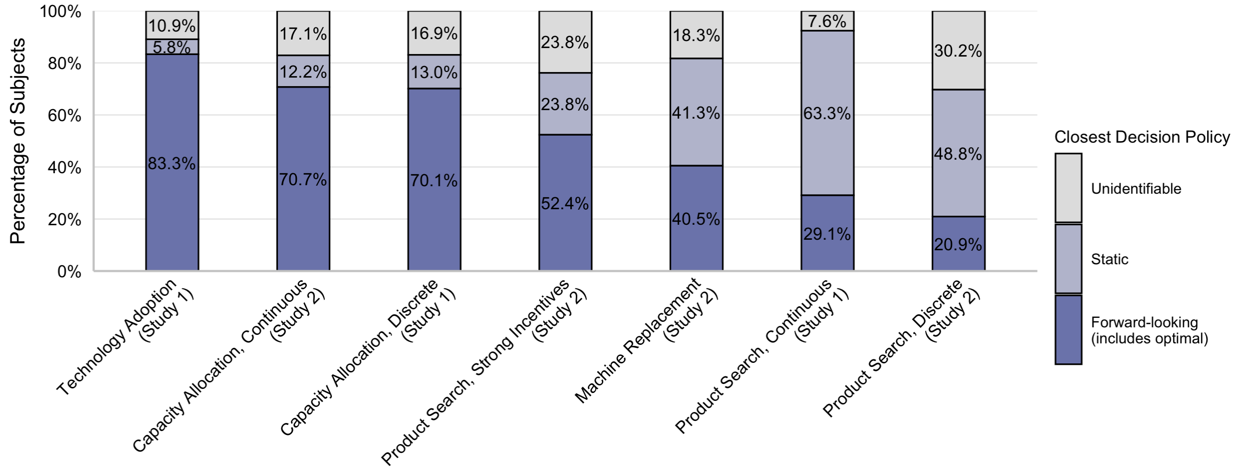
We identify subjects as "optimal" if they have at least one sale in the final period, and no sales in an earlier period. This (conservatively) identifies 10% of subjects as acting consistent with the optimal policy. Subjects' whose decisions are closer to the optimal policy (as measured by the consistency score) do earn significantly higher earnings (correlation with earnings: $\rho = 0.516, p < 0.01$). Static policies are identified for participants that consistently choose to stop and sell at a given period (or percent of the task length), or at a given failure probability. We identify 41% of subjects as using such a static decision rule. For most static subjects we cannot distinguish between having a time-based or failure probability-based rule, however for those that can be distinguished there are more than 2.5 times as many subjects following a probability-based rather than a time-based rule.[21] There are another 31% of subjects with at least two informative sequences (i.e. sales and/or failures later than observed sales) whom we can rule out as either static or optimal but who increase their average stopping period between the $T = 15$ and the $T = 20$ rounds. While this is broadly consistent with the break-even policy, only 1.6% of subjects are observed choosing to sell in a period close to the predicted round.[22] Finally, there are 18% of subjects for whom we do not have enough data to make any clear categorization. Overall, the high share of static subjects and low share of optimal subjects is more consistent with the pattern observed in the Product Search task of Study 1 than with the Technology Adoption or Capacity Allocation tasks.

Policy Comparison Between Studies Figure 11 shows the share of subjects identified as static or as forward-looking in each task of the experiment (Study 1+2). Several observations are in order (all statistical comparisons are based on two-sided tests of proportions).

[21] We note that, as in the Search problem, subjects are not choosing the "best" static policy (which would be to continue until the last period), but are choosing to stop much earlier.

[22] The only non-optimal, forward-looking policy that is separately identifiable in this task is the break-even policy, which predicts stopping in period 8 for $T = 15$ and in period 18 for $T = 20$.

Figure 11 Study 1 and Study 2: Policy Sophistication



First, eliciting choices using a continuous threshold or discrete decisions makes very little difference. The two versions of the Capacity Allocation problem (columns 2 and 3 in Figure 11 have very similar strategy profiles, with few static subjects and a high share of forward-looking subjects (n.s. for both comparisons). Similarly, both versions of the Product Search task have a similar strategy profile, with quite low shares of forward-looking subjects (n.s. for both comparisons).[23]

Second, the strong incentives parametrization of the Product Search task does shift subjects towards more sophisticated strategies. Specifically, we observe less than half as many static subjects as in the original Product Search task ($p \ll 0.01$), and more than twice as many forward looking subjects ($p = 0.012$). Compared to subjects in the original task, subjects in the strong incentives task were less likely to set a stopping threshold above the possible realizations (guaranteeing not to stop) deep into the task. In the strong incentives treatment, period 6 is the latest period where more than half of subjects set a threshold to guarantee continuing, while in the original treatment this is true for the first 10 periods. Subjects in the strong incentives treatment were also more likely to make small to medium sized changes to their stopping threshold (relative to the possible value realizations), while in the original task subjects tended to make either no change, or a large change. As we intended, there was also a stronger association between strategy sophistication and earnings than in the original task. Specifically, subjects' average deviation from the optimal policy was strongly negatively correlated with earnings ($\rho = -0.75, p \ll 0.01$); subjects identified as forward-looking earned 11.8% more than subjects identified as static (Rank sum test, $p = 0.049$).

[23] The high share of unidentified subjects in Discrete Search is due to many subjects stopping immediately in the first stage of the search (27.9 % of all decision sequences), so that identification relies on fewer observations per subject. This is because with the discrete format, the smallest possible probability of acceptance, if the subjects chooses to accept any values, is 20% (since there are five discrete alternatives). Thus the data are more "lumpy" with discrete elicitation relative to the continuous format.

While sophistication increases with incentive strength, we still see all of the stopping problems leading to less sophisticated approaches than each of the non-stopping problems. Both the original and discrete Product Search tasks have both significantly more static subjects and significantly fewer forward-looking subjects than each of the three non-stopping problems ($p < 0.01$ for all comparisons). The Machine Replacement problem has significantly fewer forward-looking subjects ($p < 0.032$) and significantly more static subjects ($p < 0.01$) than all three non-stopping problems. The strong incentives Product Search problem has fewer forward-looking subjects relative to Technology Adoption ($p < 0.01$), although the reduction relative to the other non-stopping tasks is marginally significant ($p = 0.054$ and $p = 0.086$). Hence we conclude that while certain features of the task can lead to some variations, notably the monetary returns to using forward-looking policies, there is a robust tendency for stopping problems to lead subjects to choose simpler decision rules.

Lastly, comparing subjects' strategies and earnings across the two tasks, we find quite similar results to Study 1. The correlation between a subject's deviation from optimality between Machine Replacement and the other tasks are all quite small and not statistically significant ($\rho < 0.20$ and $p > 0.20$ for all). However, subjects' earnings on the Machine Replacement task are significantly correlated with both the strong incentives Product Search task ($\rho = 0.37, p = 0.017$) and the continuous Capacity Allocation task ($\rho = 0.48, p < 0.01$). The earnings correlation with the discrete Product Search is essentially zero ($\rho = 0.01, p = 0.948$). Thus, as in Study 1 there is a relationship between a subjects' ability to achieve high earnings across tasks, but not necessarily the sophistication of the strategies they use to do so.

7. Discussion and Implications

In this section we present a broader discussion of our results and offer some implications, both for researchers interested in developing more descriptive models of dynamic decision-making, and for managers designing firm policies for consumers who face dynamic trade-offs.

Stopping vs. non-stopping problems Our first result is the striking difference in policy adoption between different dynamic decision problems. Stopping problems like our Product Search and Machine Replacement task lead most subjects to static decision rules, while non-stopping problems lead to more sophisticated forward-looking policies.

A natural question is what causes this difference. While our data do not directly speak to this, one speculative possibility seems both plausible and intuitive. Stopping problems have one *focal* decision point (the point of stopping), and that may prompt people to focus on a single key outcome: what would be a good target value to stop at in Product Search, or what breakdown percentage would be sufficiently high to prompt a sale in the Machine Replacement task? Focusing on the desired end state may predispose people towards static decision rules that achieve that state. By contrast, other dynamic problems involve many similarly important decisions (for example none of the ten uses of the free meals in the Capacity Allocation problem is likely to loom larger in importance than any other). This may prompt people to take a more holistic view on the problem, i.e., a focus on how to use their resources well, rather than any particular end state.

Optimal policy fit Our second result is that even for non-stopping problems simpler heuristics may be a better fit for many subjects than the optimal policy. Indeed, less than half of the forward-looking subjects (and hence even fewer of the whole subject pool) have the optimal policy as their *best* descriptor. Models of behavior that try to explicitly account for this heterogeneity may be more realistic, and may also help firms increase profitability. The following example illustrates the consequences of modeling a heterogeneous market of consumers as a single, optimal type.

Consider a firm offering a three part tariff (fixed fee, free units, fee per unit after free units are used) to sell its services. The firm can be, for example, a car dealer designing the terms of a lease on a new car, or a gym offering a new subscription for classes. Any consumer that accepts the terms of the contract and pays the fixed fee faces a sequence of decisions similar to our Capacity Allocation task. If the firm believes that everyone in the market is an optimizer, it should offer the following three part contract: a fixed fee of \$136, 5 free units and a per unit fee of \$6 (Column 1 in Table 3). Using a more realistic (given our data) two type split of consumers into static (30%) and reoptimization heuristic (70%) types, this contract would generate an expected total revenue of \$130 as only reoptimization types would purchase this contract. Similar calculations lead to an expected total revenue of \$189 (\$141) if the firm wrongfully assumes that the market consists wholly of static or reoptimization types (Columns 2 and 3 of Table 3). On the other hand, correctly anticipating the presence of both types leads to a menu of two incentive compatible contracts designed for each type, resulting in average total revenue of \$194. In this example, the revenue increase for offering a menu designed for two types, instead of assuming a single type, can be as high as 49% (last row of Table 3). Overall, this example illustrates that building models with more accurate and empirically grounded assumptions of the nature and heterogeneity of consumer decision types can lead to substantially different outcomes. Conversely, acting as if all consumers are optimal, or even that all consumers are forward-looking, can lead to foregone revenue opportunities.

Cross task correlations and individual markers Our last main result is that subjects seem to differ in their general ability to perform well at dynamic decision problems, as measured by the correlation in earnings across tasks. However, this is not reflected in a consistent ability to act close to the optimal policy. High-performing subjects seem to be performing well using different strategies in different tasks, rather than consistently using a sophisticated strategy close to the

Table 3 Revenue Management Application Example

	Contract optimized for single type			Contract menu	
	Optimal	Static	Reopt.	Static	Reopt.
Contract parameters:					
Fixed fee	\$136	\$133	\$193	\$133	\$183
# free units	5	5	15	5	15
Per unit fee	\$6	\$6	\$6	\$6	\$12
Total expected revenue (assuming 30/70 type split)	\$130	\$189	\$141	\$194	
Increase in expected revenue (From offering the menu)	49.01%	2.52%	37.40%	-	

Note. Calculations assume that there are 30% Static and 70% Re-optimization types in the market, that the firm chooses between 5 and 15 free units and chooses a per unit fee in $[0, \infty)$. Unit value distribution, horizon length and the remaining parameters are the same as in the Capacity Allocation task in Study 1, with a 1-to-1 point to dollar conversion rate (See Appendix A). Dollar values are rounded to the nearest integer.

optimal policy. Taken together, these results suggest that strategy choice in dynamic environments depends on the features of the problem: people may base their approach on the (cognitive) cost and the monetary benefit of using the optimal policy relative to other decision rules.

Finally, performance on a basic measure of analytic intelligence, the Cognitive Reflection Test, was significantly related to performance in the simplest task (Technology Adoption), but was less predictive of performance in more complex tasks. This result is somewhat similar to the "Threshold hypothesis" (Sawyer 2011, Jauk et al. 2013) – a classic theory in the creativity literature stating that the relationship between intelligence and creativity is linear for simpler creative tasks but less pronounced for higher-level creative endeavors. Similarly, analytic intelligence may be a necessary, but not a sufficient condition for being able to correctly solve more complex dynamic decision problems.

8. Concluding remarks

Many important operational settings involve dynamic decision problems, where choices have both immediate consequences and implications for the future. Making sophisticated choices, and particularly near-optimal choices in such settings is often complicated and cognitively demanding. Past behavioral research has explored human decision-making across a range of settings, identifying common heuristics and deviations from optimal choice. However, to date there is not yet a boundedly rational framework to describe typical behavior across a range of settings with the breadth and generality as with, for example, behavioral models of decision-making under uncertainty (prospect theory and reference-dependent preferences), time discounting, or social preferences (models of inequality aversion and social norms). This paper sets an initial step towards such a framework, by examining choices across four distinct dynamic decision contexts, designed to identify and distinguish a common set of decision policies of varying sophistication.

We identify several behavioral regularities that can form the basis for future development of a general behavioral model of dynamic decision-making. First, decision sophistication is sharply different between stopping and non-stopping problems. Subjects are substantially more likely to use simple static decision rules for stopping problems and to use sophisticated forward-looking heuristics (including some acting nearly-optimally) in non-stopping problems. Second, for stopping problems, the most common static policies relate to salient features of the problem, such as the apartment value in the Product Search problem and the failure risk in the Machine Replacement problem. For non-stopping problems, the look-ahead policy is the most common heuristic used in the simpler Technology Adoption problem, while the reoptimization heuristic is the most common for the Capacity Allocation problem. Finally, while subjects' earnings are correlated across tasks, suggesting some general ability to effectively solve dynamic decision problems, policy usage is not correlated across tasks.

These results suggest several avenues for future research. First, with the available data we can only speculate about the psychological drivers of the difference in approaches between stopping and non-stopping problems. One plausible explanation is that stopping problems are characterized by having one important key decision (when to stop) while non-stopping problems typically have many decisions of similar importance. Future experiments could verify this intuition and directly manipulate/nudge decision-makers towards being more outcome-focused or more holistic in their approach. Second, our study offers only preliminary insights into the types of cognitive abilities that enable subjects to perform systematically better across tasks. Future studies may be able to identify better predictors of performance by supplementing the Cognitive Reflection Test with other diagnostic measures (such as pattern completion and working memory tasks used by Ballinger et al. (2003, 2011)). Finally, our experiments focused on one aspect of complexity of a dynamic decision problem: the need to be forward-looking and consider both the present and future consequences of a choice. Future research could explore general features of how decision-makers construe and simplify other aspects of complexity - such as the stochastic evolution between problem states, or the construction of a consideration set from a complex set of actions.

References

Anderhub, Vital, Werner Gäuth, Wieland Mäuller, Martin Strobel. 2000. An experimental analysis of intertemporal allocation behavior. *Experimental Economics* **3**(2) 137–152.

Ascarza, Eva, Anja Lambrecht, Naufel Vilcassim. 2012. When talk is "free": The effect of tariff structure on usage under two-and three-part tariffs. *Journal of Marketing Research* **49**(6) 882–899.

Aviv, Yossi, Amit Pazgal. 2008. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & service operations management* **10**(3) 339–359.

Balcer, Yves, Steven A Lippman. 1984. Technological expectations and adoption of improved technology. *Journal of Economic Theory* **34**(2) 292–318.

Ballinger, T Parker, Eric Hudson, Leonie Karkoviata, Nathaniel T Wilcox. 2011. Saving behavior and cognitive abilities. *Experimental Economics* **14**(3) 349–374.

Ballinger, T Parker, Michael G Palumbo, Nathaniel T Wilcox. 2003. Precautionary saving and social learning across generations: an experiment. *The Economic Journal* **113**(490) 920–947.

Basten, Rob, Lijia Tan. 2021. Wait and see, or pay now? on how people decide to pay a cost to avoid a loss. *working paper* .

Bearden, J Neil, Ryan O Murphy, Amnon Rapoport. 2008. Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management* **10**(4) 625–636.

Bearden, J. Neil, Amnon Rapoport, Ryan O. Murphy. 2006. Sequential observation and selection with rank-dependent payoffs: An experimental study. *Management Science* **52**(9) 1437–1449.

Bendoly, Elliot. 2011. Linking task conditions to physiology and judgment errors in rm systems. *Production and Operations Management* **20**(6) 860–876.

Bendoly, Elliot. 2013. Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management* **31**(1-2) 62–71.

Bertsekas, Dimitri P. 1995. *Dynamic programming and optimal control*, vol. 1. Athena scientific Belmont, MA.

Bolton, Gary E, Axel Ockenfels. 2000. Erc: A theory of equity, reciprocity, and competition. *American economic review* **90**(1) 166–193.

Borenstein, Severin. 2009. To what electricity price do consumers respond? residential demand elasticity under increasing-block pricing. *UC Berkeley Working Paper Series* **30** 95.

Brown, Alexander L, Zhikang Eric Chua, Colin F Camerer. 2009. Learning and visceral temptation in dynamic saving experiments. *The Quarterly Journal of Economics* **124**(1) 197–231.

Cachon, Gérard P, Martin A Lariviere. 1999. Capacity allocation using past sales: When to turn-and-earn. *Management science* **45**(5) 685–703.

Cachon, Gérard P, Robert Swinney. 2009. Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science* **55**(3) 497–511.

Cachon, Gérard P, Christian Terwiesch, Yi Xu. 2008. On the effects of consumer search and firm entry in a multiproduct competitive market. *Marketing Science* **27**(3) 461–473.

Caplin, Andrew, Mark Dean, Daniel Martin. 2011. Search and satisficing. *American Economic Review* **101**(7) 2899–2922.

Carbone, Enrica. 2006. Understanding intertemporal choices. *Applied Economics* **38**(8) 889–898.

Carbone, Enrica, John D Hey. 2004. The effect of unemployment on consumption: an experimental analysis. *The Economic Journal* **114**(497) 660–683.

Cornand, Camille, Frank Heinemann. 2019. Experiments in macroeconomics: methods and applications. *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing.

Cox, James C., Ronald L. Oaxaca. 1989. Laboratory experiments with a finite-horizon job-search model. *Journal of Risk and Uncertainty* **2** 301–330.

Dal Bó, Pedro, Guillaume R Fréchette. 2011. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review* **101**(1) 411–29.

Diehl, Ernst, John D Sterman. 1995. Effects of feedback complexity on dynamic decision making. *Organizational behavior and human decision processes* **62**(2) 198–215.

Duffy, John, Yue Li. 2019. Lifecycle consumption under different income profiles: Evidence and theory. *Journal of Economic Dynamics and Control* **104** 74–94.

Eriksson, Kimmo, Pontus Strimling. 2010. The devil is in the details: Incorrect intuitions in optimal search. *Journal of Economic Behavior & Organization* **75**(2) 338–347.

Fehr, Ernst, Klaus M Schmidt. 1999. A theory of fairness, competition, and cooperation. *The quarterly journal of economics* **114**(3) 817–868.

Feinberg, Fred M, Joel Huber. 1996. A theory of cutoff formation under imperfect information. *Management Science* **42**(1) 65–84.

Fischbacher, Urs. 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics* **10**(2) 171–178.

Gabaix, Xavier, David Laibson, Guillermo Moloche, Stephen Weinberg. 2006. Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review* **96**(4) 1043–1068.

Gallego, Guillermo, Robert Phillips, Özge Şahin. 2008. Strategic management of distressed inventory. *Production and Operations Management* **17**(4) 402–415.

Gopalakrishnan, Arun, Raghuram Iyengar, Robert J Meyer. 2015. Consumer dynamic usage allocation and learning under multipart tariffs. *Marketing Science* **34**(1) 116–133.

Grubb, Michael D, Matthew Osborne. 2015. Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review* **105**(1) 234–71.

Hanukov, Gabi, Shoshana Anily, Uri Yechiali. 2020. Ticket queues with regular and strategic customers. *Queueing Systems* **95**.

Hathaway, Brett, Evgeny Kagan, Maqbool Dada. 2021. The gatekeeper's dilemma:'when should i transfer this customer?'. *Available at SSRN 3617356* .

Hey, John D. 1982. Search for rules for search. *Journal of Economic Behavior & Organization* **3** 65–81.

Hey, John D. 1987. Still searching. *Journal of Economic Behavior & Organization* **8** 137–144.

Hey, John D, Valentino Dardanoni. 1988. Optimal consumption under uncertainty: An experimental investigation. *The Economic Journal* **98**(390) 105–116.

Ho, Teck-Hua, Sergei Savin, Christian Terwiesch. 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science* **48**(2) 187–206.

Holt, Charles A, Susan K Laury. 2002. Risk aversion and incentive effects. *American economic review* **92**(5) 1644–1655.

Jauk, Emanuel, Mathias Benedek, Beate Dunst, Aljoscha C Neubauer. 2013. The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence* **41**(4) 212–221.

Kim, Song-Hee, Jordan Tong. 2021. Admission control bias and path-dependent feedback under diagnosis uncertainty. *Available at SSRN 3830905* .

Kim, Song-Hee, Jordan Tong, Carol Peden. 2020. Admission control biases in hospital unit capacity management: How occupancy information hurdles and decision noise impact utilization. *Management Science* **66**(11) 5151–5170.

Kremer, Mirko, Francis de Vericourt. 2020. Ticket queues with regular and strategic customers. *Working paper* .

Laibson, David. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* **112**(2) 443–478.

Lee, Tak C, Marvin Hersh. 1993. A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation science* **27**(3) 252–265.

Leider, Stephen, Özge Şahin. 2014. Contracts, biases, and consumption of access services. *Management Science* **60**(9) 2198–2222.

Levin, Yuri, Jeff McGill, Mikhail Nediak. 2009. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management science* **55**(1) 32–46.

Li, Jun, Nelson Granados, Serguei Netessine. 2014. Are consumers strategic? structural estimation from the air-travel industry. *Management Science* **60**(9) 2114–2137.

Lin, Yen-Ting, Ali K Parlaktürk, Jayashankar M Swaminathan. 2018. Are strategic customers bad for a supply chain? *Manufacturing & Service Operations Management* **20**(3) 481–497.

Long, Xiaoyang, Javad Nasiry, Yaozhong Wu. 2020. A behavioral study on abandonment decisions in multistage projects. *Management Science* **66**(5) 1999–2016.

McGill, Jeffrey I, Garrett J Van Ryzin. 1999. Revenue management: Research overview and prospects. *Transportation science* **33**(2) 233–256.

Meissner, Thomas. 2016. Intertemporal consumption and debt aversion: an experimental study. *Experimental Economics* **19**(2) 281–298.

Meyer, Robert A. 1971. Equipment replacement under uncertainty. *Management Science* **17**(11) 750–758.

Moon, Philip, Andrew Martin. 1990. Better heuristics for economic search - experimental and simulation evidence. *Journal of Behavioral Decision Making* **3**(3) 175–193.

Noussair, Charles, Kenneth Matheny. 2000. An experimental study of decisions in dynamic optimization problems. *Economic Theory* **15**(2) 389–419.

Oprea, Ryan. 2020. What makes a rule complex? *American Economic Review* **110**(12) 3913–51.

Özer, Özalp, Yanchong Zheng. 2012. Behavioral issues in pricing management .

Özer, Özalp, Yanchong Zheng. 2016. Markdown or everyday low price? the role of behavioral motives. *Management Science* **62**(2) 326–346.

Palley, Asa, Mirko Kremer. 2014. Sequential search and learning from rank feedback: Theory and experimental evidence. *Management Science* **60**(10) 2525–2542.

Phipps, A.G., R.J. Meyer. 1985. Normative versus heuristic models of residential search behavior: an empirical comparison. *Environment and Planning A* **17** 761–776.

Porteus, Evan L. 1985. Investing in reduced setups in the eoq model. *Management Science* **31**(8) 998–1010.

Puterman, Martin L. 2014. *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Romero, Julian, Yaroslav Rosokha. 2018. Constructing strategies in the indefinitely repeated prisoner's dilemma game. *European Economic Review* **104** 185–219.

Sawyer, R Keith. 2011. *Explaining creativity: The science of human innovation*. Oxford university press.

Schotter, Andrew, Yale M. Braunstein. 1981. Economic search: An experimental study. *Economic Inquiry* **19**(1) 1–25.

Schunk, Daniel, Joachim Winter. 2009. The relationship between risk attitudes and heuristics in search tasks: A laboratory experiment. *Journal of Economic Behavior & Organization* **71** 347–360.

Seale, Darryl A, Amnon Rapoport. 1997. Sequential decision making with relative ranks: An experimental investigation of the "secretary problem". *Organizational behavior and human decision processes* **69**(3) 221–236.

Seale, Darryl A, Amnon Rapoport. 2000. Optimal stopping behavior with relative ranks: The secretary problem with unknown population size. *Journal of Behavioral Decision Making* **13** 391–411.

Sethi, Suresh, Suresh Chand. 1979. Planning horizon procedures for machine replacement models. *Management Science* **25**(2) 140–151.

Sonnemans, J. 1998. Strategies of search. *Journal of Economic Behavior & Organization* **35** 309–332.

Stahl, Dale O. 1989. Oligopolistic pricing with sequential consumer search. *The American Economic Review* 700–712.

Sterman, John D. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management science* **35**(3) 321–339.

Sutton, Richard S, Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

Tirole, Jean. 2010. *The theory of corporate finance*. Princeton University Press.

Tversky, Amos, Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *science* **185**(4157) 1124–1131.

Tversky, Amos, Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty* **5**(4) 297–323.

Van Mieghem, Jan A. 1999. Coordinating investment, production, and subcontracting. *Management Science* **45**(7) 954–971.

Zwick, Rami, Amnon Rapoport, Alison KC Lo, AV Muthukrishnan. 2003. Consumer sequential search: Not enough or too much? *Marketing Science* **22**(4) 503–519.

Appendix

A. Policies

In this Appendix, we state the optimal and the applicable non-optimal policies for each problem. Policies that are not applicable for certain tasks are omitted.

A.1. Optimal policies

Optimal Policy for Technology Adoption Problem We define $V_t(n)$ as the optimal total utility from t to T starting with technology version $n \leq t$ of the product:

$$
\begin{aligned}
V_t(n) &= \max\{\text{do nothing, adapt new technology}\} \\
V_t(n) &= \max\{u_n + V_{t+1}(n), u_t - p + V_{t+1}(t),\ t < T\} \\
V_T(n) &= \max\{u_n, u_T - p\}
\end{aligned}
$$

Since the problem is deterministic, it is straightforward to show that the decision-maker adopts the new technology at time t with starting technology n if the utility of the new technology is greater than a threshold $y_t(n)$. The threshold depends on the remaining schedule of technology improvements and price of adopting new technologies.

Optimal Policy for Capacity Allocation Problem We can write the dynamic programming formulation and the value function $V_t(k)$ as

$$
\begin{aligned}
V_t(k) &= E[\max\{\text{accept meal, reject meal}\}] \\
V_t(k) &= E[\max\{X + V_{t+1}(k-1), V_{t+1}(k)\}],\ 0 < k \leq C,\ t < T \\
V_t(0) &= tE[X-p]^+,\ t \leq T
\end{aligned}
$$

Leider and Şahin (2014) show that the optimal policy is a threshold policy where the student accepts the meal at time t if and only if the value of meal is greater than the opportunity cost of free meal at time t with k units of free meals: $X \geq V_t(k) - V_t(k-1)$.

Optimal Policy for Product Search Problem We can write the dynamic programming formulation and the value function $V_t(x)$ as

$$
\begin{aligned}
V_t(x) &= E[\max\{\text{rent the max utility (x) apartment , continue search}\}] \\
V_t(x) &= E[\max\{\max(X_t, x) - tc, V_{t+1}(x_t^{max})\}],\ t < T \\
V_T(x) &= E[\max\{\max(X_T, x) - cT, 0\}]
\end{aligned}
$$

where $x_t^{max} = \max(x_{t-1}^{max}, X_t)$.

It is straightforward to show that the optimal policy is a threshold policy where the student rents the apartment at time t if and only if $x_t^{max} \geq y_t$, where y_t is increasing in t (see, for example Cox and Oaxaca 1989, for a derivation).

A.2. Static Policies

Value-based Heuristic for Technology Adoption Problem Decision-maker adopts a new model in period t if utility improvement of technology introduced at time t (u_t) compared to the utility of laptop currently owned (s_t) is greater than a threshold q. Formally, we can state these types of policies as adopt new technology if and only if $u_t - s_t \geq q$.

Time-based Heuristic for Technology Adoption Problem Decision-maker adopts a new model every $n \leq T$ periods regardless of the utility of the new models introduced.

Value-based Heuristic for Capacity Allocation Problem A decision-maker may choose a static threshold q based on an ad-hoc rule. In this case, value-based heuristic takes the form: Consume meals that have value greater than q, do not consume otherwise. Alternatively, the decision-maker may choose the "best" (highest expected profit) static rule. If the decision-maker has k tokens at the beginning of the horizon to be consumed in T periods, then the best static value-based requires solving

$$\max_{q \leq p} E[X|X > q] \min(T\bar{F}(q), k) + E(X - p)^+ \frac{(T\bar{F}(q) - C)^+}{\bar{F}(q)}$$

Notice that in this heuristic approach, student balances the expected total number of meals and total reward received over T days with the expected overage cost paid to the provider. This heuristic replaces the random number of meals with their expectation, and the decision-maker chooses the meals that are above a predetermined state independent value. Solving the optimization problem gives a static threshold $q^S(k,T)$ to be used until all free units are used. When all free units are used, the decision-maker starts rationing meals based on the unit cost of the meal, p. The optimal static threshold is given by

$$q^S(k,T) = \min\{\min\{q|\bar{F}(q) \leq \frac{k}{T}\}, p\}.$$

Time Based Heuristic for Capacity Allocation Problem A time-based heuristic for Capacity Allocation implies consuming all meals independent of their values in the first n periods. As with the value-based static heuristic, n can be chosen "optimally" to obtain the best balance between under- and overconsumption.

A.3. Dynamic Policies

k-step-look-ahead Policies For all three problems, k-step-look-ahead policies can be obtained by modifying the optimal dynamic programming formulations stated above for T periods. Instead of solving the T-period problem, decision-maker solves the k-period dynamic program optimally at time $t \geq k$, and t-period dynamic program optimally at time $t < k$.

Reoptimization Heuristic for Product Search Problem In period $t = 1, 2, ...T - 1$, use the optimal static value-based rule (updating it depending on state and time to go).

Break-even Heuristic for Capacity Allocation Problem With k free meal tokens to be consumed in the remaining t periods and a consumption threshold of v or above, the decision-maker's total net utility is $\min((tP(V \geq v), k)E[V|V \geq v]$. With no free meals in period t, the decision-maker would consume meals of value p and above which would cost $p(tP(V \geq p)$. So, the break-even heuristic finds the certainty equivalent meal value that makes the net utility in t periods equivalent to the net cost if free tokens are not available, i.e., a certain per period utility improvement is expected so that free tokens pay off at least their cost if used until the end of the horizon. If target accrual rate is any value greater than zero for the remaining time, the policy can be stated as follows: Consume meal in period t if meal value is greater than the minimum v that solves $\min((tP(V \geq v), k)E[V|V \geq v] = ptP(V \geq p)$.

Break-even Heuristic for Product Search Problem In period t, if best apartment searched has value $x_t^{\max}$, and expected value of the next apartment that will be searched is $E[X_{t+1}]$, then the decision-maker stops searching if expected value from searching the next apartment is higher than the cost of search: $E[X_{t+1}] - x_t \geq c$.

B. Study 1: Experimental Instructions and Screen Shots

Technology Adoption Task

This part of the experiment will involve the Laptop Purchase Task. In this task you will make a series of purchase decisions. There will be 14 periods, and in each period a new laptop model will be introduced. Each laptop model has a performance score - the higher the score the more valuable is the laptop to you. In some periods the performance score will remain the same as before. In other periods the performance score will improve relative to the previous period.[24]

At the beginning of the task you will receive the release schedule with the performance scores of each laptop model. Each time a new laptop model is introduced you will be asked whether you want to purchase it (replacing your current model). Each time you purchase a new laptop you will be charged its price. For example, suppose you currently have a laptop with a performance score of 10. The performance score of the new model introduced this period is 13. The price of the new model is 6.

If you choose **not to purchase**, you will:

- spend no money.
- earn 10 points this period (your current laptop performance score).
- start the next period with your current 10-score laptop.

If instead you choose **to purchase**, you will:

- spend 6 points.
- earn 13 points this period (the new laptop performance score).
- start the next period with the new 13-score laptop.

The first laptop will be given to you for free. After that you will make 14 decisions. Each time you will be asked whether or not you want to purchase a new laptop, introduced in the current period. Your overall payoff will consist of the sum of the performance score values of the laptops used in each period, minus the

[24] The instructions and sample screens have been abbreviated with quiz and example portions of the instructions. The exact sequence of screens as seen by participants, as well as the quiz questions, can be downloaded under `https://bit.ly/2sPwk18`.

charges for each purchase. For example, suppose each model costs 6 points. Then, if you have purchased a 10-score laptop in period 1, and a 13-score laptop in period 6, your total payoff in this task will be: $10 \times 5 + 13 \times 9 - 6 \times 2 = 155$ points, or $7.75.

Laptop performance scores and price (in points):

Year 0 (free of charge): 1 point,

Year 1: 6 points

Year 2: 8 points

Year 3: 9 points

Year 4: 9 points

Year 5: 9 points

Year 6: 11 points

Year 7: 12 points

Year 8: 13 points

Year 9: 14 points

Year 10: 14 points

Year 11: 14 points

Year 12: 15 points

Year 13: 15 points

Year 14: 19 points

[Example and quiz screens]

Capacity Allocation Task

This part of the experiment will involve the Meal Plan Task. For this task you will have a meal plan. Your meal plan will specify how many free meals are included with it. If you use all your free meals, you will still be able to purchase meals at the regular price (which will be specified later). You will make 20 decisions, one for each meal. Each time you will be asked whether or not you accept or reject the meal offered to you.

Not all meals have the same value - each meal will have a distinct value describing how much it is worth to you. In particular, meals will have the following values:

15 points with probability 15%

12 points with probability 25%

9 points with probability 25%

6 points with probability 25%

3 points with probability 10%

To help visualize the value probabilities, imagine that there is a bucket full of 20 balls. Two balls have 3 written on then, Five balls have 6, Five balls have 9, Five balls have 12, and Three balls have 15. Determining the value of a meal would then be like picking one of the balls randomly, assigning that value to the meal, and then putting the ball back.

In each period you choose which meal "types" you want to accept and which ones you want to reject.

- If your decision for a given meal "type" is to reject, then your payoff will remain unchanged and you will move to the next period.
- If your decision for a given meal "type" is to accept, then the value of that meal will be added to your earnings.

The payment for each meal will depend on the number of free meals remaining in your plan. The payment for each meal will be:

- one free meal unit (if you have free meals left).
- the price of that meal (if you do not have any more free meal units left).

[Example and quiz screens]

Product Search Task

This part of the experiment will involve the Apartment Search Task. Your goal in this task will be to find an apartment to rent. During the task you can visit up to 15 apartments to assess their value (e.g. size and amenities relative to rent). As you conduct the search you gradually become more familiar with the city and its neighborhoods and get better at identifying good apartments. Thus, the longer you search, the higher the apartment values you can expect to see.

In particular, you expect the apartments to have the following values:

Apartment 1: 60 points

Apartment 2: 60 points

Apartment 3: 62 points

Apartment 4: 64 points

Apartment 5: 66 points

Apartment 6: 68 points

Apartment 7: 68 points

Apartment 8: 68 points

Apartment 9: 70 points

Apartment 10: 72 points

Apartment 11: 74 points

Apartment 12: 76 points

Apartment 13: 78 points

Apartment 14: 80 points

Apartment 15: 82 points

Note however, that these values are not known with certainty at the beginning of the task - they are estimates, based on your best guess.

While you expect the apartments to become better over time, you cannot be sure about any apartment value before visiting it. In particular, each apartment's value can be up to 5 points below or 5 points above your expectation with each number in that range being equally likely. For example, for the first apartment, which you expect to be worth 60 points, the values can range between 55 and 65 points. For the third apartment, which you expect to be worth 62 points, the values can range between 57 and 67 points, and so on for the remaining apartments.

To help visualize the value probabilities, imagine that there is a bucket with 11 balls. The balls are labeled with values, in the -5/+5 range around your expectation. For example, suppose you expect an apartment to be worth 60 points. Then, the 11 balls would have the following values written on them (one value for each ball): 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65 Finding out the value of an apartment would then be like picking one of the balls at random.

Remember that while on average you expect apartments to become better, their true values are uncertain. This means, you may see a lower apartment value compared to the previous period. In each period the apartment value will be drawn independently. This means, the value drawn in the current period does not give you any additional information about the values in the next periods. You can visit as many of the 15 apartments as you want. However each apartment visit will have a cost (as will be discussed on the next screens).

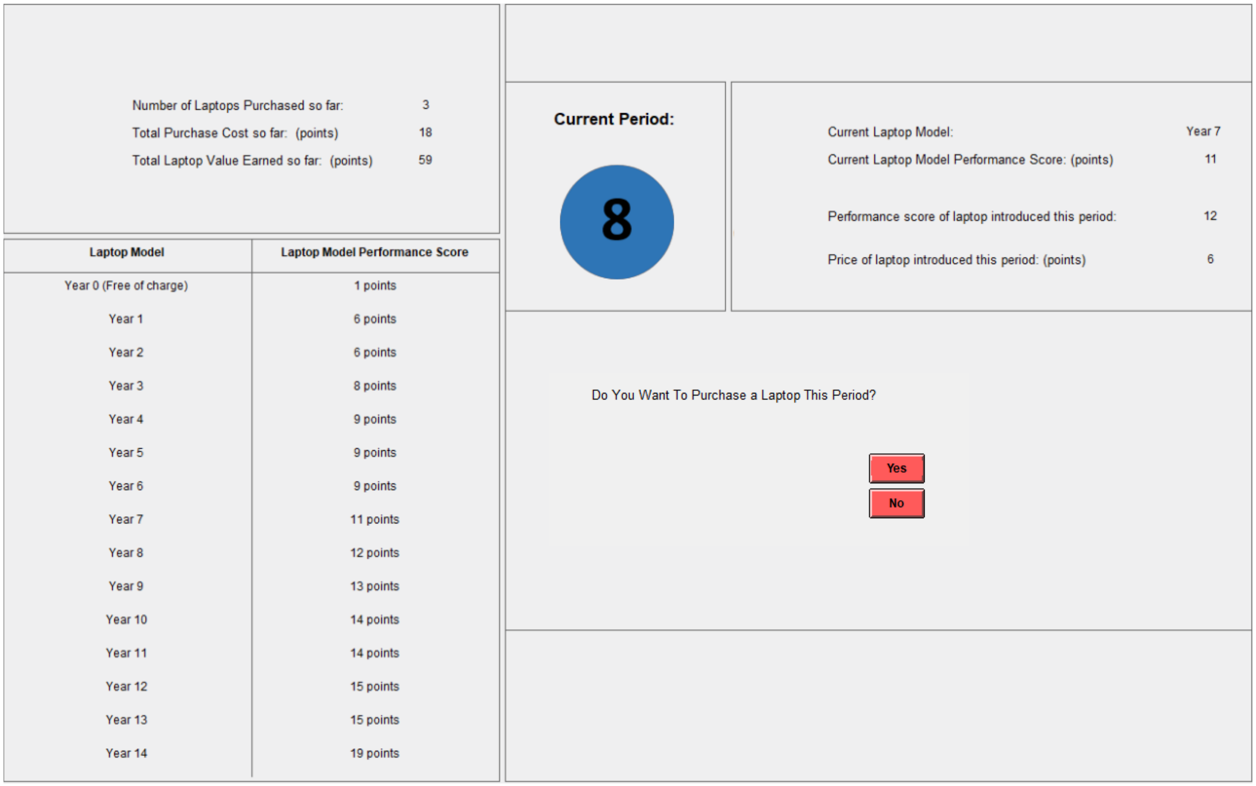
Apartments visited in the past are still available to you. In other words, when you decide to stop searching you will receive the highest apartment value that you've seen so far. For example, suppose you decide to stop after visiting three apartments, and their values were 62, 67 and 59. Suppose further that the cost of visiting one apartment is 1 point. In this case you will earn $67 - 3 \times 1 = 64$ points.

How will you choose when to stop searching? In each search period you will be asked the following question: What is the smallest apartment value that would cause you to stop search? The next apartment value will then be revealed, and your stated decision to continue (or not) will be implemented:

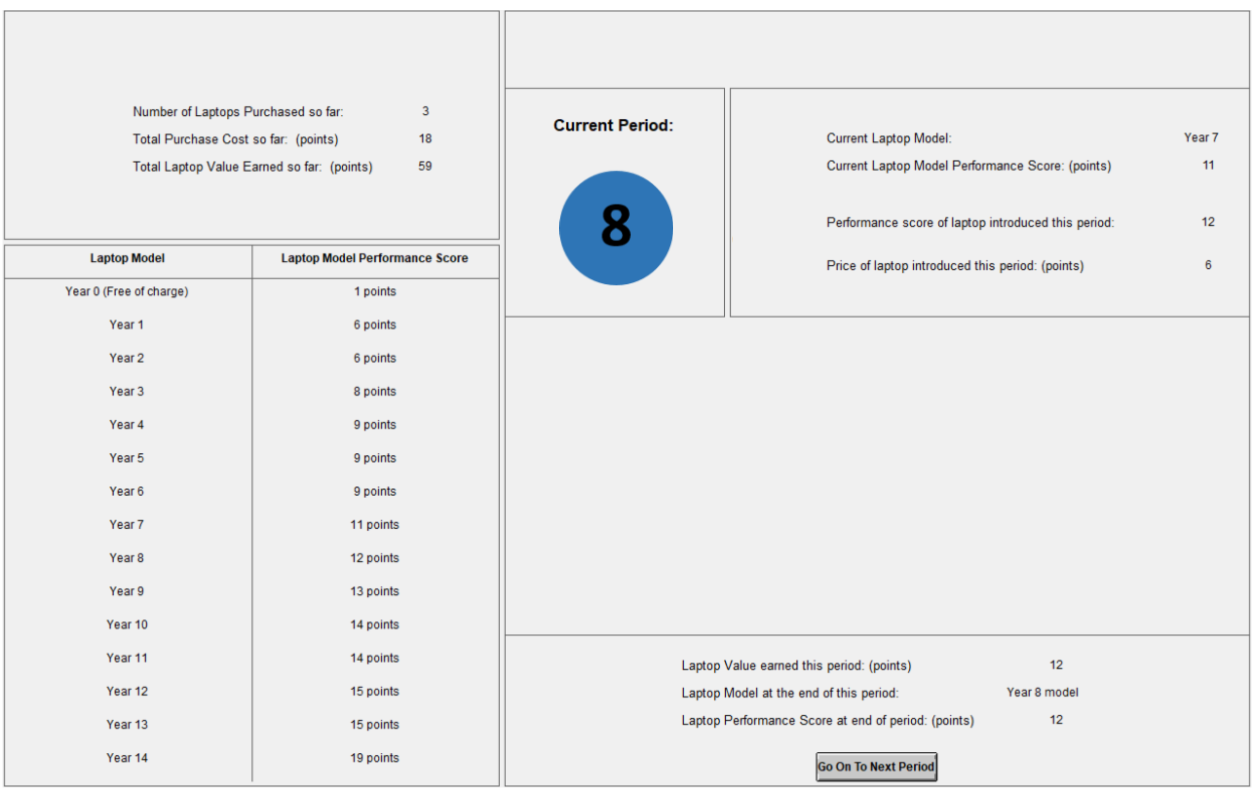
- if the revealed value is **below** your stopping value, then you will continue searching.
- if the revealed value is **above** your stopping value, then you will stop searching.

When you stop searching you will always rent the best apartment you have seen so far.

[Example and quiz screens]

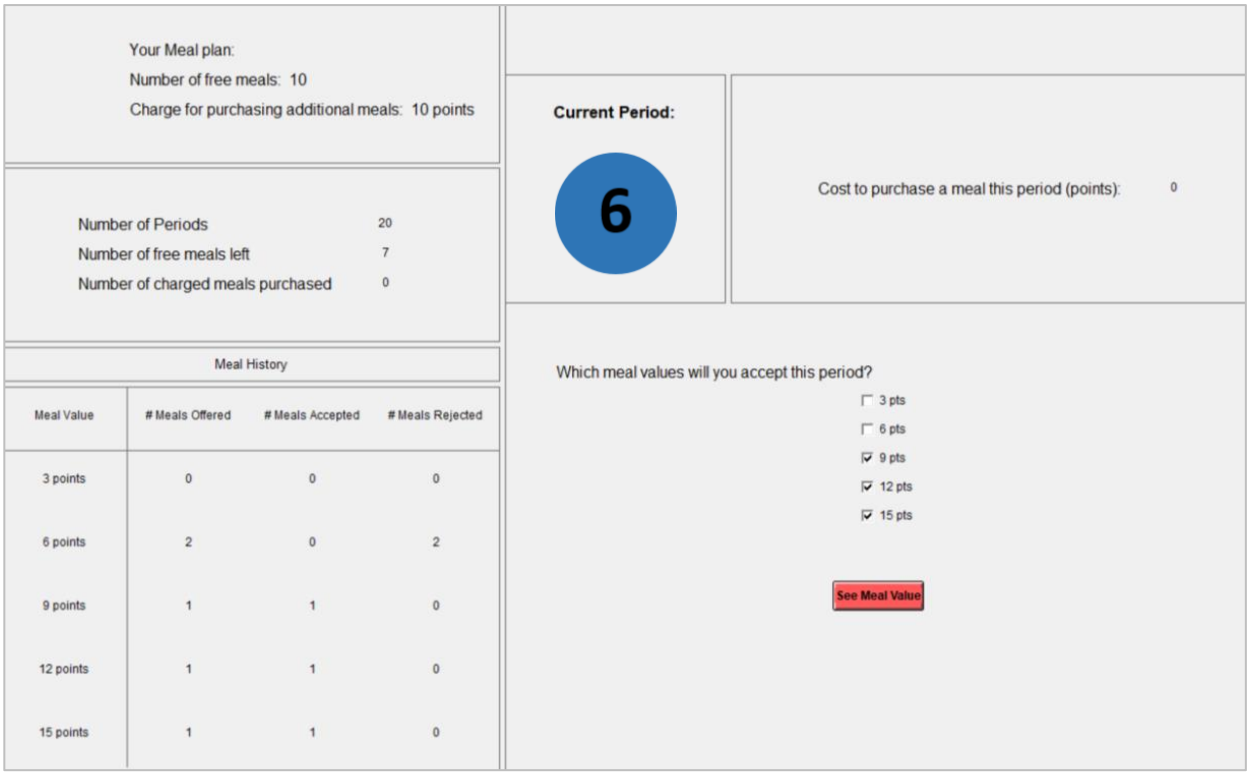


(a)

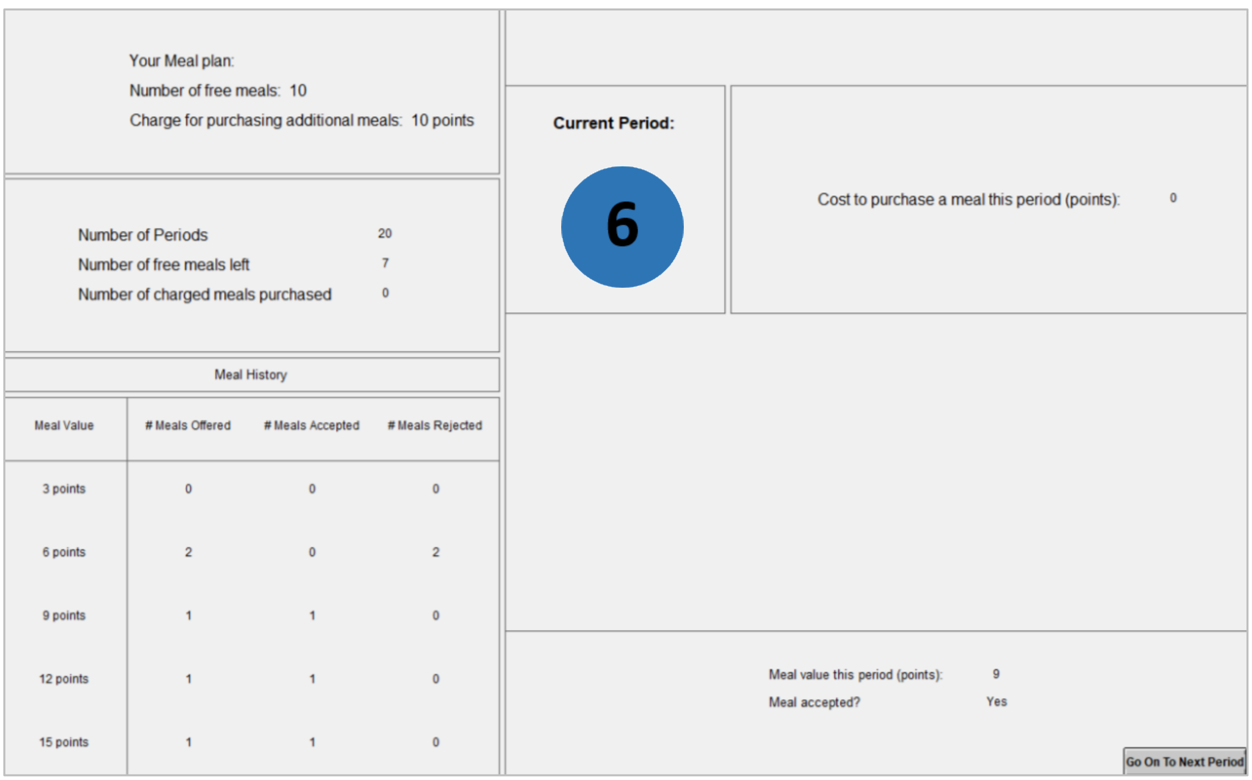


(b)

Figure B.1 **Technology Adoption: screen shots of the main decision screens. In (a) subjects are deciding to upgrade (or not) to a new model. In (b) the decision to purchase a new model has been made and subjects are about to move to the next decision period.**

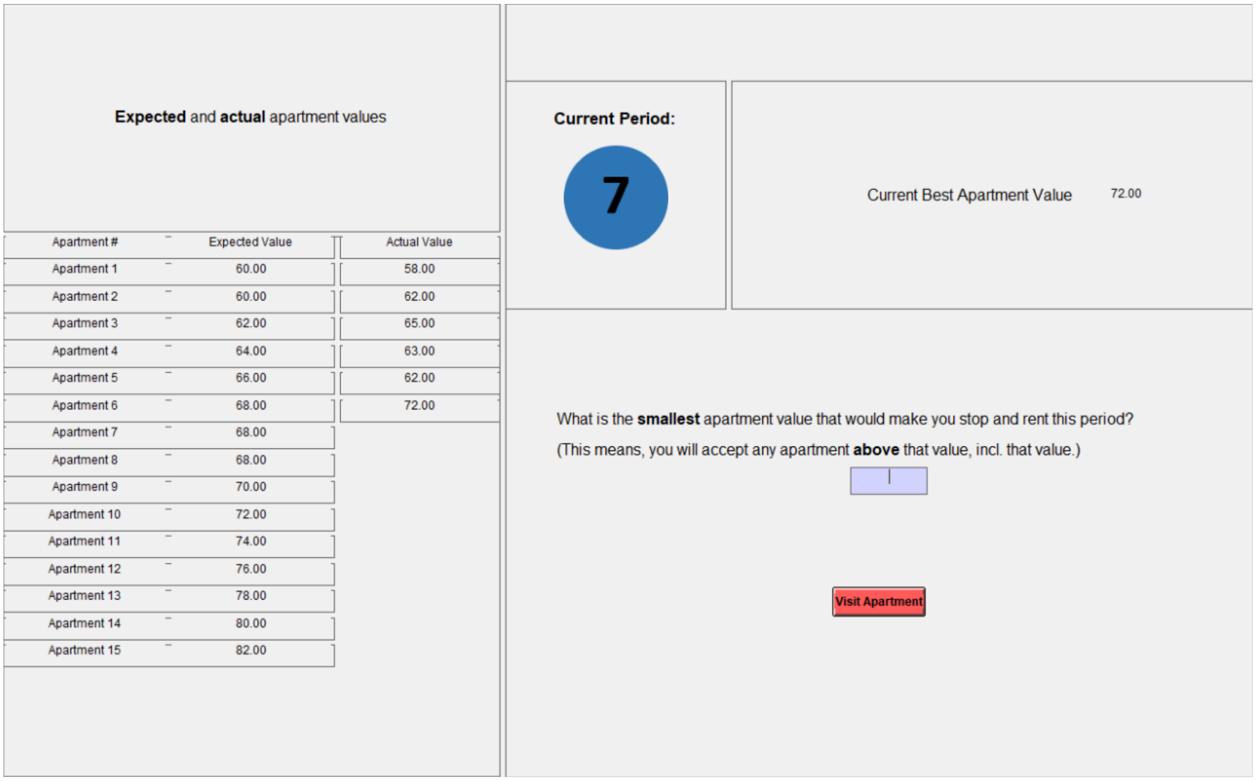


(a)

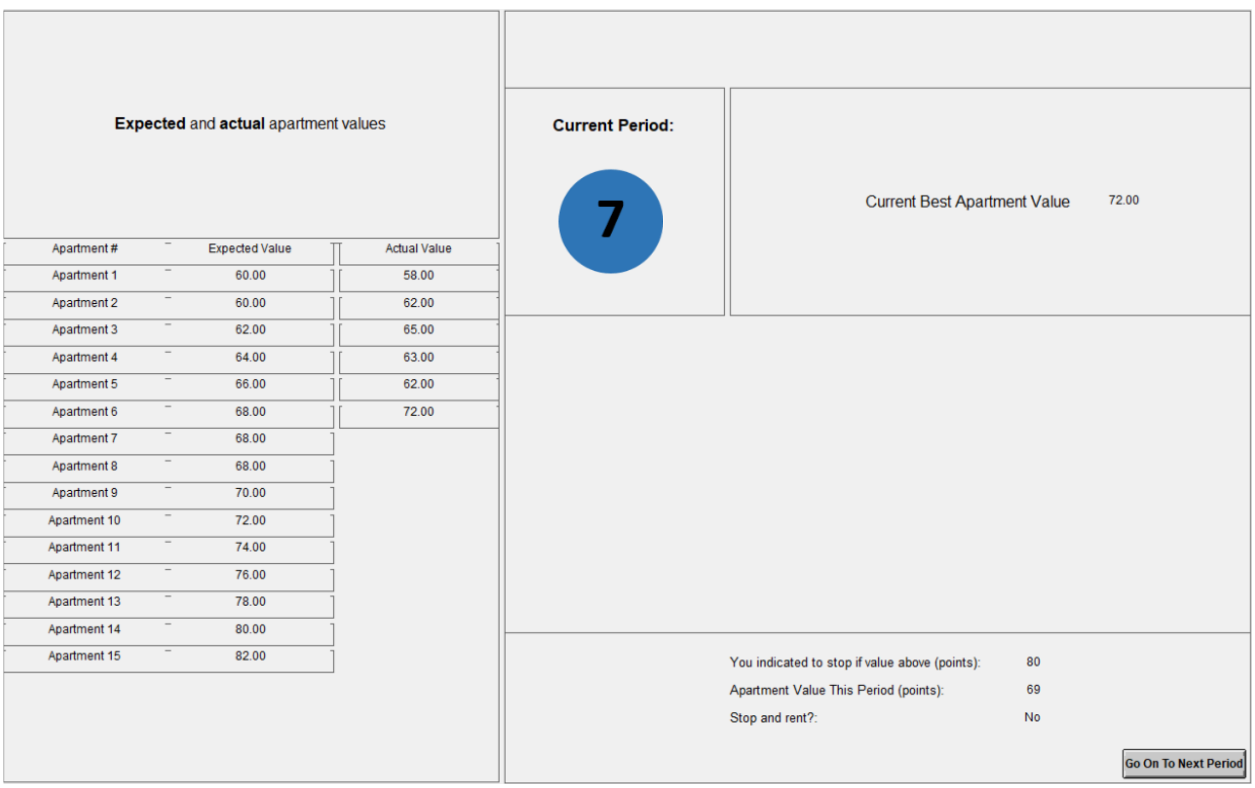


(b)

Figure B.2 Capacity Allocation: screen shots of the main decision screens. In (a) the subject is choosing the threshold for accepting/rejecting meals of different value. In (b) the realization of the random meal value and the acceptance/rejection of that value is displayed, and the subject is about to move to the next decision period.



(a)



(b)

Figure B.3 Product Search: screen shots of the main decision screens. In (a) the subject is choosing the threshold for the smallest apartment value that would lead her/him to stop searching. In (b) the realization of the random apartment value and the resulting stop/continue decision are displayed.

C. Maximum Likelihood Estimation of Policy Usage

In the main text we have used adjusted consistency scores to assign the closest decision rule to each subject (See Section 5.1 for the description of the method). In this section we use an alternative approach based on Maximum Likelihood estimation. Specifically we use an adaptation of the Strategy Frequency Estimation Method (SFEM) proposed by Dal Bó and Fréchette (2011). The construction of the method for the infinitely repeated Prisoner's Dilemma game is described in detail in Romero and Rosokha (2018). We modify this method, so that it can be used with non-binary action spaces, which is the case for some of our tasks.

SFEM is a finite-mixture estimation approach to estimate the proportion of policies used by the experiment participants. With a discrete action space that contains J decision alternatives, the distance $d_{itk}(s)$ between the action observed from subject i in period t and state s and the action that would be predicted by policy k in that state can range between 0 and $J-1$, i.e., $d_{itk} = 0, 1, ..., J-1$. If we denote by M_d the matrix that contains the number of times the distance of size $d = 0, 1, ..., J-1$ was observed for each subject $i = 1, 2, ..., N$, and for each policy $k = 1, 2, ..., K$, then we have $J-1$ matrices, with each matrix collecting the number of times each distance was observed in the subject population. Then, the proportion of policies can be estimated by maximizing the following likelihood expression:

$$\mathcal{L}(\phi, \beta) = ln\big(\phi' \cdot P\big) \cdot \mathbf{1},$$

where the first argument of the likelihood function, ϕ is the vector of proportions that maximize the likelihood of the choices observed in the data. The second argument of the likelihood function is the vector of population-average probabilities, β that maximize the joint probability of the observed choices:

$$\begin{aligned} P &= \beta_0^{M_0} \circ \beta_1^{M_1} \circ \cdots \circ \beta_{J-1}^{M_{J-1}}, \\ \sum_{d=0}^{J-1} \beta_d &= 1, \\ \beta_d &\geq \beta_{d+1}, \qquad d = 0, 1, ..., J-1 \end{aligned}$$

where β_d denotes the estimate of the population-average probability that a distance of size d is observed, and $\circ$ denotes element-wise product. The probabilities β_d must sum up to 1, and are constrained to be increasing with the inverse of the distance.

Table C.1 summarizes the bootstrapped estimates for the subset of non-static subjects for the Technology Adoption and the Capacity Allocation tasks in Study 1. Note that we do not estimate strategies for Product Search because the majority of subjects are static in Product Search.[25] Overall, the results match the proportions presented in Figure 8. In both tasks optimal strategy is the second-best match, accounting for about a quarter to a third of non-static subjects. Look-ahead policies are the best fit for most subjects in Technology Adoption, while the reoptimization heuristic is the best fit for most subjects in Capacity Allocation. Lastly, note that the sum of the β_d parameters is between 0.8 and 0.85 in each column indicating that collectively, the policies are a good descriptor for our subject population.

[25] We cannot use the SFEM method for static subjects because many static subjects essentially use a unique version/parametrization of the static policy, i.e., a different static threshold.

Table C.1 Maximum Likelihood Estimation Results

	Technology Adoption	Capacity Allocation
Proportion of subjects matched to policy		
$\phi_{Optimal}$	0.267	0.326
	(0.042)	(0.068)
ϕ_{LKH1}	0.205	0.125
	(0.037)	(0.082)
ϕ_{LKH2}	0.006	0.044
	(0.007)	(0.030)
ϕ_{LKH3}	0.359	0.011
	(0.045)	(0.015)
$\phi_{Break-even}$	0.086	0.118
	(0.026)	(0.054)
$\phi_{Reoptimization}$	0.076	0.375
	(0.021)	(0.104)
Population likelihood parameters		
β_0	0.809	0.454
	(0.008)	(0.031)
β_1	-	0.391
		(0.052)
LL	-3843.126	-3816.063

Note. Parameters (estimated shares of subjects following each policy) and standard errors are reported for 100 bootstrapped samples. The sample includes all non-static subjects. β_0 is constrained to be above 1/2 for Technology Adoption, and above 1/3 for Capacity Allocation.

D. Study 2: Experimental Instructions and Screen Shots

D.1. Experimental instructions: Machine Replacement Problem

This part of the experiment will involve the **Car Sale Task**. Your goal in this task will be to receive the maximum possible value out of the car you own. Your car is getting old, and you are starting to get worried that it might break down. You would like to sell it before that happens (your car will be worthless after a breakdown), but you would also like to get as much use out of it before selling it. The longer you continue to drive it, the more likely a breakdown becomes. The game will have several periods. In each period you will have the opportunity to sell the car.

- If you keep the car, and it does not break down you will get points, and can continue using the car.
- If you keep the car, and it breaks down you will get no additional points, and the game ends.
- If you sell the car you also get points (as will be described later), and the game ends.

As you car gets older, it becomes more and more likely to break down. However, while you expect that the likelihood of a breakdown increases over time, you cannot be sure when (and whether) it will happen. For example, Suppose the likelihood of a breakdown is 5%, and you decide to keep (not sell) the car. This means you will be able to keep using the car for at least one more period with probability 95%. To help visualize the probabilities, imagine that there is a bucket with 20 balls. The balls are labeled either "Break" or "No break". If the probability of failure is 5%, then 19 balls would have "No Break" written on them, and one would be labeled "Break". Finding out whether the car has failed, or not, would then be like picking one of the balls at random (and then returning it to the bucket). Remember that while on average you expect the probability of a break-down to increase, the timing of a break-down is uncertain. This means, you may see a break-down relatively early in the game, even though the probabilities of a break-down are low in the beginning. In each period the break-down can happen independently from the previous periods. This means, just because your car has not broken down yet, does not mean it will not break down in the next period(s). If you decide to sell the car, you will no longer be at risk of losing it to a break-down. The points you receive from selling the car will be added to the usage value you have accumulated so far. Whenever you decide to sell the car, you will still be able to use it one more time prior to selling. For example, suppose you decide to sell the car during the 5th period (that is, after using it for 4 periods). Suppose further that you receive 7 points in each period that you use the car, and the sale price is 40. Then, if the car does not break down in any of the first 4 periods, you will receive $7 \times 5 + 40 = 75$ points.

D.2. Parameters

The parameters were chosen as follows:

- Stopping problem with incremental payoffs

  Round 1-2: $T = 20, u = 12, S = 32, P = [0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.07, 0.07, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.16]$

  Round 3-4: $T = 15, u = 7, S = 40, P = [0.01, 0.01, 0.02, 0.02, 0.02, 0.08, 0.08, 0.08, 0.08, 0.09, 0.09, 0.1, 0.12, 0.12, 0.13]$

- Capacity Allocation task with continuous action space

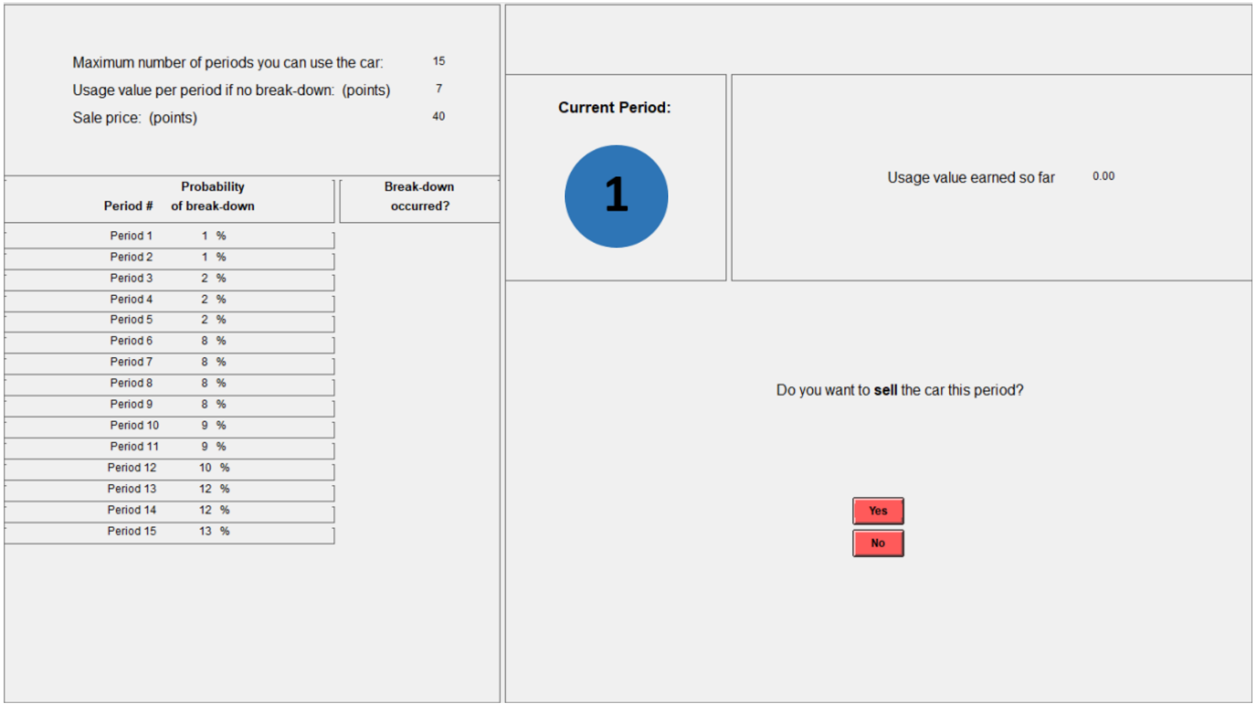
  $T = 20, V$ discrete uniform$\{1,10\}, c = 10, p = 5$

- Product Search task with stronger incentives

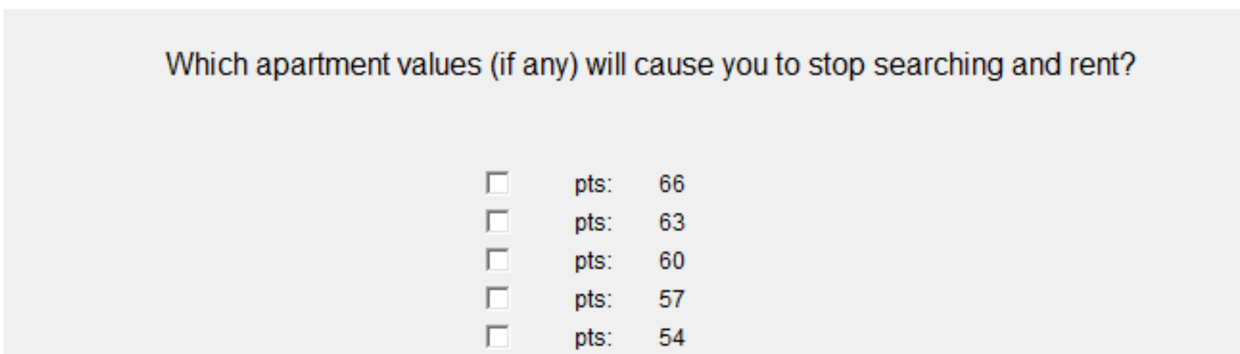
  $T = 15, V = [6, 8, 10, 13, 16, 19, 21, 21, 21, 24, 24, 26, 26, 26, 28], c = 1,$ value distribution is discrete uniform around the means, with zero mean and values of $\{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$.

- Product Search task with discrete action space

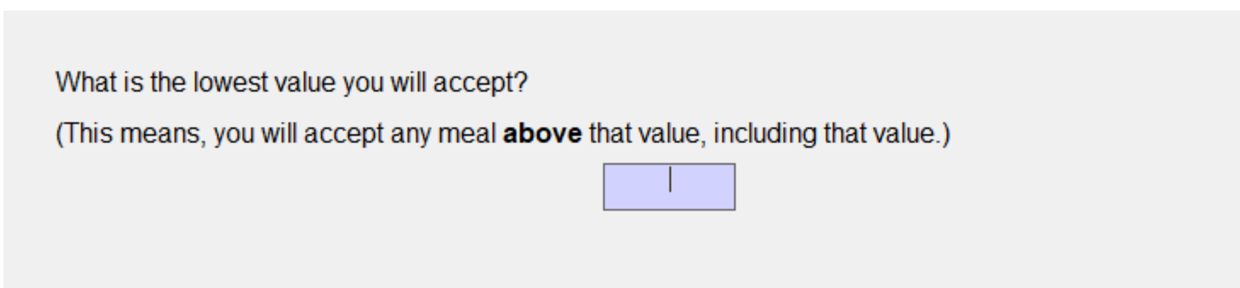
  $T = 15, V = [60, 62, 64, 66, 68, 70, 72, 72, 72, 74, 74, 76, 76, 76, 78], c = 1,$ value distribution is discrete uniform around the means, with zero mean and values of $\{-6, -3, 0, 3, 6\}$.



(a)



(b)



(c)

Figure D.1 **Screen shots of the main decision screens in Study 2. Panel (a) shows the decision screen in the Machine Replacement task. Panel (b) shows the action space in in the Product Search task with discrete action space. Panel (c) shows the action space in the Capacity Allocation task with continuous action space.**