Management Science

Contracts, Biases, and Consumption of Access Services

Stephen Leider, Özge Şahin

MANAGEMENT SCIENCE
Vol. 60, No. 9, September 2014, pp. 2198–2222
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1889

Contracts, Biases, and Consumption of Access Services

Stephen Leider
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109, leider@umich.edu

Özge Şahin
Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202, ozge.sahin@jhu.edu

We study theoretically and empirically the consumption of access services. We demonstrate that consumption is affected by contract structure (pay-per-use versus three-part tariffs) even if the optimal consumption plans are identical. We find that, although there is extensive individual heterogeneity, on average, consumers' choices follow a structure that is similar to a nearly optimal heuristic and correctly react to imbalances between the number of free calls and call opportunities remaining. However, consumers use the free units too quickly, leading to overconsumption and lost surplus. These errors are partially driven by mistaken beliefs about the value distribution. We also measure subjects' willingness to pay for a contract with free access units, and we find that nearly half of subjects are willing to pay at least the full per-unit price, with a substantial fraction willing to overpay. In response, the optimal firm strategy offers a three-part tariff at a very small discount, which increases revenue by 8%–14% compared to only offering a pay-per-use contract.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2013.1889.

Keywords: economics; behavior and behavioral decision making; marketing; pricing; microeconomics; intertemporal choice; decision analysis; applications; industrial organization; market structure; firm strategy; market performance

History: Received December 22, 2011; accepted August 5, 2013, by Teck-Hua Ho, behavioral economics. Published online in *Articles in Advance* July 17, 2014.

1. Introduction

Nonlinear pricing of access (subscription) services such as telecommunication, car leasing, club memberships, and product warranties has received a great deal of attention from both researchers and practitioners. Research up to date studies how firms should structure the tariffs (pay-per-use, two-part tariff, three-part tariff, and unlimited usage) and examines the drivers of consumer's tariff choice. Wilson (1993) reviews the literature on profit- and welfare-maximizing tariff structures. The fundamental assumptions of much of this literature are that consumers are rational decision makers who choose the surplus-maximizing tariff and the pricing structure does not influence consumer's value for the service. However, recent studies (Thomas and Morwitz 2005, Soman 2001) show that pricing structures themselves may affect the usage decisions. Indeed, Bertini and Wathieu (2008) state that pricing can transform, as well as capture, the utility of an offer.

The interaction effect between the tariff structure and usage is documented by recent research on telecommunication services (Ascarza et al. 2012) and health clubs (DellaVigna and Malmendier 2006). These studies show that consumers do not necessarily choose the tariff that leads to the lowest billing rate for a given amount of consumption. Ascarza et al. (2012) find significant differences in total consumption by consumers who choose a two-part and a three-part tariff contract that cannot be explained by a change in the budget constraint. In a different context, Iyengar et al. (2011) find that consumers have lower utility for two-part tariffs compared to pay-per-use tariffs, which results in lower usage of service for their particular application. These results suggest that consumers make mistakes while maximizing their surplus from the consumption of access services, and pricing structures influence the types of mistakes they make. Identifying the mistakes and the behavioral biases that determine this consumption behavior has important ramifications for pricing and optimal contract design.

When choosing an access contract, consumers are generally uncertain about how much they will use the service and/or how valuable usage opportunities are. This uncertainty makes the consumption problem quite complicated. In this paper, we investigate how different contractual forms (pay-per-use or three-part tariff) affect in-contract consumption decisions. In contrast to the earlier literature, we model the consumption process in detail and analyze the consumer behavior *after* the purchase of the contract. In particular, we are interested in answering the following questions: (1) Is there a plausible heuristic that consumers use to make daily consumption decisions? (2) Given the heuristic used, does contractual form affect consumption behavior? (3) Do decision biases lead to substantially different consumption behavior under certain contractual forms? (4) What are the implications of the interaction effect between contractual form and consumption for the firms?

To that end, we develop a theoretical model of consumption behavior for an individual who faces a sequence of consumption opportunities and has been endowed with a price contract. This model identifies the optimal consumption policy, as well as a nearly optimal heuristic. In addition to these two dynamic policies, we consider a myopic and a static heuristic that were studied in the existing literature. Furthermore, we identify how consumption would shift under biases such as the overestimation or underestimation of consumption values, risk aversion, regret aversion, the sunk cost fallacy, and the taxi-meter affect (a physiological transaction cost such as distaste for payment at the time of consumption).

We test our theoretical model by conducting a laboratory experiment where subjects repeatedly perform a dynamic consumption "cell phone task." In each task, subjects are endowed with a contract that provides 0, 10, or 20 free phone calls (and charged an access fee) and then receive 30 calls whose value is drawn randomly. For each call, subjects decided whether to answer the call (and either use a free call or be charged the per-unit cost). To study how the structure of subjects' decisions compare to the four policies we consider, we first ask which policy structures are more consistent with the aggregate patterns in individual choices. We also explore individual heterogeneity and match each individual's profile of choices to the four policies. Additionally, we examine within the structure of subjects' choices whether they are correctly "calibrated" in the sense of using their allotment of free calls at the optimal rate.

When we study the aggregate patterns in consumer choice, both the optimal dynamic policy and the proposed nearly optimal dynamic heuristic describe the average structure of subjects' choices well, with subjects adjusting their threshold in the correct direction as the optimal threshold (and nearly optimal heuristic) changes as the remaining time and included free units in the contract deplete. We find that the nearly optimal heuristic matches their average choices somewhat more closely than the optimal policy. However, in line with previous empirical research, they are too aggressive in using their free calls, which leads to suboptimally answering too many calls and answering too many calls of low value. These mistakes cost subjects up to 20% of their payoff, and subjects continue to make these mistakes even after repetition of the consumption task. When we explore the individual profiles, we find that a substantial fraction of subjects consistently use the static and myopic policies and an equally high fraction are consistently using dynamic policies. When the policies are distinguishable, our proposed dynamic heuristic provides the best unique description of the subject decisions among myopic, static, optimal dynamic, and heuristic dynamic policies. Overall, we find that many subjects are reasonably sophisticated, in that they use dynamic choice rules that follow the patterns of the optimal dynamic policy; however, they consistently exhibit an overuse bias. Therefore, on average, consumers can best be thought of as moderately sophisticated—neither being fully myopic nor fully optimal.

Additionally, we measure subjects' beliefs about the value distribution, and we show that mistaken beliefs can explain part of the overusage effect. Specifically, subjects with a contract providing 20 free calls who overestimate the frequency of low-value calls (and therefore underestimate the value of future calls) are more liberal in using their free calls. However, mistaken beliefs do not lead to overusage in the 10 Calls treatment. In a second experiment, we provide subjects with the full call value distribution and again find that subjects significantly overuse their free units.

We then run a third experiment where we can directly measure subjects' willingness to pay for the contract with free calls (instead of the pay-per-use contract). When we offered the 10 Calls contract, we find that almost 50% of subjects are willing to pay full price or more (i.e., pay in advance at least as much as it would cost to answer the calls under pay-per-use contract), with 21% of subjects willing to pay more than the pay-per-use price. Strikingly, this latter group increases over time to 27% in the fourth repetition. When we offered the 20 Calls contract, more than 40% of subjects are willing to pay at least the pay-per-use price, with 8% willing to pay more. Prepurchasing the 20 Calls contract sacrifices substantial option value relative to the pay-per-use contract, because only 51% of subjects receive 20 or more calls worth at least $0.35.

We also find that in the 10 Calls treatment, the subjects who report the highest willingness to pay for free calls also answer the fewest calls under a pay-per-use contract. Therefore, offering a contract with a three-part tariff has three benefits to the firm: extracting revenue via the monthly fee from consumers who value free calls highly, sorting consumers out of pay-per-use who will be low usage customers, and increasing the usage of consumers with free calls. We do not find a similar sorting effect in the 20 Calls treatment, indicating that the sorting behavior may be affected by the menu of contracts available.

We then calculate the optimal access fee (i.e., the revenue-maximizing price for the firm given average consumer behavior under each kind of contract) and find that the optimal discount from the full pay-per-unit cost is very small. With the optimal fee, the firm increases revenue by 14% by offering the 10 Calls contract, and increases revenue by 8% by offering the 20 Calls contract, compared to only offering a pay-per-use contract. The 10 Calls contract increases firm revenue more than the 20 Calls contract because of both greater overvaluation of the contract and greater revenue generated from consumer overusage of free calls.

2. Literature Review

We first survey the existing literature on price discrimination in access service industries. Most of the literature in this area assumes there are multiple types of consumers that differ in their taste for consumption, and that a monopolist firm offers a menu of pricing contracts to induce consumers to self-select into the appropriate contract given their type. Examples of nonlinear pricing contracts used in the telecommunications and utilities markets include pay-per-use contracts; two-part tariffs with an access fee and a per-unit usage price; three-part tariffs with an access fee, some number of free units, and a pay-per unit usage price; and unlimited usage contracts (for a review of the economics literature on nonlinear pricing, see Wilson 1993 and Tirole 1988).

Typically consumers must select the pricing contract significantly in advance of the consumption decisions, which introduces uncertainty about future demand and consumption valuations. Instead, the consumer must rely on an estimate over her usage during the contract duration. Many papers have analyzed the effects of demand uncertainty and measured its effects on pricing. Miravete (2002) estimates a structural econometric model of demand for fixed-line telephone service for a provider that offers a two-part tariff and a flat-rate tariff, allowing for uncertain future consumption. Lambrecht et al. (2007) find that it is ex ante optimal to choose a tariff with a higher usage allowance than would be optimal if they were not uncertain over their demand.

The advance pricing and revenue management literature also suggests nonlinear pricing solutions for one-time use services such as event tickets and air transportation (Xie and Shugan 2001, Gallego and Şahin 2010). The fundamental assumption of all these papers is that consumers are rational decision makers who seek to maximize their surplus. Moreover, this literature focuses on the contract purchase decisions rather than ex post consumption behavior. Recently an operations management literature has developed studying access service pricing using a queuing framework where the service system may be congested. Most of this literature studies pay-per-use pricing, with a few exceptions that study subscriptions (access fee with unlimited usage). Randhawa and Kumar (2008) compare per-use pricing with subscription pricing that imposes usage limits (similar to Netflix's policy). Cachon and Feldman (2011) compare pay-per-use and subscription pricing when there are congestion costs. Bitran et al. (2008) study two-part tariffs where the firm's pricing policy and service level (quality) affects the dynamics of their system over time through customer satisfaction. In our study, consumers do not experience congestion costs, and we compare ex post consumption behavior under a pay-per-use contract and a three-part tariff. These types of contracts are common in car leasing, telecommunication services, and utilities where system congestion is rarely an issue for the consumer. Behavioral research on optimal stopping problems, such as when to stop a job search or when to adopt a new technology, is also related to this work. The majority of this research is in the context of labor economics (Cox and Oaxaca 1989, Schotter and Braunstein 1981) and the decision maker has a single position to fill (single unit to consume). Cox and Oaxaca (1989) show that decision makers employ complex but suboptimal policies that are structurally similar to the optimal. Bearden et al. (2007) study the closest optimal control problem (a standard revenue management problem where the seller maximizes his revenue from multiple units over a finite horizon by accepting and rejecting the offers for multiple units that arrive sequentially over time) to ours. Different than Bearden et al. (2007), we directly elicit the strategy of the consumer before each decision, which allows us to identify the consumer's policy more clearly. Moreover, we theoretically show how behavioral biases would change the optimal and heuristics policies as well as perform several diagnostic tasks to identify the behavioral biases in decision making.

Another body of work focuses on decision biases and mistakes in tariff choice. This literature has found that the consumers often make mistakes in tariff choice (Kridel et al. 1993, Miravete 2002, Train et al. 1987, DellaVigna and Malmendier 2006, Nunes 2000, Grubb 2009). In particular, consumers exhibit a biased preference for choosing a flat rate contract (unlimited usage plans) over a pay-per-use option even if it leads to a lower consumption value. Lambrecht and Skiera (2006, p. 213) identify risk aversion, demand overestimation, and a distaste for paying per consumption ("taxi-meter" effect) as possible causes of the flat rate bias. DellaVigna and Malmendier (2006) show that health club users overestimate their future usage by more than 100% and subsequently tend to choose flat rates over pay-per-use contracts. Several other papers consider the effect of self-control problems on optimal nonlinear pricing (DellaVigna and Malmendier 2004, Oster and Scott Morton 2005, Esteban et al. 2007, Plambeck and Wang 2013). Note that all of these papers study static contract choice, whereas we study the dynamic consumption decisions and mistakes of consumers after the contract choice.

Recent empirical work has shown that within-month consumption is strongly affected by the contract terms beyond what can be explained by the change in marginal prices and budget constraints. In particular, Ascarza et al. (2012) estimate that within individuals, demand satiation increases by 31.5% under a three-part tariff (after controlling for budget effects). Ascarza et al. (2012) point out that a pricing plan may have attributes that alter and influence the consumer's usage decisions. To that end, we model consumers who are uncertain about their consumptions and make usage decisions taking into account the remaining balance of included free units and time in their contract. This is an improvement over the existing literature, which typically looks into post-tariff behavior. Grubb and Osborne (2014) and Yao et al. (2012) are two exceptions that analyze in-contract consumer decisions using cellular phone data. We discuss how our findings support and differ from those in §8.

In summary, our work is the first experimental paper we are aware of that focuses on the effect of decision heuristics and biases on the post-tariff choice consumption decisions. We compare three-part tariffs to pay-per-use and focus on the impact of overestimation or underestimation of consumption values, risk aversion, regret aversion, and the sunk cost and taxi-meter effects on the dynamic consumption decisions and heuristics. Finally, we examine firm's optimal contract.

3. Consumer Behavior and Theoretical Predictions

We will first present our theoretical model and predictions.[1] We study a three-part tariff (x, K, p), where x is the access fee, K is the number of free units (initial allowance), and p is the nonnegative per-unit fee for any consumption over initial allowance. The consumer pays fixed cost x for the right to use the service and K free units in T periods. If her consumption turns out to be more than K, she pays a per-unit fee p for each additional unit. Notice that pay-per-unit contract $(0, 0, p)$ is a special case. We are interested in in-contract consumption behavior and how the contract terms influence this behavior. We do not theoretically study the contract purchase decisions and the optimal menu of contracts.

Consumers are uncertain about their exact consumption levels and the value of each consumption opportunity, V. Whenever a consumption opportunity arises, consumers observe the actual value of the opportunity and decide whether to consume a unit of service. First we consider a risk-neutral rational consumer. Consumption opportunities arise sequentially over time. The risk-neutral rational consumer extracts utility v from the consumption of each free unit. If she uses the service when she does not have any free units, she pays pay-per-unit price p resulting in a net benefit of $v-p$. We first study the optimal dynamic consumption policy.

3.1. Stochastic Dynamic Consumption Model and Optimal Policy

Most individuals use shortcuts and heuristics when making daily consumption decisions. To understand the consumption problem and identify plausible heuristics, we first solve the discrete time optimal control problem of an unboundedly rational consumer who holds an (x, K, p) contract with a duration of T periods. Then we derive heuristics that have similar structural properties with this dynamic optimal policy.

We assume the value of each consumption V follows a distribution such that a consumption opportunity with value v or less arises with probability $F(v) = P(V \leq v)$ at time t, and the consumer adjusts her consumption strategy dynamically over time. Each consumption uses one unit of service. With k units left and t periods to go, the optimal expected utility is given by

$$J(k,t) = E[\max\{V + J(k-1, t-1), J(k, t-1)\}], \quad k > 0,\ t \geq 1,$$
$$J(k,t) = tE(V-p)^+, \quad k \leq 0,\ t \geq 0.$$

If there are t periods to go until the contract coverage ends, the consumer observes the value of the service and decides whether to use the service or not. If a free unit is used, then her expected utility is $V + J(k-1, t-1)$, where $J(k-1, t-1)$ is the optimal expected utility with $k-1$ free units and $t-1$ periods to go. Otherwise, her expected utility is given by $J(k, t-1)$. If she has no free units left and there are t periods to go until the contract expires, she uses the service only if the value of the consumption opportunity is greater than the pay-per-unit fee, $V \geq p$,

[1] Additional details and proofs are included in the online appendix, available at http://www-personal.umich.edu/~leider/Papers/AccessServicesOnlineAppendix.pdf.

resulting in the expected utility $tE(V-p)^+$. With some algebra, we can write the optimal expected utility as $J(k,t) = J(k,t-1) + E[\max\{V - \Delta J(k,t-1), 0\}]$, where $\Delta J(k,t) = J(k,t) - J(k-1,t)$ with boundary condition $J(k,0) = 0$.

Theorem 1 shows that the decision maker uses a threshold policy to ration the consumption opportunities and characterizes the optimal *stochastic dynamic threshold* (SDT). She uses the service if the value of the service is greater than the threshold $\Delta J(k,t)$. The value $\Delta J(k,t)$ is the opportunity cost of using the kth remaining free unit when there are t periods to go. The threshold, $\Delta J(k,t)$, is a function of the valuation distribution V, pay-per-unit fee, the number of remaining free units, and remaining time until contract expires.

Theorem 1 (Stochastic Dynamic Threshold Policy).[2] *It is optimal to use the service if and only if* $V \geq \Delta J(k,t)$. *Moreover,* (i) $J(k,t)$ *is increasing in* k *and* t, (ii) $\Delta J(k,t)$ *is decreasing in* k *and is increasing in* t, (iii) $\Delta J(k,t)$ *is increasing in* p.

The first part of Theorem 1 states that the consumer utility is higher if she has more free units and more time to use the free units. The last two parts show that the threshold is decreasing in the number of free units, increasing in the remaining time t and per unit price p. This implies that the consumer becomes more conservative in her usage as the remaining time in her contract and the pay-per-unit fee increase, whereas she becomes more liberal in her usage as the number of free units in the contract increases.

Although a rational individual who has infinite computational ability would use the optimal stochastic dynamic threshold policy stated in Theorem 1, it is more likely that she uses a heuristic in making consumption decisions as computing and updating the optimal threshold over time requires solving a nontrivial stochastic optimization problem. Next we will use our understanding of the structure of the optimal policy to develop near optimal dynamic heuristics. We discuss three heuristics, some of which are proposed by the earlier literature in the following section.

3.2. Heuristic Policies

In this section, we study three heuristic policies in increasing complexity: (i) a myopic policy, (ii) a deterministic static threshold (DST) policy, and (iii) a deterministic dynamic threshold (DDT) policy. The myopic and a version of the static policy are previously studied by Liebman and Zeckhauser (2004) and Borenstein (2009), respectively. Here we show the connection of the deterministic static threshold policy to the optimal policy, and then we propose an alternative to this totally static (inattentive) heuristic, the deterministic dynamic threshold policy. This alternative decision rule is structurally similar to the optimal policy but easier to compute and results in higher consumer surplus than myopic and deterministic static policies as it updates the policy over time (hence, it is a deterministic but attentive policy). In §5, we investigate which of the three heuristic policies and the stochastic optimal policy match consumption decisions better in a cell phone usage experiment.

3.2.1. Myopic Policy. With the myopic policy, also called spotlighting (Liebman and Zeckhauser 2004), consumers focus on the instantaneous costs and payoffs in the current period without considering the effects of the current period decision on the remaining decisions (i.e., they ignore the opportunity cost of the current decision). Thus, consumers with an (x, K, p) contract will use a free unit for any consumption opportunity that has a positive value (i.e., use threshold zero to filter the consumption opportunities) if they have free units and will use a unit if the value of the consumption opportunity is greater than the per-unit cost p if there are no free units left (i.e., use p as the threshold when they run out of free units). This policy ignores the opportunity cost of using a free unit now and the possibility of seeing higher-valued consumption opportunities in the future.

3.2.2. Deterministic Static Threshold Policy. The stochastic static threshold policy, first proposed by Borenstein (2009) and Grubb and Osborne (2014), assumes that the consumer picks a threshold at the beginning of the consumption horizon and uses this threshold to filter the consumption opportunities over time.[3] We derive a deterministic static heuristic policy that is asymptotically optimal for the stochastic optimal control problem stated in §3.1 as the free units in the contract and the number of consumption opportunities grow. The threshold is static because it is calculated only once at the beginning of the contract duration and deterministic in the sense that it ignores the uncertainty in the number of future consumption opportunities and assumes the number of opportunities is known (in expectation).[4] During the course of the contract, the consumer is inattentive and does not

[2] Papastavrou et al. (1996) show a similar result for the problem with $p = \infty$. We omit the proof of this result. The proof is similar to Papastavrou et al. (1996) and available from the authors for continuous and discrete valuation distributions as well as for a continuous time model with Poisson arrivals.

[3] In the context of electricity markets, Borenstein (2009, p 4) explains this policy as the "behavioral rule" consumers use to make decisions about their consumption patterns before the consumption period begins. During the consumption period, exogenous shocks to the quantity demanded occur, but consumers do not change their behavior in response.

[4] Note that in our experiment, the number of call opportunities is fixed and known to our subjects.

track usage but simply makes all calls valued above the threshold.[5]

A static heuristic is appealing because if an individual uses the same static threshold (behavioral rule) in all decisions until the expiration of the contract, this simplifies the utility maximization problem stated in the previous section. With this policy, the consumer uses the service if its value is greater than threshold q if she has free units. Then she filters the consumption opportunities by p if she has no free units. The expected utility of a free unit given the value of consumption is greater than q is $E(V \mid V > q)$. If the consumer has to pay for the service, the expected utility is $E(V-p)^+$. Combining these two terms, we obtain the expected utility of a consumer who uses the *same* static threshold rule:

$$J^s(k,t) = \max_{q\le p} J^s(k,t,q) = E(V \mid V > q) E\min(D_q, k) + \frac{E(V-p)^+}{\bar F(p)} E(D_q - k)^+ \frac{\bar F(p)}{\bar F(q)}.$$

We assume here the consumption opportunities arise in every period, and therefore the number of answered calls, D_q, is a binomial random variable with parameters t and $\bar F(q)$. This optimization problem is still a nontrivial stochastic optimization problem.[6] To simplify the heuristic even further, we assume that the consumer considers only the expectation of the number of units consumed if threshold q is used (i.e., ignores the fact that D_q is a binomial random variable with parameters t and $\bar F(q) = P(V \ge q)$, and assumes that the total number of consumption opportunities filtered with q is equal to the expectation of D_q, $t\bar F(q)$). Replacing stochastic demand D_q with its expectation, we obtain the following *deterministic* approximation[7] to the consumer utility problem, which we use to derive the DST:

$$J^d(k,t) = \max_{q\le p} J^d(k,t,q) = E(V \mid V > q)\min(t\bar F(q), k) + E(V-p)^+ \frac{(t\bar F(q) - k)^+}{\bar F(q)}.$$

The first term is the utility derived from the consumption of free units, and the second term accounts for the expected utility derived from the paid units. Notice that deterministic approximation $J^d(k,t)$ provides an upper bound on the utility given by the stochastic problem.[8] We call the policy derived from this problem the *deterministic static threshold policy*. We define $q(k,t) = \min\{q \ge 0 \mid P(V > q) \le k/t\}$ as the free-unit clearing threshold. Notice that q cannot be larger than the per-unit price p, so we have $q(k,t,p) = \min\{p, q(k,t)\}$.

Theorem 2. (i) *The policy $q(k,t,p)$ maximizes $J^d(k,t,q)$,* (ii) *$q(k,t,p)$ is decreasing in k and increasing in t,* (iii) *$J^d(k,t)$ is increasing and jointly concave in k and t for any fixed p, and* (iv) *$J^d(k,t)$ is decreasing in p for any fixed (k,t).*

Theorem 2 shows that a free-unit clearing threshold bounded by a pay-per-unit fee maximizes the deterministic approximation $J^d(k,t,q)$. Given a contract (x,K,p) to be used in T periods, we define the deterministic static threshold as $q^{\text{DST}}(k,t,p) = q(K,T,p)$ for any $k \le K$ and $t \le T$. This quantity is calculated at the beginning of the contract duration and is not updated. Notice that the DST policy mimics the structure of the stochastic dynamic optimal policy (Theorem 1) although it is significantly easier to calculate compared to both the stochastic optimal dynamic threshold policy and the stochastic static threshold policy (the optimal solution of $J^s(k,t,q)$). It requires a simple comparison of the ratio of the expected number of consumption opportunities to the number of free units (k/t) with the likelihood of receiving a call that has value greater than q^{DST}. This suggests that the individual should answer the top $100k/t$ percentage of the highest-valued calls. As the expiration time T decreases, expected number of consumption opportunities decreases, therefore the consumer chooses a lower threshold at the beginning of the horizon and uses the free units more liberally. If the contract has a fewer number of free units, the consumer chooses a higher threshold (use the free units more conservatively). One can also show that this deterministic static heuristic is asymptotically optimal (as T and K increases) for the stochastic optimal control problem discussed above.

3.2.3. Deterministic Dynamic Threshold Policy. This heuristic builds upon the analysis and discussion of the static threshold policy. Whereas the stochastic optimal dynamic threshold policy assumes completely attentive and sophisticated decision makers, the deterministic static threshold policy lies on the other end of the spectrum by assuming inattentive

[5] The customer is attentive to no longer having any free calls and will adjust her policy in this case.

[6] Solutions to structurally similar problems are given in Borenstein (2009) and Grubb and Osborne (2014).

[7] Savage (2012, p. 11) calls this type of deterministic approximations of stochastic variables the "flaw of averages" and discusses extensively when they are good approximations.

[8] We can see that the deterministic utility is an upper bound to the stochastic utility by realizing $\min(D_q,k) = D_q - (D_q-k)^+$, $E(V \mid V > q) - (E(V-p)^+)/(\bar F(q)) \ge 0$ for $q \le p$ and using the Jensen's inequality for convex functions: $J^s(k,t,q) = E(V \mid V > q)(ED_q - E(D_q-k)^+) + E(V-p)^+((E(D_q-k)^+)/(\bar F(q))) = E(V \mid V > q)E(D_q) - E(D_q-k)^+(E(V \mid V > q) - (E(V-p)^+)/(\bar F(q))) \le E(V \mid V > q)t\bar F(q) - (t\bar F(q)-k)^+(E(V \mid V > q) - (E(V-p)^+)/(\bar F(q))) = J^d(k,t,q)$.

decision makers (they do not change their behavior in response to the fact that opportunity cost of using each free unit changes over time). We expect decision makers to adjust their behavior over time as they use up their free units. The question is how they can adjust their behavior in a computationally inexpensive way. As we discuss above, DST requires a simple computation that can be easily repeated without increasing the computational burden on the decision maker or require higher analytical sophistication. Therefore, we consider the dynamic version of the deterministic heuristic discussed above. Consumers who use the deterministic dynamic threshold policy re-solve $J^d(k, t, p)$ at every consumption opportunity and recalculate the deterministic static threshold $q(k, t, p)$ that would apply for the remaining time. Given a contract (x, K, p), a consumer uses the threshold $q^{\mathrm{DDP}}(K, T, p) = q(K, T, p)$ at the beginning of the horizon, but updates this threshold to $q^{\mathrm{DDP}}(k, t, p) = q(k, t, p)$ as the free units used $(k < K)$ and the remaining time in the contract decreases $(t < T)$. This means that the consumer does adjust based on the length of time remaining and the number of free units left, but does not account for the effect of possible future adjustments on the current policy while adjusting the current threshold. This increases the expected utility and partially captures the dynamic nature of the optimal policy.

Next we study how behavioral biases alter the heuristic policies.

3.3. Overestimation or Underestimation of Call-Value Frequency and Overconfidence

It has been shown that mistaken beliefs on the likelihood of future events may have significant impact on individuals' decisions (Loewenstein et al. 2003; Eliaz and Spiegler 2006, 2008). In our setting, consumer beliefs about the value and the number of consumption opportunities may affect their consumption behavior. The following theorem shows how the optimal and the deterministic static thresholds of a consumer who overestimates or underestimates demand or the valuation distribution differs from a rational consumer's threshold.

Theorem 3. *Suppose that V is the true valuation distribution and W is the consumer's perception of her valuation distribution. If V first-order stochastically dominates W, $V \succeq W$, then $\Delta J_V(k, t) \geq \Delta J_W(k, t)$ and $q_V^{\mathrm{DDT}}(k, t, p) \geq q_W^{\mathrm{DDT}}(k, t, p)$.*

If the consumer estimates the upper tail of the valuation distribution to be lighter (heavier) than it really is, she is more aggressive (conservative) in using free units than she is if her estimates are correct with the optimal policy, deterministic static policy, and dynamic deterministic policy.[9] Overestimation or underestimation of the upper tail results in lower consumer surplus than consumers would obtain if they use the true distribution. It is easy to see that overestimation or underestimation has no effect on the myopic policy.

3.4. Risk Aversion

Previous studies on phone tariff choice have also emphasized the importance of risk aversion. Miravete (2003) and Train et al. (1989) find that consumers who are uncertain about their usage rate tend to choose flat rate phone plans to protect themselves from the downside risk of paying too much if their usage rate turns out to be high. Other researchers (e.g., Nunes 2000) do not find a relationship between tariff choice and risk aversion. To see the effect of risk aversion on in-contract usage behavior, we study how the threshold policy changes if consumers are risk averse in the following theorem. We model risk-averse expected utility by introducing diminishing marginal utility for money.

Theorem 4. (i) *Suppose that $U(y) = -e^{-\gamma y}$ with $\gamma \geq 0$ for $y \geq 0$. A consumer with higher risk aversion uses a lower optimal threshold for any k and t, $\Delta J_{\gamma_1}(k, t) \leq \Delta_{\gamma_2} J(k, t)$ for $\gamma_1 \geq \gamma_2$.*

(ii) *Suppose that $U(y)$ is increasing and concave for $y \geq 0$. A risk-averse consumer uses a lower optimal threshold than a risk-neutral individual: $q_{RA}^{\mathrm{DDT}}(k, t, p) \leq q^{\mathrm{DDT}}(k, t, p)$.*

Theorem 4 states that an individual becomes more liberal in her usage as her risk aversion increases if she uses the optimal policy or a deterministic threshold policy (DST or DDT). Risk aversion has no affect on the myopic policy.

3.5. Sunk Cost and the Taxi-Meter Effect

A rational consumer takes into account only current and future costs and benefits while making consumption decisions. However, the psychology and behavioral economics literatures show that individuals often incorrectly pay attention to sunk costs while making decisions (see, e.g., Arkes and Blumer 1985). In access services, if the contract has an access fee and a number of free units, the sunk cost of the access fee might affect the consumption decisions of some individuals. Consumers may then feel disutility proportional to the number of residual free units at the contract expiration time. Although the deterministic static threshold policy should lead consumers to use all free units in expectation, if consumers care about sunk costs, the additional asymmetric utility cost for

[9] Because DST is a special case of DDT with $(k = K, t = T)$, we only state the theorem for the DDT.

Table 1 Summary of Theoretical Predictions

Policy	Threshold at (k, t), $k > 0$	Static (inattentive)	Dynamic (attentive)	Usage with bias: Overestimation	Usage with bias: Risk aversion	Usage with bias: Sunk cost
Stochastic dynamic threshold (SDT)	$\Delta J(k, t)$	No	Yes	Conservative	Liberal	Liberal
Myopic	0	Yes	No	No change	No change	No change
Deterministic static threshold (DST)	$q^{DST} = q(K, T, p)$	Yes	No	Conservative	Liberal	Liberal
Deterministic dynamic threshold (DDT)	$q^{DDT} = q(k, t, p)$	No	Yes	Conservative	Liberal	Liberal

having excess units (compared to having too few units) may lead them to consume the service more aggressively to reduce this disutility.

On the other hand, if the consumer has no free units, or uses all the free units before the contract expiration, she has to pay p whenever she uses the service. Prelec and Loewenstein (1998) argue that coupling the payment with the consumption decreases the utility derived from the service/product. Mental accounting assumes that consumers attribute the disutility of payment for a good directly to the utility derived from its consumption (Prelec and Loewenstein 1998, Soman 2001). Paying per use lessens the utility from consumption, because the distaste of paying is attributed to the consumption at the time of usage. In contrast, payments in advance of consumption decouple consumption from payment. Several other papers in the literature (e.g., Lambrecht and Skiera 2006) call this the taxi-meter effect and suggest it may be one of the biases that consumers face when choosing among tariffs. Here, we consider whether the consumer acts differently if she has to pay each time she uses the service. We employ the imputed cost and benefit concept as described by Prelec and Loewenstein (1998) to model the taxi-meter affect. If V is the utility from the consumption and ρp is the experience utility lost due to the imputed cost, the net surplus is given by $V - (1+\rho)p$. A consumer chooses to consume if the value of the service is greater than $(1+\rho)p$. If the consumer has the taxi-meter bias (i.e., if $\rho > 0$), the consumer acts more conservative than a rational individual when consuming costly units (i.e., if $p > 0$), but acts rationally when consuming free units ($p = 0$).

Table 1 summarizes the theoretical predictions discussed in this section. Next, we will test these predictions using a cell phone experiment.

4. Experiment 1: Design

To examine the structure of the consumers' consumption decisions of access services and how the four policies we identified previously relate to their decision rules, we designed a laboratory experiment to simulate the cell phone consumption problem. Subjects performed four consumption decision tasks, as well as several tasks designed to identify biases in subject beliefs, risk aversion, regret aversion, and the sunk cost fallacy. Lastly, subjects filled out a brief demographic questionnaire.

4.1. Cell Phone Consumption Task

In the simulated cell phone consumption task, subjects received 30 phone calls. The calls had one of five possible values (drawn randomly and independently) for answering the call: \$0.15, \$0.30, \$0.45, \$0.60, or \$0.75.[10] To allow subjects to have potentially biased beliefs about the distribution of call values, subjects were not told the exact probabilities of each call value.[11] Instead, before the first consumption task, subjects were told that call values would be drawn independently in each period from the same distribution throughout the experiment, and then were allowed to draw sample outcomes from the distribution. Subjects were allowed to draw as many samples as they wished before continuing with the experiment.[12]

In each period, we used the strategy method to elicit from the subjects whether they would answer each type of phone call. That is, they were asked (for each call value) if they would want to answer the call or not, and they were told that their conditional strategy would be used to answer the call. This allows us to observe a subject's complete consumption strategy for each phone call, rather than only observing the outcome of their decision. Subjects were then told the actual call value and whether they had answered it (according to their stated strategy).

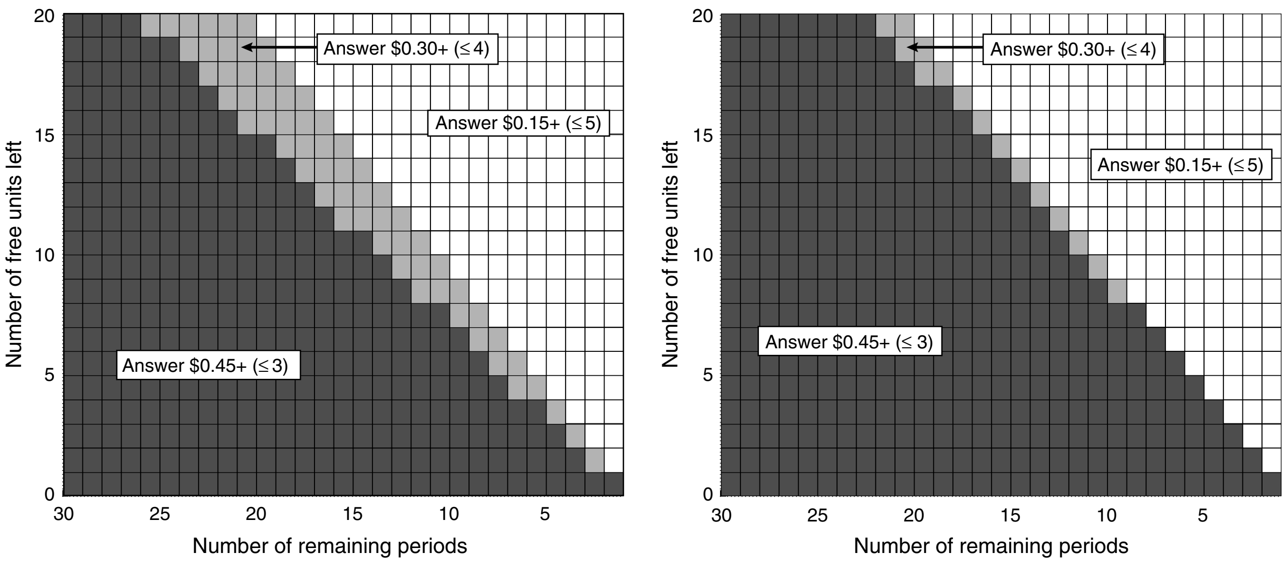
Subjects participated on one of three treatments that defined their cell phone plan. In the "0 Calls" treatment, subjects began with zero free calls, but had no "monthly fee." In the "10 Calls" treatment, subjects began with 10 free phone calls and had a \$3.50 monthly fee deducted from their payoff at the end of the task. In the "20 Calls" treatment, subjects

[10] The call values had the following probabilities: $P(\$0.75) = 0.15$, $P(\$0.60) = 0.25$, $P(\$0.45) = 0.25$, $P(\$0.30) = 0.25$, $P(\$0.15) = 0.10$.

[11] This assumption is also realistic—consumers are unlikely to know in advance the exact probability that they will receive a certain number of important phone calls in the coming month. Instead, consumers must base their beliefs on previous experience of how likely it is they will use a given number of minutes.

[12] To speed up the experiment, subjects saw 10 random outcomes at a time, and they also saw a table of the number of observations of each call value from all of the sampled outcomes.

Figure 1 **Stochastic Dynamic Threshold Policy (Left) and Deterministic Dynamic Threshold Policy (Right)**



Note. Each policy both by the call values answered (e.g., \$0.45+) and the call types answered (e.g., ≤ 3) are shown.

began with 20 free calls, and had a \$7.00 monthly fee. In all three treatments, subjects had to pay \$0.35 to answer a call if they did not have any free calls left. Subjects could see both the plan details, as well as the current period and the current number of free calls left, throughout the decision task. The 10 Calls and 20 Calls treatments allow us to study different aspects of behavior. In the 20 Calls treatment (depending on the sequence of call valuations), subjects may face decisions where the optimal SDT and DDT policies differ, whereas in the 10 Calls treatment, the overusage of free calls is more likely to cause subjects to earn lower payoffs (since, on average, they should have a lower period of pay-per-use).

Figure 1 displays what the optimal SDT and DDT policies for a given optimal SDT and DDT policies for a given number of free calls left and number of remaining periods for the specific call-value distribution used in the experiment.[13] The two policies are very similar, with the optimal policy answering the \$0.30 call in more cases than the

After receiving all 30 calls, subjects were informed of their monthly fee, the total value of all calls answered, the total charges for answering calls, and their overall payoff. One of the cell phone tasks was selected randomly for payment.

4.2. Beliefs About the Value Distribution

At the start of each consumption task, we elicited subject beliefs about the upper and lower tails of the call-value distribution. Subjects were asked to guess how many of the 30 calls would be \$0.75 calls and how many would be \$0.15 calls. For each guess, subjects had \$0.50 added to their task payoff if they were correct or \$0.25 if they were within one of the correct answer.

4.3. Risk Aversion

To measure risk aversion, after the fourth consumption task, subjects were asked to perform the paired lottery choice task from Holt and Laury (2002).[14] Subjects were asked to make 10 choices between a "safe" lottery and a "risky" lottery. Both lotteries had two potential outcomes (with the risky lottery having a larger difference between the payoffs), and both had the same probability of high- and low-value outcomes. The probability of the high outcome increased in 10% increments from 10% to 100%. For example, in one decision the safe lottery was (30% chance of \$2.00, 70% chance of \$1.60), whereas the risky lottery was (30% chance of \$3.85, 70% chance of \$0.10). Therefore, the safe lottery has a higher expected value when the probability of the high outcome is small, and the risky lottery has a higher expected value when the probability of the high outcome is large. Following Holt and Laury (2002), we use the number of safe lottery choices as a measure of risk aversion.[15] One of the lottery decisions was randomly selected for payment.

[13] We identify the SDT based on the value function described in §3.1.

[14] We include the diagnostic measures of risk aversion, regret aversion, etc. at the end of the experiment to avoid any potential contamination of subjects' consumption behavior—which is the main focus of our study. However, given that these measures have largely no effect on consumption behavior, we do not feel that contamination across tasks distorted behavior in our experiment.

[15] As in Holt and Laury (2002), many of our subjects do not have a single decision where they switch from choosing the safe lottery to choosing the risky lottery.

4.4. Regret Aversion

We use two measures of regret aversion, based on Zeelenberg et al. (1996) and Zeelenberg and Beattie (1997). Both measures exploit the fact that a regret-averse individual does not like to discover that the choice she made led to a worse outcome than another possible option. Our first measure presents subjects with two additional lottery choices, with the probabilities set (based on the subject's previous lottery choices) so that the subject should be roughly indifferent between the two lotteries. However, for these two choices, the subject will be informed of the outcome of the safe lottery or the risky lottery, in addition to whichever lottery they chose. Therefore, a regret-averse individual should choose the safe lottery for the first choice, and the risky lottery in the second choice—i.e., she should choose the lottery that she will already be informed about. This means that the subject will not be able to compare the outcomes and therefore will avoid regret.

The second measure uses two ultimatum game choices. For the first game, proposers will only be told if the responder accepted or rejected her offer. For the second game, proposers will also be told the smallest offer the responder would have accepted. Zeelenberg and Beattie (1997) show that in the second case, regret-averse proposers make more aggressive (i.e., lower) offers to avoid the regret of making a higher offer than is necessary to avoid rejection. All subjects make decisions both as proposers and as responders for both games.

4.5. Sunk Cost

To identify subjects exhibiting the sunk cost fallacy, we asked subjects to make a decision for a hypothetical scenario adapted from Arkes and Blumer (1985). In this scenario, subjects were told that they had accidentally bought tickets for two ski trips on the same weekend. They had paid more for one trip, but expected to enjoy the other trip more. They were told they could not return either ticket, and were asked to choose which trip they would go on. Therefore, subjects who say they would go on the less enjoyable trip that they had paid more for exhibit the sunk cost fallacy.

4.6. Cognitive Ability

For a subset of our sessions, we also included a simple measure of cognitive ability. We asked subjects to solve the three question *cognitive reflection task* from Frederick (2005). Each of the three questions has an intuitive, but incorrect, answer, whereas seeing the correct answer takes somewhat deeper thinking. Frederick argues that the cognitive reflection task (CRT) score is a simple measure of a kind of cognitive ability that correlates well with decision-making heuristics and biases such as present-biased intertemporal preferences and risk seeking to avoid losses. The CRT score also correlates well with SAT and ACT scores. Subjects in our experiment were paid \$0.25 for each correct answer.

5. Experiment 1: Results

We had a total of 104 students at the University of Michigan participate as subjects, with 36 subjects in the 0 Calls treatment, 36 subjects in the 10 Calls treatment, and 32 subjects in the 20 Calls treatment.[16] Sessions lasted approximately 50 minutes, and subjects earned \$12.94 on average.

We will begin by first examining subject decisions for individual calls, then consider the structure of their decision rule throughout the cell phone task (including how such decision rules relate to the four policies we identified previously), and, finally, we will consider how well calibrated subjects' decision rules are, on average, compared to the optimal policy.

5.1. Single Call Decisions

We first examine subjects' consumption decisions. Figure 2 displays for each call type within each treatment the percentage of decisions to answer the call. We display the answer rates for decisions with and without free calls separately. Answer rates without free calls are similar across all treatments: subjects answered the three highest call types in 70% to 90% of decisions.[17] When subjects have free calls, they answer the three highest-value calls in 80% to 90% of decisions. However, subjects answer the lowest-value calls at substantially higher rates: between 30% and 50% of the time.

Columns (1)–(5) of Table 2 report the results of regressing subjects' answer policy for each call type on treatment dummies (with and without free calls), task dummies, and the number of remaining periods. We include subject random effects and cluster the standard errors at the subject level.[18] As in Figure 2, differences between treatments in answering the \$0.75 call when the subject has no free calls left. There is no significant difference between treatments for the \$0.60 and \$0.45 calls; however, subjects are significantly more likely to answer \$0.30 and \$0.15 when they have free calls remaining.

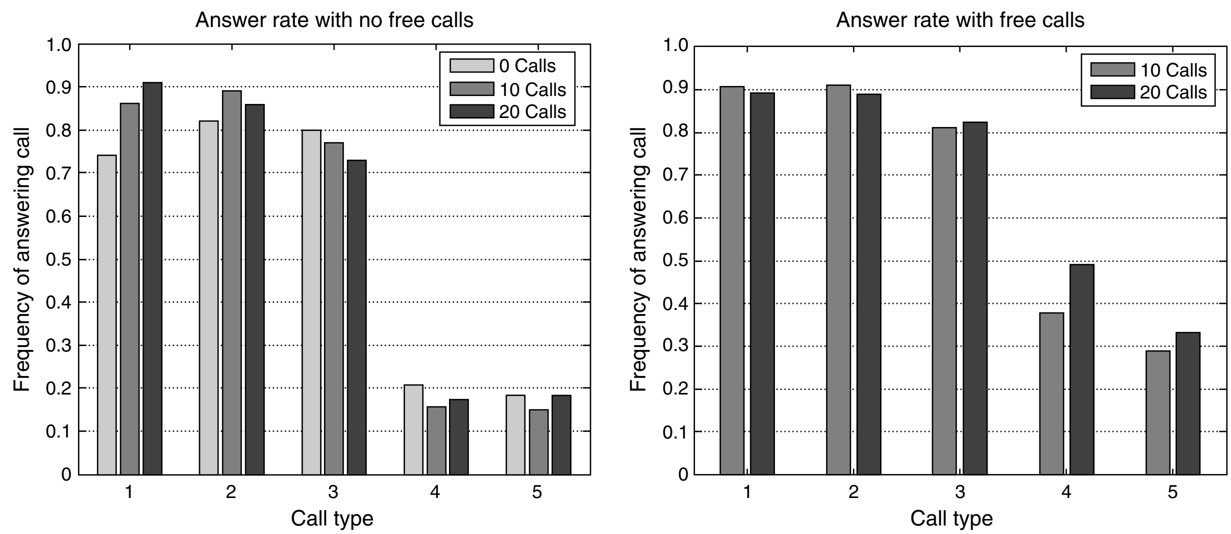
We next examine whether subjects use a threshold policy. Table 3 displays for each treatment the

[16] Three subjects in the 10 Calls treatment had to be dropped from the analysis because of a technical problem in the first session.

[17] Although the lower answer rate of the \$0.75 call in the 0 Calls treatment is a bit surprising, it does not outliers—35% of subjects choose not to answer the \$0.75 call in at least one period. However, this result is not robust—in Experiment 2, subjects in the 0 Calls treatment answer the \$0.75 call 90% of the time (see Figure 5), whereas they answered the two lowest call types in 10% to 20% of decisions.

[18] We do the same for all the regressions reported in this paper.

Figure 2 Answer Rates With and Without Free Calls



Note. The $0.75 call is denoted as call type 1, the $0.60 call is denoted as call type 2, etc.

Table 2 Answer Policy

	Answer call type					Answer policy	
	$0.75	$0.60	$0.45	$0.30	$0.15	(No. of types answered)	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Periods Left	0.000158	0.00111	0.00160	0.00110	−0.000433	−0.00489	0.00254
	(0.000579)	(0.000735)	(0.00116)	(0.00151)	(0.00140)	(0.00348)	(0.00225)
10 *Calls Treatment and* 0 *Calls Left*	0.119	0.0744	−0.0250	−0.0442	−0.0386	−0.180	−0.132
	(0.0781)	(0.0574)	(0.0700)	(0.0720)	(0.0710)	(0.183)	(0.182)
20 *Calls Treatment and* 0 *Calls Left*	0.166**	0.0467	−0.0578	−0.0202	−0.00568	−0.301	−0.195
	(0.0767)	(0.0904)	(0.0968)	(0.0732)	(0.0956)	(0.188)	(0.180)
10 *Calls Treatment and* 1+ *Calls Left*	0.167**	0.0799	−0.000700	0.165**	0.112	0.594***	0.416*
	(0.0676)	(0.0524)	(0.0698)	(0.0755)	(0.0736)	(0.203)	(0.218)
20 *Calls Treatment and* 1+ *Calls Left*	0.150**	0.0667	0.0194	0.281***	0.148**	0.657***	1.031***
	(0.0665)	(0.0549)	(0.0575)	(0.0606)	(0.0660)	(0.175)	(0.186)
10 *Calls Treatment*: *# Calls Left*							0.0223*
							(0.0134)
20 *Calls Treatment*: *# Calls Left*							−0.0352***
							(0.00885)
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.695***	0.769***	0.739***	0.193***	0.186***	3.099***	2.982***
	(0.0568)	(0.0438)	(0.0525)	(0.0525)	(0.0533)	(0.143)	(0.139)
Observations	12,120	12,120	12,120	12,120	12,120	9,681	9,681
Number of subjects	101	101	101	101	101	100	100

Notes. Standard errors clustered at the subject level reported in parentheses, with columns (1)–(5) estimated jointly as seemingly unrelated regressions. The dependent variable for columns (1)–(5) is an indicator variable if the subject answered that call type, for columns (6) and (7) it is the subject's answer threshold (i.e., the number of call types answered). The constant term reflects the policy in the 0 Calls treatment. The specification is ordinary least squares (OLS) with subject random effects, and in columns (6) and (7) the observations are restricted to periods where the subject used a threshold policy.

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

percentage of periods in each task where subjects used a threshold policy in an individual period to answer calls,[19] as well as the percent of subjects who use a threshold policy throughout the task.[20] Overall, subjects used a threshold policy for the majority of their decisions; however, somewhat fewer subjects used a threshold for every decision in a

[19] That is, we identify periods where a subject's answer choices are such that the value of every answered call type is higher than the value of every unanswered call type.

[20] The second total column shows the percentage of subjects who use a threshold policy in each period of all four tasks.

Table 3 Usage of Threshold Policies

	% of decisions with threshold policy					% subjects always using threshold policy				
Treatment	Task 1	Task 2	Task 3	Task 4	Total	Task 1	Task 2	Task 3	Task 4	Total
0 Calls	66	64	72	79	71	53	58	69	72	53
10 Calls	77	89	91	91	87	58	79	85	88	55
20 Calls	78	79	83	91	83	47	69	69	75	44

task.[21] Furthermore, subjects appeared to learn to use a threshold policy: Subjects in the 10 Calls and 20 Calls treatments were significantly more likely to use a threshold policy throughout the experiment in task 4 than they were in task 1 (test of proportions: $p < 0.01$ and $p = 0.02$). Subjects in the 0 Calls treatment were also somewhat more likely to use a threshold policy throughout ($p = 0.09$). Additionally, cognitive ability appears to play a role in whether a subject consistently uses a threshold policy. In the 0 Calls treatment, subjects who got zero correct in the CRT used a threshold throughout the task in 46% of tasks, compared to 78% for subjects who got all three questions correct (nonparametric test for trends: $p < 0.01$). Similarly, in the 10 and 20 Calls treatments, only 38% (13%) who got zero correct consistently used a threshold, compared to 95% (75%) who got three questions correct ($p < 0.01$ for both).

Given that most subjects use a threshold policy, we can characterize the five call answer decisions of these subjects by their answer threshold. We will describe an answer policy by the number of call types the subject has chosen to answer. For example, an answer policy equal to 2 means the subject wishes to answer any call worth \$0.60 or more, whereas an answer policy equal to 5 means the subject wishes to answer all five call types. Figure 3 shows for each treatment what the average answer policy was in what the average answer policy was in each period, as well as the average number of free calls remaining in each period.

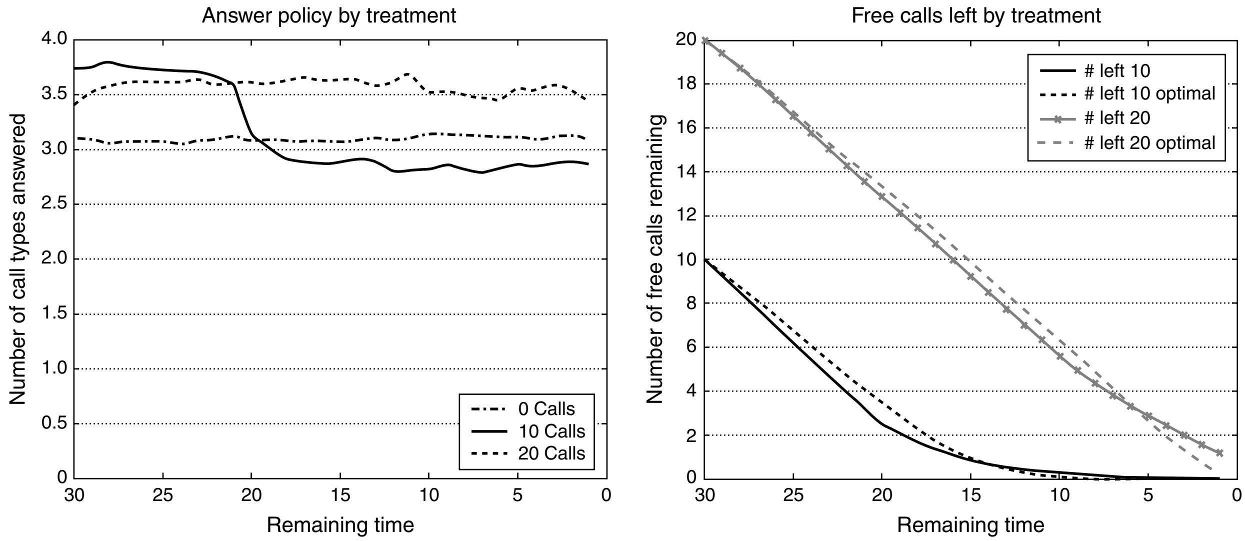
Subjects in the 0 Calls treatment choose, on average, a policy very close to their dominant strategy (answering any call worth at least \$0.45) throughout the task. Subjects choose to answer fewer call types only 6% of the time, and choose to answer more call types only 10% of the time. In the 10 Calls treatment, subjects begin by answering, on average, all but the lowest value of calls. This quickly uses up all of their free calls—subjects, on average, use their last free call in period 15. Afterward subjects act very similarly to those in the 0 Calls treatment, on average answering any call worth at least \$0.45. Similarly, in the 20 Calls treatment, subjects begin by answering, on average, slightly more than 3.5 call types. However, because their free calls last much longer (until period 26 on average), subjects in this treatment continue to answer approximately 3.5 call types, on average, throughout the task. Overall, the average answer policy is 0.5 higher when subjects have free calls left (2.96 with zero calls left versus 3.56 with one or more calls left).

In column (6) of Table 2, we regress the answer policy on the same set of variables as the previous specifications. The answer policy with zero free calls is not significantly different from three for any of the treatments in any period ($p > 0.10$ for all comparisons). The answer policy with one or more free calls is significantly larger than 3.0 for both the 10 Calls and 20 Calls treatments ($p < 0.01$ for both comparisons), and the coefficients for the 10 Calls and 20 Calls treatments are not significantly different ($p = 0.75$). In column (7) we add an additional control for the number of free calls the subject has left. Answer policies are slightly less conservative with more free calls in the 10 Calls treatment, whereas they are slightly more conservative in the 20 Calls treatment. In both treatments, however, answer policies are significantly higher for any number of free calls than they are with zero free calls. The average answer policy does not differ significantly between the 10 Calls treatment and the 20 Calls treatment when subjects have their full allotment of free calls ($p = 0.16$), however they do begin to differ significantly as subjects have relatively few free calls left—the estimated policy with one call left in the 20 Calls treatment is 3.98, compared to 3.42 in the 10 Calls treatment ($p < 0.01$).[22]

[21] Although in these data the subjects in the 0 Calls treatment are less likely to use a threshold policy, this does not appear to be a robust result. Subjects in Experiments 2 and 3 with a 0 Calls contract use a threshold approximately as often as other subjects. In Experiment 3, this is true for the subset of subjects who faced very high random prices (ruling out sorting effects).

[22] These changes in subjects' policies as they run low on free calls does not appear to be driven by differences in the composition of subjects who do and do not reach decision nodes where they have few calls remaining. Every subject runs out of calls in the 10 Calls treatment. In the 20 Calls treatment, 66% of subjects run out, and 90% of subjects end up with three or fewer calls. If we rerun specification (7) only with subjects who end up having three or fewer calls we find essentially the same results (10 *Calls Treatment* × # *Free Calls*: $\beta = 0.0241$, s.e. $= 0.0134$, $p = 0.072$; 20 *Calls Treatment* × # *Free Calls*: $\beta = -0.0296$, s.e. $= 0.00903$, $p = 0.001$). Additionally, if we look within individuals at the change in their average answer policy for their first three free calls and their last three free calls, we find the same results: Subjects in the 10 Calls treatment are significantly less conservative when they have many free calls compared

Figure 3 **Answer Policy and Free Call Usage**



5.2. Decision Rule Structure

We next want to examine what structure exists in the answer choices subjects make throughout a cell phone task—in particular, how often, when, and why do subjects change their answer choices from one period to the next. Overall, subjects in the 0 Calls treatment change their choices in 11% of periods, and choices change 9% of the time in the 10 Calls treatment and 15% of the time in the 20 Calls treatment. However, part of this is driven by the inconsistent usage of a threshold policy. When subjects use a threshold policy in two consecutive periods, they only change their policy 1%, 5%, and 7% of the time, respectively. For subjects who use a threshold policy in every period of a cell phone task, the average number of policy changes is 0.21 in the 0 Calls treatment, 1.33 in the 10 Calls treatment, and 2.08 in the 20 Calls treatment.

In Table 4 we show the fraction of subjects who use a threshold policy for every period whose profile of answer policies can be characterized into one of several simple forms.[23] Note that very few subjects change their answer policy while they have no free calls (all such subjects are in the "all others" category). Almost 90% of subjects in the 0 Calls treatment do not change their policy at all—consistent with the optimal policy. For the 10 Calls treatment, roughly a third of subjects use a single policy (with 80% of those having an answer policy of three throughout), whereas another third have one policy with free calls and a second without (of those, 63% have an answer policy of five with free calls and a policy of three without, and another 11% have a policy of four with free calls and three without). In the 20 Calls treatment, only 8% use a single policy, with the two most common policy types having a single change upon using up the last free call (53% have an answer policy of five with free calls and a policy of three without, and 41% have a policy of four with free calls and three without), or having multiple changes with free calls (with 73% of these subjects starting with a relatively conservative answer policy—i.e., high threshold—and ending with a more liberal one). Thus, most subjects have relatively few policy changes in the 0 and 10 Calls treatments (where we expect few changes based on the optimal policies), whereas subjects in the 20 Calls treatment tend to change more often.

We can also look at when the policy changes are occurring within the cell phone task. We will focus on the 10 Calls and 20 Calls treatments. Sixteen

Table 4 **Common Answer Policy Structures**

Answer policy structure	0 Calls (%)	10 Calls (%)	20 Calls (%)
Zero changes	89.01	35.29	8.43
One change, after last free call	0	34.31	20.48
One change, with free call	0	4.90	12.05
Two changes, one with free calls, one after last free call	0	1.96	10.84
Two changes, all with free calls	0	0.98	9.64
Three or more changes, one after last free call, zero after that	0	8.82	6.02
Three or more changes, all with free calls	0	0	18.07
All others	10.99	13.73	14.46

Note. This table includes only subjects who use a threshold policy in all periods.

to few (signrank test: $p < 0.001$), whereas subjects in the 20 Calls treatment are significantly more conservative with many free calls ($p < 0.001$).

[23] For this taxonomy, "after last free call" means that change in policy occurs in the period immediately following the use of the last free call.

percent of changes occur in the period immediately after the subject uses his last free call. While subjects have free calls left, they change their answer policy 21% of the time when the optimal policy changes, compared to 5% when the optimal policy does not change—suggesting subjects are more likely to change in response to the incentives captured in the optimal policy. Furthermore, 37% of the changes that occurred when the optimal policy did not change caused the subject to now match the optimal policy. While subjects have free calls left, they are more likely to change their policy when they did not answer a call last period (10% change without answering a call last period versus 4% change with answering)—with almost all of those changes to make the policy more liberal (i.e., lower the threshold and answer more call types).

5.3. Comparing Decision Rule Structure to the Optimal Policies

Next, we want to consider how the structure of subject's decisions compare to the four policies we discussed in §3. We will make this comparison in two levels. First, we will ask which policy structure is most consistent with the aggregate patterns in individual choices. Second, we will explore individual heterogeneity and identify how many individual profiles across all choices can be well described by each policy.

5.3.1. Aggregate Structure of Choices. First, it is clear from Table 2 that the *myopic heuristic* is not a good descriptor of average behavior across all subjects, since the average answer policy with free calls differs significantly in both treatments from the myopic policy of answering all calls ($p < 0.01$ for both). Table 4 similarly suggests that a *deterministic static policy*[24] does not describe the aggregate patterns of subject choices—46% of subjects in the 20 Calls treatment (where we expect policy adjustments to occur) use multiple thresholds while they have free calls remaining, and 68% of subjects use multiple thresholds with free calls remaining when they face a call sequence where the stochastic dynamic policy would change.[25] Furthermore, subjects appear to adjust their answer policy in response to the underlying incentives captured in either the *stochastic dynamic policy* or the *deterministic dynamic policy*. We can see this further in Table 5, which reports the results of regressing subjects' answer policies on dummy variables for the optimal SDT policy (odd-numbered columns) or the optimal DDT policy (even-numbered columns). In both cases, the average policy increases significantly when the optimal policy changes from three to four or five (i.e., the subject has many free calls left relative to the number of remaining periods), even when controlling separately for the number of free calls left. Hence, on average, subjects are increasing their answer policy significantly when the optimal DDT and SDT policies indicate they should.[26] Additionally, note that the coefficient on having at least one free call is approximately 0.5, and that the responses to increases in the optimal policy are correspondingly 0.5 smaller than we would expect—this is an initial indication that subjects are (on average) biased toward being too liberal in using their free calls. We analyze this bias in more depth below.

Both of the dynamic policies appear to be good proxies for the overall pattern of subjects' answer policies. Subjects' answer policies match the optimal SDT exactly in 62.4% of decisions, including 41.6% of decisions with one or more free calls. Similarly, decisions exactly match the optimal DDT in 63.9% of decisions, including 44.8% of decisions with one or more free calls. If we want to identify which policy structure is a better match for subject behavior, we need to focus on the 20 Calls treatment, since that is the only treatment where a sample path can lead to different optimal policies. In the 20 Calls treatment, the optimal stochastic and deterministic policies differ in 12% of observations where the subject has free calls, and 49% of cell phone tasks involve at least one such period. When the stochastic and deterministic policies differ, subjects match the SDT in 9.8% of decisions, whereas they match the DDT in 50.7% of decisions (this difference is significant: signed rank test $p < 0.01$). Furthermore, if we compare a regression specification that predicts the 20 Calls treatment with either just the optimal SDT policy or just the optimal DDT policy, a Vuong test indicates that the optimal deterministic policy provides a significantly better fit ($p < 0.01$). This suggests that both of the optimal dynamic policies are good predictors of the average behavior of all subjects; however, the deterministic heuristic matches average behavior more closely.

[24] Note that in our experiment, the optimal (i.e., without bias) deterministic static policy is to answer calls of value \$0.45 or higher in every period. We discuss biased versions of our policy types in §5.4; however, we note here that the biased version of the deterministic static policy (i.e., answer calls of value \$0.30 or higher until running out of free calls) describes only 3% of our subjects in the 10 Calls treatment and 5% of subjects in the 20 Calls treatment.

[25] Similarly, 88% of subjects use multiple thresholds with free calls when the deterministic dynamic policy would change.

[26] While we look at average behavior here, we find consistent results at the individual level. If we compare within a subject the average answer policy used when the optimal policy is three, four, or five, we find that answer policies increase significantly the optimal policy changes from three to four and again from four to five for both the DDT and SDT (signrank test: $p < 0.001$ for all comparison). This suggests that our results for aggregate behavior are not due to selection effects from subjects with different kinds of policies being differentially likely to reach decisions where the optimal policy has increased.

Table 5 Comparison to Optimal Policy

Variables	(1)	(2)	(3)	(4)	(5)	(6)
# Periods Left	0.0184***	0.0176***	0.00174	0.000965	0.00696***	0.00642***
	(0.00348)	(0.00354)	(0.00240)	(0.00247)	(0.00227)	(0.00223)
Optimal Policy = 4	0.400***	0.520*	0.226**	0.401	0.200**	0.412
	(0.107)	(0.314)	(0.107)	(0.308)	(0.102)	(0.304)
Optimal Policy = 5	1.553***	1.453***	1.263***	1.206***	1.170***	1.120***
	(0.283)	(0.276)	(0.296)	(0.287)	(0.296)	(0.287)
10 *Calls Treatment and* 0 *Calls Left*			−0.135	−0.140	−0.101	−0.105
			(0.182)	(0.182)	(0.182)	(0.182)
20 *Calls Treatment and* 0 *Calls Left*			−0.316	−0.313	−0.233	−0.229
			(0.193)	(0.191)	(0.186)	(0.185)
10 *Calls Treatment and* 1+ *Calls Left*			0.545***	0.551***	0.422*	0.421*
			(0.200)	(0.200)	(0.216)	(0.216)
20 *Calls Treatment and* 1+ *Calls Left*			0.499***	0.528***	0.797***	0.829***
			(0.193)	(0.188)	(0.202)	(0.197)
10 *Calls Treatment*: *# Calls Left*					0.0154	0.0162
					(0.0132)	(0.0132)
20 *Calls Treatment*: *# Calls Left*					−0.0268***	−0.0275***
					(0.00865)	(0.00873)
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	2.937***	2.964***	2.998***	3.009***	2.915***	2.923***
	(0.0838)	(0.0841)	(0.138)	(0.138)	(0.139)	(0.139)
Observations	9,681	9,681	9,681	9,681	9,681	9,681
Number of subjects	100	100	100	100	100	100

Notes. Standard errors clustered at the subject level reported in parentheses. The dependent variable is the number of call types answered (e.g., answer \$0.15 and above is denoted as "5," \$0.30 and above is denoted as "4," etc.). Odd-numbered columns use the optimal SDT policy as the independent variable, and even-numbered columns uses the optimal DDT policy. The specification is OLS with subject random effects, and the observations are restricted to periods where the subject used a threshold policy.

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

5.3.2. Individual-Level Comparison of Policies. We next look each subject's profile of 30 choices, and ask which policy structure can best describe that subject. For each subject and each policy structure, we calculate the total absolute difference between a subject's policy in each period and the predicted policy for that period. Panel A of Table 6 reports the fraction of subjects who exactly match each policy (i.e., zero total difference), as well as the fraction of subjects who have that policy as their closest match (i.e., smallest total difference). Note that these matches are not exclusive, as two (or more) policies can coincide along a sample path. We see that there is a reasonable amount of heterogeneity—particularly in the 20 Calls treatment (where the policies are more likely to differ).

Because many subjects can be explained by more than one policy type, we break this analysis down further for the 20 Calls treatment in panel B. Here we show the fraction of subjects that have each combination of closest matches that we observe in our data; 55% of our subjects have a unique closest match—with the myopic and DDT policies representing approximately 20% each. Another 10% are equally close to DDT and either static (DST) or SDT, and about 25% are equally close to all three. Finally, 10% are equally consistent with all four policy types. Overall, we have 30% of our subjects who can only be described by the static and myopic policies (and are not consistent with a dynamic policy), 30% of our subjects who can only be described by a dynamic policy (and are not consistent with the static and myopic policies), and 40% of subjects whose choices are consistent with both the static and dynamic structures. Hence, although the analysis above indicates that a dynamic policy type is the best single descriptor of the aggregate pattern of choices, there is a substantial fraction of subjects who are definitively using a static policy type.

Panel C reemphasizes the unique description of behavior that our proposed DDT heuristic policy provides. We make pair-wise comparisons of the static, DDT, and SDT policies, and calculate the fraction of subjects who are consistent with one policy but not the other. When DDT is distinguishable from the static policy, almost two and a half times as many subjects are described by DDT. Similarly, when DDT is distinguishable from SDT, more than two and a half times as many subjects follow DDT. By contrast, when SDT and static are distinguishable, roughly equal numbers of subjects follow each type.

Table 6 Individual-Level Policy Identification

Panel A: Identifying individual types

	Myopic	DST	DDT	SDT
10 Calls				
Exact match	21.57%	29.41%	29.41%	29.41%
Closest match	40.20%	65.69%	65.69%	65.69%
20 Calls				
Exact match	10.84%	6.02%	7.23%	4.82%
Closest match	30.12%	50.60%	62.65%	46.99%

Panel B: Breakdown of 20 Calls closest matches

% match one type		% match two or more types	
Myopic	19.28%	DST and DDT	7.23%
DST	8.43%	DDT and SDT	2.41%
DDT	18.07%	DST, DDT, and SDT	24.10%
SDT	9.64%	All policies	10.84%
Total	55.42%	Total	44.58%

Panel C: 20 Calls closest match model comparison

DDT vs. DST		SDT vs. DST		DDT vs. SDT	
DDT but not DST	20.48%	SDT but not DST	12.05%	DDT but not SDT	25.30%
DST but not DDT	8.43%	DST but not SDT	15.66%	SDT but not DDT	9.64%

Notes. "Exact match" denotes a subject whose profile of choices perfectly coincides with a particular policy. "Closest match" indicates which policy (or policies) has the smallest total absolute difference from the subject's choices.

5.4. Comparing Decision Rule Calibration to the Optimal Policy

In addition to the decision policy structure, we also examine where the average answer threshold is relative to the optimal SDT policy. We have already seen in Figure 2 and Table 2 that subjects are quite liberal in answering low-value calls. Figure 4 shows for each treatment the average difference between subjects' actual answer policy and the optimal policy in each period, as well as the percentage of subjects who have overused their free calls in each period. Subjects in the 10 and 20 Calls treatments answer too many calls early in the task (when they still have free calls left), and subjects use a policy close to the optimum when they do not have any free calls. Overall, in the 10 Calls treatment, between 40% and 50% of subjects have too few calls during the first half of the task, whereas in the 20 Calls treatment, between 30% and 50% of subjects have too few free calls throughout the task.

We confirm these results by regressing the difference between the actual and optimal SDT policy on the number of remaining periods, treatment dummies, and dummies for having one or more free calls

Figure 4 Average Deviation from Optimal Policy and Average Overusage of Free Calls

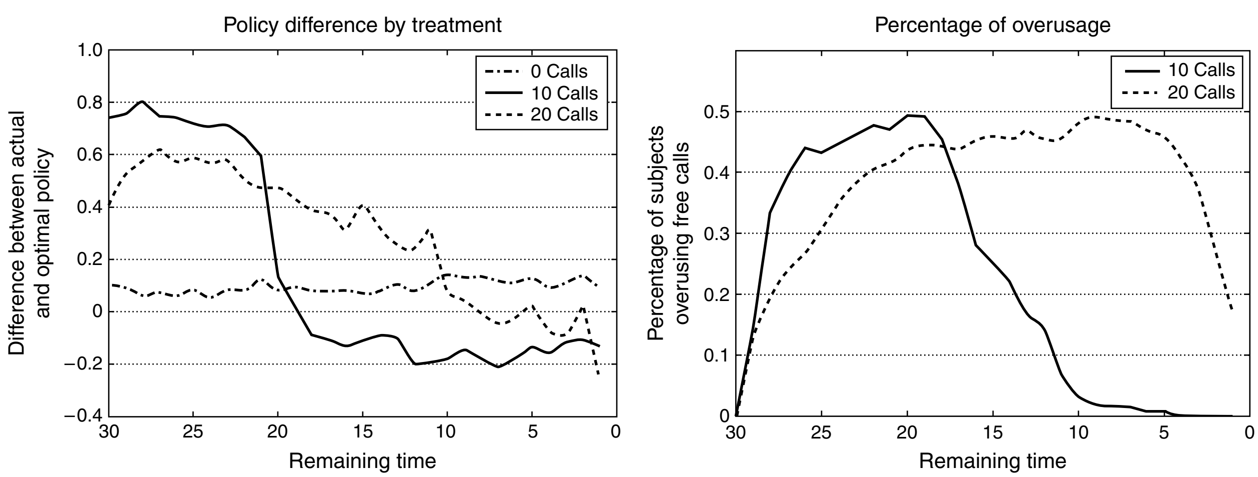


Table 7 Deviations from Optimal Policy

	Difference from optimal policy				Incorrectly answer	
	Stochastic	Deterministic	Stochastic	Deterministic	\$0.30	\$0.15
Variables	(1)	(2)	(3)	(4)	(5)	(6)
# Periods Left	0.00732***	0.00489*	0.0120***	0.00961***	0.00614***	0.00258
	(0.00251)	(0.00262)	(0.00333)	(0.00303)	(0.00177)	(0.00161)
10 *Calls Treatment and* 0 *Calls Left*	−0.0959	−0.114	−0.0650	−0.0829	−0.0117	−0.0191
	(0.181)	(0.181)	(0.181)	(0.182)	(0.0724)	(0.0715)
20 *Calls Treatment and* 0 *Calls Left*	−0.341*	−0.323*	−0.273	−0.255	0.0362	0.0280
	(0.201)	(0.196)	(0.197)	(0.192)	(0.0747)	(0.0982)
10 *Calls Treatment and* 1+ *Calls Left*	0.505**	0.522***	0.429**	0.424**	0.145*	0.0973
	(0.198)	(0.199)	(0.213)	(0.215)	(0.0754)	(0.0787)
20 *Calls Treatment and* 1+ *Calls Left*	0.324	0.438**	0.546***	0.669***	0.547***	0.308***
	(0.206)	(0.199)	(0.210)	(0.207)	(0.0719)	(0.0889)
10 *Calls Treatment*: # *Calls Left*			0.00742	0.0113	−0.00263	−0.00115
			(0.0136)	(0.0135)	(0.00481)	(0.00478)
20 *Calls Treatment*: # *Calls Left*			−0.0209**	−0.0218**	−0.0243***	−0.0146***
			(0.00876)	(0.00850)	(0.00403)	(0.00454)
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.0841	−0.0513	−0.159	−0.126	0.115**	0.139**
	(0.141)	(0.141)	(0.147)	(0.145)	(0.0547)	(0.0556)
Observations	9,681	9,681	9,681	9,681	12,120	12,120
Number of subjects	100	100	100	100	101	101

Notes. Standard errors clustered at the subject level reported in parentheses, with columns (5) and (6) estimated jointly as seemingly unrelated regressions. The dependent variable in columns (1) and (3) is the difference between the subject's answer threshold and the optimal stochastic dynamic threshold. The dependent variable in columns (2) and (4) is the difference between the subject's answer threshold and the optimal deterministic dynamic threshold. The dependent variable in columns (5) and (6) is an indicator variable that equals 1 if the subject answered the given call type when the optimal stochastic policy would not. The specification is OLS with subject random effects, and the observations in specifications (1)–(4) are restricted to periods where the subject used a threshold policy. The constant term reflects the policy in the 0 Calls treatment.

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

left. The results are reported in columns (1) and (3) of Table 7. In columns (2) and (4), we also report the difference between the actual policy and the optimal DDT policy. For both benchmark optimal policies, the average subject answers significantly more call types than is optimal in both the 10 and 20 Calls treatments, with the size of the deviation decreasing in the 20 Calls treatment as the subject has fewer free calls remaining.[27] In both treatments, the average deviation from the optimal policy is approximately 0.5 (with no significant difference in the deviation between the treatments—$p > 0.10$ in all specifications). In columns (5) and (6), we conduct a similar analysis using the full data set (i.e., for observations both with and without a threshold policy). We construct an indicator variable that equals 1 if the subject answered the \$0.30 call (or the \$0.15 call) when the optimal policy did not, and then regress that variable on the same set of controls. We find very similar results—subjects with free calls are significantly more likely to incorrectly answer low-value calls (for \$0.30 calls this is significant in both treatments, whereas for \$0.15 calls this is only significant 20 Calls treatment).

This overusage of free calls, in turn, leads many subjects to answer too many calls in total and therefore earn less from the cell phone consumption task then the optimal policy would provide. Table 8 reports for each treatment the number of answered

[27] We find similar results within an individual subject if we compare average policy differences when the subject is using his first three free calls versus his last three free calls. For both the DDT and SDT, the average policy difference is higher when the subject is using his last few calls (signrank: $p < 0.001$ for both).

Table 8 Comparison to Optimal Policy Outcomes

	Optimal policy		All subjects		
Treatment	Answered calls	Avg. payoff (\$)	Answered calls	Avg. payoff (\$)	% sub-optimal
0 Calls	19.40	4.37	17.45	3.23	89
10 Calls	19.20	4.33	19.11	3.55	90
20 Calls	21.05	4.33	21.20	3.17	95
	Optimal policy		**Always use threshold**		
Treatment	Answered calls	Avg. payoff (\$)	Answered calls	Avg. payoff (\$)	% sub-optimal
0 Calls	19.32	4.35	19.82	4.28	82
10 Calls	18.98	4.26	19.60	3.97	87
20 Calls	21.12	4.38	22.24	3.95	94

calls and the average payoff under the optimal policy, as well as the averages observed for all subjects and for subjects who used a threshold policy throughout the task.

Subjects in the 0 Calls treatments earn significantly less than the optimal policy (signrank test: $p < 0.01$). The majority of this difference is because subjects who do not use threshold policies answer too few calls ($p < 0.01$), however even those who always use a threshold policy earn less than the optimal amount ($p < 0.01$). In the 10 Calls treatment, subjects who consistently use a threshold policy earn higher payoffs, but even these subjects answer too many calls ($p = 0.05$) and earn significantly less than they would if they used the optimal policy ($p < 0.01$). In the 20 Calls treatment, subjects who consistently use a threshold also answer more calls than is optimal ($p < 0.01$), and both threshold using and nonusing subjects earn significantly less than the optimal payoff ($p < 0.01$ for both), although subjects who use a threshold policy earn 90% of the optimum. Furthermore, in both the free calls treatments, subject misallocate their calls: the average value the answered calls of threshold-using subjects is significantly lower than the optimum ($p < 0.01$ for both), i.e., subjects answer too many low-value calls relative to high-value calls.[28]

5.5. Determinants of Overuse

We now examine what behavioral factors can help explain why subjects overuse their free calls. In our data, subjects tend to overestimate the frequency of both high-value \$0.75 and low-value \$0.15 calls, even after controlling for their initial outcome sample. On average, subjects guessed 0.85 more high-value calls and 2.17 more low-value calls than they should expect based on the proportion observed in the sample—both are significantly greater than 0 (t-test: $p < 0.01$, $p < 0.01$); 35% of subjects overestimate the number of high-value calls by at least one, whereas 66% of subjects overestimate the number of low-value calls by at least one. Furthermore, many subjects incorrectly believe the distribution is symmetric: 30% guessed that there would be an equal number of low and high-value calls.

Table 9 replicates Table 7 while including the difference between subjects' guesses and the expected number of calls as an additional control (we use both the true expectation and the sample-based expectation). Subjects in the 20 Calls treatment who mistakenly overestimate the number of low-value calls (which implies undervaluing future consumption opportunities) are significantly more aggressive in using free calls. For the average amount of overestimate, this leads to a predicted increase in the answer policy of 0.11 to 0.13. There is a corresponding decrease in the estimated coefficient for the main treatment effect. This suggests that the mistaken beliefs mechanism explains part (but not all) of the overusage of free calls in the 20 Calls treatment. Although the coefficients on beliefs in the 10 Calls treatment are directionally consistent with this mechanism, they are not significant.

Subjects' mistaken beliefs appear to be persistent throughout the experiment. Although beliefs become somewhat more accurate after the first task, the distribution of the \$0.75 and \$0.15 beliefs are not significantly different in the last two tasks (rank-sum test $p > 0.60$ for both), with the average of both beliefs significantly larger than zero (t-test: $p < 0.01$ for both). This suggests that subjects are not approaching correct beliefs over the experiment. Furthermore, the error in subjects' beliefs was not significantly correlated with the number of sample outcomes they examined.[29]

We also consider risk aversion, regret aversion, the sunk cost fallacy, and cognitive ability as alternative mechanisms.[30] None of these factors can explain our overusage results. Risk aversion, regret aversion, and the sunk cost fallacy are not significant predictors of behavior in any treatment; the only significant effect of cognitive ability is that subjects with lower CRT scores are significantly more conservative (and in fact too conservative) in the 0 Calls treatment.[31] Together this suggests that the overuse bias is partially driven by mistaken beliefs, and otherwise unrelated to the other behavioral factors we measure.

6. Experiment 2

We conducted a second experiment that replicates the first experiment, but provides subjects with the exact distribution of call values (instead of having them draw sample outcomes from the distribution). Although this is arguably less realistic than the experience-based design of our main experiment, providing a complete description allows us to test whether our results are an artifact of giving subjects only incomplete information about the value distribution. A total of 36 students participated, with 18 each in the 0 Calls and 10 Calls treatments.

[28] We find similar results by comparing actions to the deterministic dynamic threshold—more than 65% of subjects received a suboptimal payoff in the 10 Calls and 20 Calls treatments, due to answering significantly too many low-value calls ($p < 0.01$ for both).

[29] The median sample size was 100 outcomes. The rank correlations between guess and sample size were $\rho = 0.0492$ and $\rho = -0.0714$ ($p = 0.32$ and $p = 0.15$, respectively).

[30] Regression results are available upon request from the authors.

[31] This is somewhat puzzling, since the 0 Calls treatment is a much simpler consumption problem. However, because the CRT measures the subjects' depth of thinking, it may be proxying for boredom and impatience with the task, rather than calculating ability.

Table 9 Effect of Mistaken Beliefs

	Difference from optimal policy			
	Stochastic		Deterministic	
Variables	(1)	(2)	(3)	(4)
# Periods Left	0.0115***	0.0115***	0.00915***	0.00915***
	(0.00326)	(0.00326)	(0.00297)	(0.00296)
10 *Calls Treatment and* 0 *Calls Left*	−0.103	−0.100	−0.119	−0.117
	(0.200)	(0.197)	(0.200)	(0.198)
20 *Calls Treatment and* 0 *Calls Left*	−0.432*	−0.436*	−0.401*	−0.402*
	(0.230)	(0.228)	(0.223)	(0.221)
10 *Calls Treatment and* 1+ *Calls Left*	0.400*	0.403*	0.397*	0.400*
	(0.233)	(0.230)	(0.235)	(0.231)
20 *Calls Treatment and* 1+ *Calls Left*	0.396*	0.391*	0.531**	0.529**
	(0.239)	(0.237)	(0.233)	(0.232)
10 *Calls Treatment*: *# Calls Left*	0.00795	0.00793	0.0117	0.0117
	(0.0136)	(0.0136)	(0.0135)	(0.0135)
20 *Calls Treatment*: *# Calls Left*	−0.0196**	−0.0196**	−0.0206**	−0.0206**
	(0.00873)	(0.00872)	(0.00844)	(0.00844)
$0.75 *Guess—E*[# $0.75] *and* 0 *Calls Treatment*	0.0156	0.0161	0.0162	0.0168
	(0.0177)	(0.0175)	(0.0177)	(0.0174)
$0.75 *Guess—E*[# $0.75] *and* 10 *Calls Treatment*	−0.0230	−0.0239	−0.0237	−0.0246
	(0.0340)	(0.0339)	(0.0344)	(0.0343)
$0.75 *Guess—E*[# $0.75] *and* 20 *Calls Treatment*	−0.00851	−0.00647	−0.00348	−0.00171
	(0.0326)	(0.0325)	(0.0318)	(0.0316)
$0.15 *Guess—E*[# $0.15] *and* 0 *Calls Treatment*	−0.0170	−0.0179	−0.0171	−0.0180
	(0.0299)	(0.0295)	(0.0302)	(0.0298)
$0.15 *Guess—E*[# $0.15] *and* 10 *Calls Treatment*	0.0109	0.0112	0.0108	0.0111
	(0.0231)	(0.0231)	(0.0232)	(0.0232)
$0.15 *Guess—E*[# $0.15] *and* 20 *Calls Treatment*	0.0498**	0.0498**	0.0435*	0.0437*
	(0.0226)	(0.0226)	(0.0259)	(0.0258)
Task Controls	Yes	Yes	Yes	Yes
Constant	−0.154	−0.156	−0.122	−0.125
	(0.166)	(0.163)	(0.164)	(0.161)
Observations	9,681	9,681	9,681	9,681
Number of subjects	100	100	100	100

Notes. Standard errors clustered at the subject level reported in parentheses. The dependent variable in columns (1) and (2) is the difference between the subject's answer threshold and the optimal stochastic dynamic threshold. The dependent variable in columns (3) and (4) is the difference between the subject's answer threshold and the optimal deterministic dynamic threshold. The specification is OLS with subject random effects, and the observations in specifications (1)–(4) are restricted to periods where the subject used a threshold policy. The constant term reflects the policy in the 0 Calls treatment. Columns (1) and (3) use the true expectation, and columns (2) and (4) use the expectation derived from the subjects' sample of outcomes.

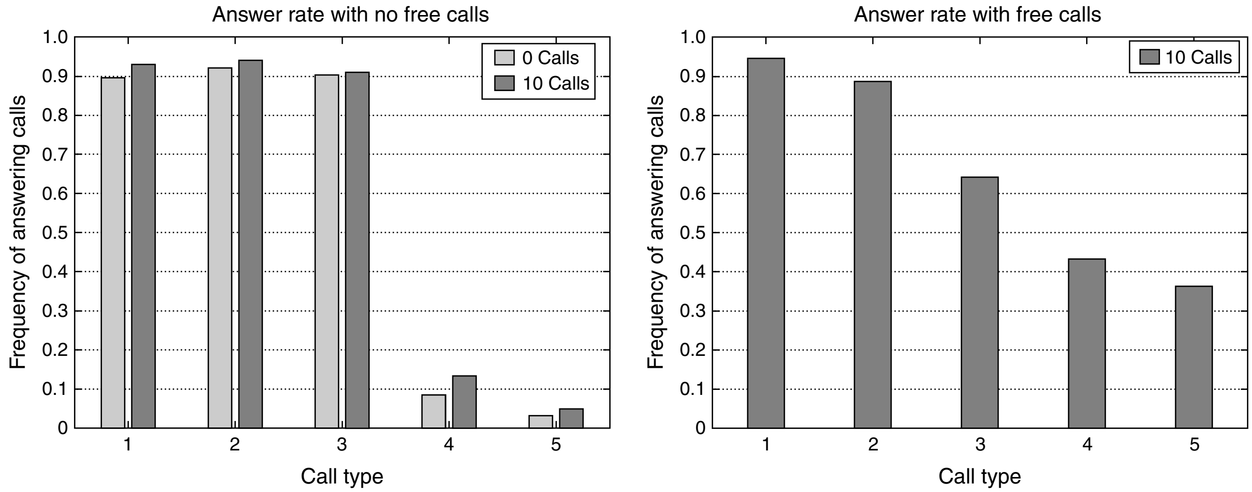
$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

6.1. Results

Figure 5 displays the answer rates for each type of call in each for each type of call in each treatment. Subjects' decisions are largely similar to our previous results, including a substantial increase in the answer rate of the $0.30 and $0.15 calls when subjects have free calls. One difference is that subjects answer the $0.45 call only 64% of the time with free calls, which is lower than in Experiment 1. We also find that subjects consistently use threshold policies, as in Experiment 1: 87% of decisions in the 0 Calls treatment and 96% of decisions in the in the 10 Calls treatment are threshold policies.

Subjects in this experiment have substantially more accurate beliefs than subjects in our previous experiment. Subjects overestimate the number of 0.75 calls by 0.66 on average (compared to 0.99 in the Experiment 1, rank-sum test $p > 0.20$) and overestimate the number of 0.15 calls by 1.03 on average (compared to 2.23 in the Experiment 1, rank-sum test $p < 0.01$). Only 22% of guesses overestimate the number of high-value calls by at least one (compared to 35%, test of proportions $p < 0.01$), and only 53% overestimate the number of low-value calls by at least one (compared to 66%, test of proportions $p < 0.01$). It is important to note that these mistakes are likely of a different nature compared to the mistaken beliefs in Experiment 1. Subjects in that experiment were making a mistake of inference and/or memory, whereas subjects in this experiment have full information, and so are making a mistake of calculation.

Figure 5 Answer Rates With and Without Free Calls



To test whether subjects overuse their free calls, we again measure the deviation from the optimal policy as the difference between a subject's answer policy and the optimal stochastic dynamic policy for each call decision. We then replicated the analysis from Table 7 by regressing the deviation from the optimal policy on treatment controls and mistaken beliefs—these results are reported in Table 10.[32] As column (1) demonstrates, we find at least as large an overuse bias for free calls as in the original experiment. This large deviation from the optimal policy remains when we include controls for beliefs. We do find an effect of mistaken beliefs, however the sign of the effect is reversed from our main experiment. Together these results confirm that the overuse bias result is robust to the sampling paradigm. Additionally, they provide further support for our interpretation that mistaken beliefs explain only a part of the overuse bias effect.

7. Experiment 3

We conducted an additional experiment to measure subjects' willingness to pay for a contract with free calls instead of having the pay-per-use contract. Procedures were the same as in Experiment 1, except that at the beginning of each task subjects were asked to state the largest monthly fee that they would be willing to pay for a 10 Calls contract or a 20 Calls contract. A random fee was then generated, and the subject was given a contract with 10 Calls (or 20 Calls) and the random fee if their WTP was greater than the fee. If the subject's WTP was smaller than the fee, the subject played the task with the 0 Calls contract and no fee. We again measured risk aversion, the sunk cost fallacy, and cognitive ability.[33] In this experiment, 77 subjects participated, with 55 providing WTP for 10 free calls, and 22 providing WTP for 20 free calls.

Table 10 Deviations from Optimal Policy

Variables	(1)	(2)	(3)
# Periods Left	0.00582*	0.00719*	0.00734**
	(0.00308)	(0.00380)	(0.00368)
10 *Calls Treatment and* 0 *Calls Left*	−0.0842	−0.0769	−0.0379
	(0.213)	(0.217)	(0.189)
10 *Calls Treatment and* 1+ *Calls Left*	0.747**	0.843**	0.857**
	(0.304)	(0.362)	(0.350)
10 *Calls Treatment*: *# Calls Left*		−0.0198	−0.0174
		(0.0234)	(0.0227)
$0.75 *Guess—E*[# $0.75] *and* 0 *Calls Treatment*			−0.0132
			(0.0343)
$0.75 *Guess—E*[# $0.75] *and* 10 *Calls Treatment*			0.000231
			(0.0180)
$0.15 *Guess—E*[# $0.15] *and* 0 *Calls Treatment*			−0.00676
			(0.0203)
$0.15 *Guess—E*[# $0.15] *and* 10 *Calls Treatment*			−0.0774
			(0.0484)
Task Controls	Yes	Yes	Yes
Constant	−0.283*	−0.301*	−0.266**
	(0.153)	(0.162)	(0.129)
Observations	3,957	3,957	3,957
Number of subjects	36	36	36

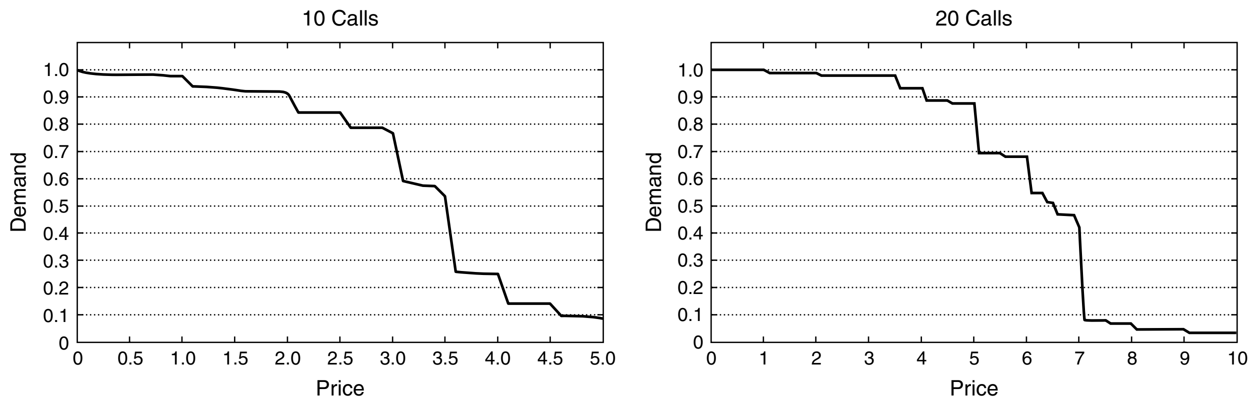
Notes. Standard errors clustered at the subject level reported in parentheses. The dependent variable is the difference between the subject's answer threshold and the optimal stochastic dynamic threshold. The specification is OLS with subject random effects, and the observations are restricted to periods where the subject used a threshold policy. The constant term reflects the policy in the 0 Calls treatment. Column (3) uses the true expectation of the value distribution.

*$p < 0.10$; **$p < 0.05$.

[32] We find similar results using the difference between the actual policy and the optimal deterministic dynamic threshold.

[33] The two regret aversion measures were eliminated to keep the session length approximately the same.

Figure 6 **Willingness to Pay Distribution for 10 and 20 Calls Contracts**



7.1. Results

Consumption behavior is similar to that observed in Experiment 1. In the 10 Calls treatment, subjects answered 18.95 calls, on average, when they had the 0 Calls contract, and they answered 19.91 calls under the 10 Calls contract. Similarly, subjects in the 20 Calls treatment answered 19.91 calls with the 0 Calls contract, and 23.61 calls with the 20 Calls contract. The average value of answered calls is $0.53 with the 10 Calls contract (compared to $0.54 in Experiment 1) and $0.52 with the 20 Calls contract (compared to $0.52 in Experiment 1).

Figure 6 shows the demand function that is implied by the distribution of subjects' willingness to pay for each contract (i.e., the fraction of subjects willing to pay at any given price). The mean willingness to pay for 10 free calls was $3.23, with a median of $3.49; 21% of subjects are willing to pay more than the "face value" of $3.50, whereas 29% were willing to pay exactly face value. For 20 free calls, the mean willingness to pay was $6.14, with a median of $6.50; 8% of subjects were willing to pay more than $7.00, with another 34% of subjects willing to pay exactly $7.00. Although fewer subjects overpay, note that paying the full pay-per-use price of $7.00 is a larger mistake for the 20 Calls contract, as the subject loses a substantial option value. Only 51% of subjects received at least 20 high-value calls, whereas every subject received at least 10 calls worth paying the per-use cost for.[34] It is important to note that willingness to pay actually increases slightly from an average of $3.14 in the first task to $3.32 in the fourth for the 10 Calls contract, and increases from $6.17 to $6.23 for the 20 Calls contract. Furthermore, the percent of subjects willing to overpay increases significantly from 13% in task one to 27% in task four for the 10 Calls contract. In the 20 Calls treatment, 5% of subjects were willing to overpay in both the first and fourth periods, with 45% willing to pay at least full price in both periods. This bias toward free units during tariff choice aligns with previous results and is consistent with consumers anticipating feeling a taxi-meter effect during consumption (despite our previous results indicating that a taxi-meter effect does not affect consumption behavior).

Table 11 reports the results of regressing subjects' willingness to pay on various individual characteristics. Although we have previously shown that mistaken beliefs significantly influence consumption decisions, they do not appear to affect subjects' willingness to pay for free units of access. This is useful from the firm's perspective, since the kind of mistaken beliefs that lead to overuse of free calls that the individual is endowed with could potentially reduce their willingness to pay ex ante for free calls if individuals use these beliefs in evaluating potential contracts. However, it appears that the decision process to select a contract does not rely upon beliefs in the same way as the decision process to use a contract. We also again find largely no effect of other behavioral factors such as risk aversion, the sunk cost fallacy, and cognitive ability, although there is a significant increase in WTP for 20 Calls among subjects who exhibit the sunk cost fallacy.

We do find that some aspects of contract choice affect subjects' consumption patterns. Table 12 reports the results of regressing the number of calls a subject answered on the randomly generated monthly fee (odd columns) or the subject's willingness to pay (even columns). We report observations for the 0 Calls contract and the 10 Calls contract for the 10 Calls treatment in columns (1) and (2) and columns (3) and (4), respectively, and the 0 Calls contract and

[34] Note that this means undervaluing the 10 Calls contract is also a mistake, since there is negligible option value in this setting. We demonstrate in Table 11 that WTP in this case is not driven by beliefs, risk attitudes, or cognitive biases. This suggests that undervaluing the contract may reflect either an overestimate of the option value or an aversion to the fixed fee.

Table 11 Willingness to Pay

	10 Calls treatment				20 Calls treatment			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$0.75 *Guess—E*[# $0.75]	0.0368 (0.0324)				−0.0336 (0.0543)			
$0.15 *Guess—E*[# $0.15]	0.000489 (0.0298)				0.0294 (0.0396)			
Risk Aversion		−0.0584 (0.0435)				−0.223 (0.175)		
Sunk Cost			0.106 (0.175)				0.756** (0.371)	
CRT Score				−0.00942 (0.0730)				0.0572 (0.238)
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	3.051*** (0.111)	3.480*** (0.257)	3.101*** (0.130)	3.151*** (0.174)	6.114*** (0.235)	7.367*** (1.033)	6.067*** (0.286)	6.059*** (0.453)
Observations	220	220	220	220	88	88	88	88
Number of subjects	55	55	55	55	22	22	22	22

Notes. Standard errors clustered at the subject level reported in parentheses. The dependent variable is the subject's willingness to pay for 10 (20) free calls. The specification is OLS with subject random effects.
$^{**}p < 0.05$; $^{***}p < 0.01$.

20 Calls contract for the 20 Calls treatment in columns (5) and (6) and columns (7) and (8), respectively. Neither the price paid nor the subjects' willingness to pay appear to affect consumption under either the 10 Calls or 20 Calls contracts. However, subjects in the 10 Calls treatment with a 0 Calls contract who faced a high fee for the 10 Calls contract answer significantly fewer calls than those who faced a lower fee. Although a portion of the effect could be driven by the mere exposure to the price, it appears that this effect can substantially be explained by a sorting effect: Subjects with a higher willingness to pay for free calls consume significantly fewer calls under a pay-for-use contract (i.e., this subset of subjects exhibit a taxi-meter effect on consumption). Therefore, preselling access units can have three beneficial effects for the firm: it extracts revenue from high-value consumers, it leads to increased usage of the service, and it screens out from the pay-per-use contract those customers who are least profitable under that contract. We did not, however, find an analogous sorting effect in the 20 Calls treatment, so the screening benefit may depend on the menu of contracts that the firm offers.

7.1.1. Optimal Pricing. We can identify the revenue maximizing fee for the 10 Calls contract based on three factors: (1) the observed distribution of consumer willingness to pay, (2) the overusage bias for consumers who purchase the 10 Calls contract, and (3) the estimated effects of price on behavior estimated in Table 12 for subjects who end up with the pay-per-use contract. Based on our data for the willingness to pay for the 10 Calls contract, the optimal fee is $3.49—a very small discount relative to the per-call price. This leads to 54% of consumers purchasing the contract, generating an average revenue of $6.88 per customer[35]—a 14.0% increase over the estimated average revenue of $6.04 if the firm does not offer a presale contract. Any price between $3.01 and $4.00 would lead to a revenue increase of up to 10%, and any price above $2.35 would increase revenue over not preselling. Because the WTP distribution shifts slightly, we also examine the optimal price given the WTP distribution from task 4. We again find that $3.49 is the optimal price, with an estimated increase in revenue of 14.2%. If we use the consumption–WTP relationship instead of the consumption–price relationship, we also find that the optimal price of $3.49.

If we do not include the effect of price (or WTP) on consumption in the 10 Calls treatment, we still find the correct price, but would overestimate the potential revenue from not preselling. Under this model, preselling units would appear to only increase revenue by 3.0%. It is therefore important to properly account for the effect of the menu of contracts on resulting consumption behavior, as this misestimate of the benefits of preselling could adversely affect firm decisions.

We can similarly find the optimal fee for the 20 Calls contract given our data. As before, we find

[35] The overusage bias would lead customers with the 10 Calls contract to answer, on average, 10.03 calls beyond their initial endowment of 10 free calls, yielding the firm a total revenue of $7.00 from this segment. For consumers with the 0 Calls contract, the price effect would lead them to answer, on average, 19.28 calls, generating a total revenue of $6.75.

Table 12 Consumption and Willingness to Pay

	10 Calls treatment				20 Calls treatment			
	0 Calls contract		10 Calls contract		0 Calls contract		20 Calls contract	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Random Fee	-1.333^{**}		0.184		-0.151		-0.230	
	(0.662)		(0.384)		(0.258)		(0.225)	
WTP		-0.882^{*}		0.183		0.0447		0.152
		(0.526)		(0.621)		(0.280)		(0.278)
Constant	23.93^{***}	21.33^{***}	19.57^{***}	19.26^{***}	20.94^{***}	19.52^{***}	24.46^{***}	22.75^{***}
	(2.481)	(1.515)	(0.781)	(2.097)	(1.706)	(1.535)	(0.844)	(1.872)
Observations	81	81	139	139	34	34	54	54
Number of subjects	47	47	55	55	19	19	21	21

Notes. Standard errors clustered at the subject level reported in parentheses. The dependent variable is the total number of calls answered. The specification is OLS with subject random effects.
$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

that the optimal price is \$6.95—a very small discount relative to the pay-per-use contract.[36] This fee leads to 47% of consumers purchasing the contract, generating an average revenue of \$7.54—an 8.3% increase over the estimated revenue of \$6.97 if the firm only offers the pay per use contract. Any price between \$6.39 and \$7.00 would lead to an increase in revenue of at least 5%, whereas any price of at least \$6.00 would increase revenue at least some amount. If we use the WTP distribution from task 4, we again find that \$6.95 is the optimal price, with an estimated increase in revenues of 8.9%.

It is interesting to note that if a firm were to only offer one contract with free calls, it would prefer to offer the 10 Calls contract. The 10 Calls contract leads to almost twice the increase in revenue as the 20 Calls contract. The 10 Calls contract provides two advantages to the firm. First, more consumers overvalue the 10 Calls contract than the 20 Calls contract, leading to greater uptake of the contract at the optimal price. Second, the firm benefits from consumer overusage by not providing too many free calls, so that there is a greater number of periods that consumers pay the per-use fee.

8. Discussion and Concluding Remarks

In this paper, we consider several plausible decision heuristics that consumers may use in consuming access services. A deterministic dynamic policy that accounts for the number of remaining free units and the amount of time left before the expiration of the contract can lead to expected consumption utility that is close to the optimal stochastic dynamic policy, but is much easier to calculate. We also consider how various decision biases (such as mistaken beliefs about the value distribution, risk aversion, the sunk cost fallacy, and the taxi-meter effect) could affect consumption decisions.

We then test our model using a dynamic consumption experiment modeled on cell phone services. We find that a majority of subjects use a threshold policy in making consumption decisions, with choices matching both the stochastic and deterministic policies in many cases. Additionally, the deterministic dynamic policy provides the best single description of average consumer behavior. However, subjects use free calls too quickly, leading to average payoffs significantly below the expected benefit under the optimal policy. Many subjects exhibit behavioral biases that significantly affect behavior: subjects who overestimate the lower tail use free units more liberally. Furthermore, these mistakes persist throughout the experiment. We also measure subjects' willingness to pay for free calls, and find that a substantial number are willing to overpay. This leads to the optimal price involving only a very small discount, and that offering the optimal three-part tariff contract increases revenue by approximately 8% to 14%.

In our study, we find support for results of Ascarza et al. (2012). They indicate that the satiation level of individuals on a three-part tariff is, on average, 31.5% greater than on a two-part tariff. We also find subjects overuse minutes when they are on a three-part tariff compared to their usage when they are on a pay-per-use contract. However, Ascarza et al. (2012) do not model the dynamics of consumer decision making or test possible explanations for overuse. They explain overuse by the additional utility that individuals may obtain from three-part tariffs since three-part tariffs may result in greater enjoyment in usage.

[36] To calculate the optimal price for the 20 Calls contract, we do not include an effect of WTP on consumption under the 0 Calls contract, because there was no significant effect observed in the 20 Calls treatment.

Grubb and Osborne (2014) estimate a structural model of contract choice and usage in cellular phone services on a data set of individual cellular phone bills. Their paper is very interesting and shares some of the same insights as our experiment. However, our results suggest caution in making some of the structural assumptions used in Grubb and Osborne (2014). On the positive side, we find supportive evidence that consumers use heuristics to simplify the consumption problem, and that many consumers may have miscalibrated decision rules. On the other hand, we find that a substantial fraction of subjects adjust their threshold,[37] and we do not find that individuals make optimal consumption decisions given their beliefs, nor do we find that individuals learn in a sophisticated fashion. Instead, we find that subjects overconsume beyond what their beliefs justify, and that both mistaken beliefs and overconsumption persist over time.

One important implication of our results for the broader literature is to caution against the common assumption that the same biases drive both tariff choice and consumption decisions. Although it is a natural assumption that the same decision process that determines how much a consumer values access units ex ante also determines how the consumer uses them, our results suggest that this need not be the case. In our experiment there is only a weak connection between a consumer's willingness to pay for free calls and the subsequent consumption decisions. Moreover, the effect of biased beliefs about the value distribution that partially explains consumption behavior plays no role in determining willingness to pay. Thus, it seems that consumers use two distinct decision processes one for tariff selection and another for consumption. This may be particularly important for empirical research, where researchers typically must make strong identifying assumptions about consumers' decisions processes to address consumer heterogeneity and selection, as well as the inability to observer consumer value distributions, beliefs, etc.

Finally, some of the earlier literature suggests hyperbolic discounting as the potential source of overusage of services (e.g., Yao et al. 2012). Our experiment cannot speak directly to this mechanism, because all consumption decisions occur within a short span of time. Although it is likely that hyperbolic discounting is indeed a contributing factor to many observed examples of overconsumption behavior, it is of note that we still find significant overusage when subjects have free minutes in settings that rule out hyperbolic discounting.

[37] In Grubb and Osborne's (2014) data, consumers were not able to find out how many minutes they had left, whereas in our experiment, subjects remaining calls were always displayed. Hence, we believe that Grubb and Osborne's assumption of complete inattentiveness is likely a good assumption for their data, but may not generalize to settings where consumers can more easily track their usage.

While our experiment is focused on the consumption decisions given a specific contract, two natural extensions are to examine further how consumers may choose between potential contracts given these consumption biases, and how firms should respond to these biases in consumption behavior in choosing what contracts to offer and how to price the contracts to maximize profit.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2013.1889.

Acknowledgments

The authors thank the editors and reviewers for their suggestions, which have significantly improved the paper.

References

Arkes HR, Blumer C (1985) The psychology of sunk costs. *Organ. Behav. Human Decision Processes* 35(1):124–140.

Ascarza E, Lambrecht A, Vilcassim N (2012) When talk is "free": The effect of tariff structure on usage under two- and three-part tariffs. *J. Marketing Res.* 49(6):882–899.

Bearden JN, Murphy RO, Rapoport A (2007) Decision biases in revenue management: Some behavioral evidence. *Manufacturing Service Oper. Management* 10(4):625–636.

Bertini M, Wathieu L (2008) Attention arousal through price partitioning. *Marketing Sci.* 27(2):236–246.

Bitran G, Rocha e Oliveira P, Schilkrut A (2008) Managing customer relationships through pricing and service quality. Working Paper 750, IESE Business School, Barcelona, Spain.

Borenstein S (2009) To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing. Working paper, University of California, Berkeley, Berkeley.

Cachon G, Feldman P (2011) Pricing services subject to congestion: Charge per-use fees or sell subscriptions? *Manufacturing Service Oper. Management* 13(2):244–260.

Cox JC, Oaxaca RL (1989) Laboratory experiments with a finite horizon job-search model. *J. Risk Uncertainty* 2(3):301–330.

DellaVigna S, Malmendier U (2004) Contract design and self-control: Theory and evidence. *Quart. J. Econom.* 119(2):353–401.

DellaVigna S, Malmendier U (2006) Paying not to go to the gym. *Amer. Econom. Rev.* 96(3):694–719.

Eliaz K, Spiegler R (2006) Contracting with diversely naive agents. *Rev. Econom. Stud.* 73(3):689–714.

Eliaz K, Spiegler R (2008) Consumer optimism and price discrimination. *Theoret. Econom.* 3(4):459–497.

Esteban S, Miyagawa E, Shum M (2007) Nonlinear pricing with self-control preferences. *J. Econom. Theory* 135(1):306–338.

Frederick S (2005) Cognitive reflection and decision making. *J. Econom. Perspect.* 19(4):25–42.

Gallego G, Şahin Ö (2010) Revenue management with partially refundable fares. *Oper. Res.* 58(4, part 1):817–833.

Grubb MD (2009) Selling to overconfident consumers. *Amer. Econom. Rev.* 99(5):1770–1807.

Grubb MD, Osborne M (2014) Cellular service demand: Biased beliefs, learning, and bill shock. *Amer. Econom. Rev.* Forthcoming.

Holt CA, Laury SK (2002) Risk aversion and incentive effects. *Amer. Econom. Rev.* 92(5):1644–1655.

Iyengar R, Jedidi K, Essegaier S, Danaher PJ (2011) The impact of tariff structure on consumer retention, usage, and profitability of access services. *Marketing Sci.* 30(5):820–836.

Kridel DJ, Lehman DE, Weisman DL (1993) Option value, telecommunication demand, and policy. *Inform. Econom. Policy* 5(2):125–144.

Lambrecht A, Skiera B (2006) Paying too much and being happy about it: Existence, causes and consequences of tariff-choice biases. *J. Marketing Res.* 43(2):212–223.

Lambrecht A, Seim K, Skiera B (2007) Does uncertainty matter? Consumer behavior under three-part tariffs. *Marketing Sci.* 26(5):698–710.

Liebman JB, Zeckhauser RJ (2004) Schmeduling. Unpublished manuscript, Harvard University, Cambridge, MA. http://www.hks.harvard.edu/jeffreyliebman/schmeduling.pdf.

Loewenstein G, O'Donoghue T, Rabin M (2003) Projection bias in predicting future utility. *Quart. J. Econom.* 118(4):1209–1248.

Miravete EJ (2002) Estimating demand for local telephone service with asymmetric information and optional calling plans. *Rev. Econom. Stud.* 69:943–971.

Miravete EJ (2003) Choosing the wrong calling plan? Ignorance and learning. *Amer. Econom. Rev.* 93(1):297–310.

Nunes J (2000) A cognitive model of people's usage estimations. *J. Marketing Res.* 37(4):397–409.

Oster SM, Scott Morton FM (2005) Behavioral biases meet the market: The case of magazine subscription prices. *Adv. Econom. Anal. Policy* 5(1):1–30.

Papastavrou J, Rajagopalan S, Kleywegt AJ (1996) The dynamic and stochastic knapsack problem with deadlines. *Management Sci.* 42(12):1706–1717.

Plambeck EL, Wang Q (2013) Implications of hyperbolic discounting for optimal pricing and information management in service systems. *Management Sci.* 59(8):1927–1946.

Prelec D, Loewenstein G (1998) The red and the black: Mental accounting of savings and debt. *Marketing Sci.* 17(1):4–28.

Randhawa R, Kumar S (2008) Usage restriction and subscription services: Operational benefits with rational users. *Manufacturing Service Oper. Management* 10(3):429–447.

Savage SL (2012) The flaw of averages: Why we underestimate the risk in the face of uncertainty (John Wiley & Sons, Hoboken, NJ).

Schotter A, Braunstein YM (1981) Economic search: An experimental study. *Econom. Inquiry* 19(1):1–25.

Soman D (2001) Effects of payment mechanisms on spending behavior: The role of rehearsal and immediacy of payments. *J. Consumer Res.* 27(4):460–474.

Thomas M, Morwitz VG (2005) Penny wise and pound foolish: The left-digit effect on price cognition. *J. Consumer Res.* 32(1): 54–64.

Tirole J (1988) *The Theory of Industrial Organization* (MIT Press, Cambridge, MA).

Train KE, Ben-Akiva M, Atherton T (1989) Consumption patterns and self-selecting tariffs. *Rev. Econom. Statist.* 71(1):62–73.

Train K, McFadden DL, Ben-Akiva M (1987) The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *RAND J. Econom.* 18(1): 109–123.

Wilson R (1993) Nonlinear pricing. (Oxford University Press, New York).

Xie J, Shugan SM (2001) Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Sci.* 20(3):219–243.

Yao S, Chen Y, Mela CF, Chiang J (2012) Determining consumers' discount rates with field studies. *J. Marketing Res.* 49(6): 822–841.

Zeelenberg M, Beattie J (1997) Consequences of regret aversion 2: Additional evidence for effects of feedback on decision making. *Organ. Behav. Human Decision Processes* 72(1):63–78.

Zeelenberg M, Beattie J, van der Pligt J, de Vries NK (1996) Consequences of regret aversion: Effects of expected feedback on risky decision making. *Organ. Behav. Human Decision Processes* 65(2):148–158.