PNAS
RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES
OPEN ACCESS

Resolving Feynman's restaurant problem reveals optimal solutions and human strategies

Brian Christian[a], Evan M. Russek[b], and Thomas L. Griffiths[c,d,1]

Edited by Ayelet Fishbach, The University of Chicago, Chicago, IL; received April 30, 2025; accepted March 31, 2026 by Editorial Board Member Timothy D. Wilson

In the 1970s, physicist Richard Feynman turned lunch with a friend into a math problem—how to optimize dish selection over multiple meals—but his handwritten notes remained a mystery for decades. Here we present the fully deciphered problem and solution, prove its optimality, generalize it to related problems, and compare the results to human behavior. The optimal policy specifies decreasing thresholds for switching from exploring new dishes to exploiting the best, with thresholds varying based on the distribution of the quality of dishes. We connect these results to the existing psychological literature on optimal stopping problems, which has explored close variants on Feynman's problem, and use our generalization of the solution to explore how the underlying distribution of the quality of the options influences people's choices. A preregistered experiment with 2,520 participants shows that people adopt thresholds that decrease linearly with the proportion of trials remaining, consistent with the observation of linear thresholds in other optimal stopping problems. However, we show that people tend to explore more than predicted by linear thresholds, and that different distributions of quality result in thresholds with the same slope but different intercepts. These results indicate that people adapt linear thresholds used in optimal stopping tasks in a way that is sensitive to the underlying distribution—a simple strategy that we show is nearly as effective as Feynman's solution.

decision-making | optimal stopping | exploration-exploitation

In the late 1970s, the physicist Richard Feynman sat down for lunch with his friend Ralph Leighton at a Thai restaurant called Indra in Glendale, California. Leighton was trying to decide whether to order his running favorite (the ginger chicken), or try something new that had a chance of being even better. Feynman turned the dilemma into a math problem—and solved it. Unfortunately Feynman never published his analysis. All that remained from the conversation were his handwritten notes, which Leighton happened to keep (Fig. 1). The notes remained inscrutable for decades, until we managed to decipher them and reconstruct Feynman's original problem and solution.

Feynman focused on individual dishes at a given restaurant, whereas we imagine restaurants in a given city, although the formulation is identical. Each restaurant has a fixed value between 0 and 1 (alternatively, we can evaluate the percentile rank of each restaurant), but this value is not known until the first time it is sampled. The values are assumed to follow a Uniform distribution, with any value between 0 and 1 being equally likely. Of course this is a vast simplification of the full complexity of human dining behavior—where restaurant quality can change over time, and where our complex, dynamic preferences include factors like satiety, novelty, and social influence (1–6)—but its core question captures a recognizable and familiar human predicament: how should someone visiting the city for a certain number of nights decide which restaurant to visit each night in order to maximize the sum of these values?

Feynman's restaurant problem is an instance of what is known as an optimal stopping problem (7, 8). As such, it falls in the same category as the famous secretary problem (9), in which an interviewer seeks to maximize the probability of hiring the best candidate for a position but can only evaluate those candidates relative to one another. This problem can be translated to the dining setting by assuming the goal is to maximize the probability of selecting the best restaurant over a series of meals. However, Feynman's problem differs from the classic secretary problem in three ways: the distribution from which the restaurants are drawn is known, the diner is able to return to restaurants that they visited previously, and the goal is to maximize the total score across nights rather than the probability of identifying the single best option.

Feynman's restaurant problem is also closely related to the finite-horizon multi-armed bandit problem (10, 11), in which a decision-maker is presented with a set of options

Significance

Richard Feynman described a decision-making problem and its solution in handwritten notes, but the meaning of the notes has been a mystery for almost 50 y. We decipher the problem and solution from Feynman's notes, and prove that Feynman's solution is optimal. We generalize his result and find closed-form solutions for other distributions, and then turn to ask the question of how humans actually solve such decision-making problems. In a preregistered experiment with 2,520 participants, we find definitive evidence that humans use a decision threshold that decreases linearly with the proportion of trials remaining, achieving performance remarkably close to the optimal solution found by Feynman.

Author affiliations: [a]Department of Experimental Psychology, University of Oxford, Oxford OX1 3EL, United Kingdom; [b]Department of Psychology, Hunter College, City University of New York, New York, NY 10065; [c]Department of Computer Science, Princeton University, Princeton, NJ 08540-5233; and [d]Department of Psychology, Princeton University, Princeton, NJ 08540

Author contributions: B.C., E.M.R., and T.L.G. designed research; performed research; contributed new analytic tools; analyzed data; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.F. is a guest editor invited by the Editorial Board.

[1] To whom correspondence may be addressed. Email: tomg@princeton.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2509612123/-/DCSupplemental.

that differ in their payoffs (such as different arms of a gambling machine) and seeks to maximize the total payoff received from trying those options a fixed number of times. Again, this could be translated to the dining context, treating the restaurants as the different options. The key difference here is that in the multiarmed bandit problem the payoffs are usually stochastic, with a distribution around the true value, while in Feynman's problem the true value of a restaurant is directly observed. Like the multiarmed bandit problem, Feynman's restaurant problem creates a tension between exploring new options and exploiting knowledge acquired so far, but does so without dealing with uncertain observations.

Optimal stopping problems often arise in everyday life, appearing not just in choosing what to eat, but in finding a home, deciding who to marry, selecting a parking spot, and knowing when to quit a job (12). Extensive literatures have explored how people solve variants of the secretary problem (13–17) and the multiarmed bandit problem (18–25). Feynman's restaurant problem is a valuable addition to this canon: by removing uncertain observations, it makes it possible to study the explore–exploit tradeoff in a particularly pure manner, and the existence of closed-form expressions for the optimal policy facilitates its comparison to human behavior. In fact, previous behavioral experiments have used optimal stopping problems that are similar to (26–28) or variants on (29 and 30) Feynman's restaurant problem (for a detailed breakdown see *Discussion*). Here, we make use of the optimal solutions we derived for multiple variants of this problem, together with an innovative experimental design that allows us to get an unusually clear picture of people's behavior and to draw direct parallels between results in the psychological literature and the solution found by Feynman. As a consequence, we hope to not just re-solve the problem that Feynman first posed more than 40 y ago, but to resolve the question of how people perform such tasks.

Results

Optimal Policy. Our analysis of Feynman's notes revealed that he had assumed that the value of the options x has a Uniform distribution $p(x) = 1$ for $x \in [0, 1]$, so that any value between 0 and 1 is equally likely. He then argued that the optimal policy is to try different options each night until encountering an option with a value that exceeds the threshold t_n, where n is the number of nights remaining, with

$$t_n = \frac{\sqrt{n}}{\sqrt{n}+1}, \qquad [1]$$

and thereafter exclusively choose that option for the remaining nights. We reconstructed this problem from Feynman's notes and provide a full derivation and proof of the optimality of Feynman's solution in *SI Appendix*.

Our characterization of Feynman's problem and its optimal solution made it possible to express that solution in a generalized form. As a result, we are able to obtain analytical solutions to the restaurant problem beyond the Uniform case. To explore the nature of these solutions, we derived closed-form solutions for three additional distributions of value: Exponential, Power Law, and Triangular (shown in Fig. 2*A*). These distributions provide a nice contrast with the Uniform, having a rapidly decreasing exponential tail (Exponential), a slowly decreasing "heavy" tail (Power Law), and an unusual skew toward larger values (Triangular). As a consequence, they provide a useful set for exploring the form of the optimal thresholds and for evaluating human behavior on this task.

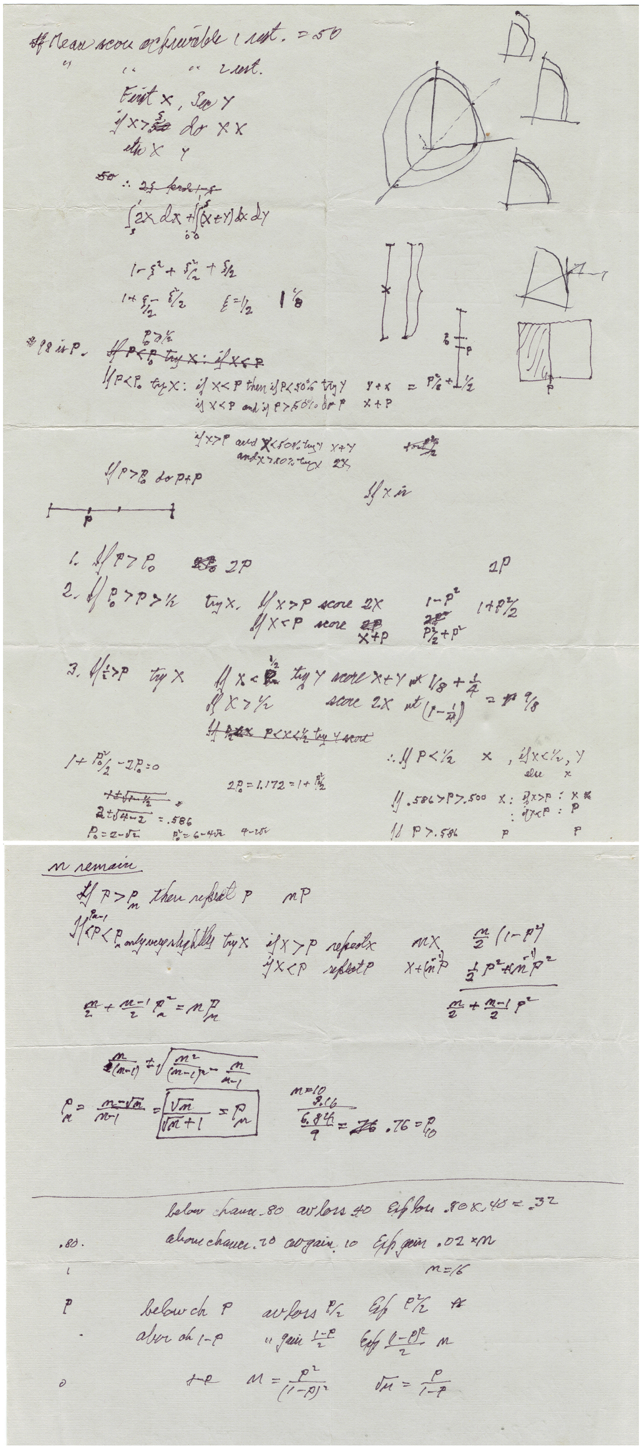

Fig. 1. Richard Feynman's handwritten notes detailing the restaurant problem. Courtesy of Ralph Leighton. Image credit: Richard P. Feynman (estate).

The Exponential distribution assumes $p(x) = e^{-x}$ for $x \in [0, \infty)$. We show that this results in the optimal threshold

$$t_n = 1 + W[\frac{n-1}{e}], \qquad [2]$$

where $W[\cdot]$ is the Lambert W function. The Power Law distribution we analyze assumes $p(x) = 2x^{-3}$ for $x \in [1, \infty)$.

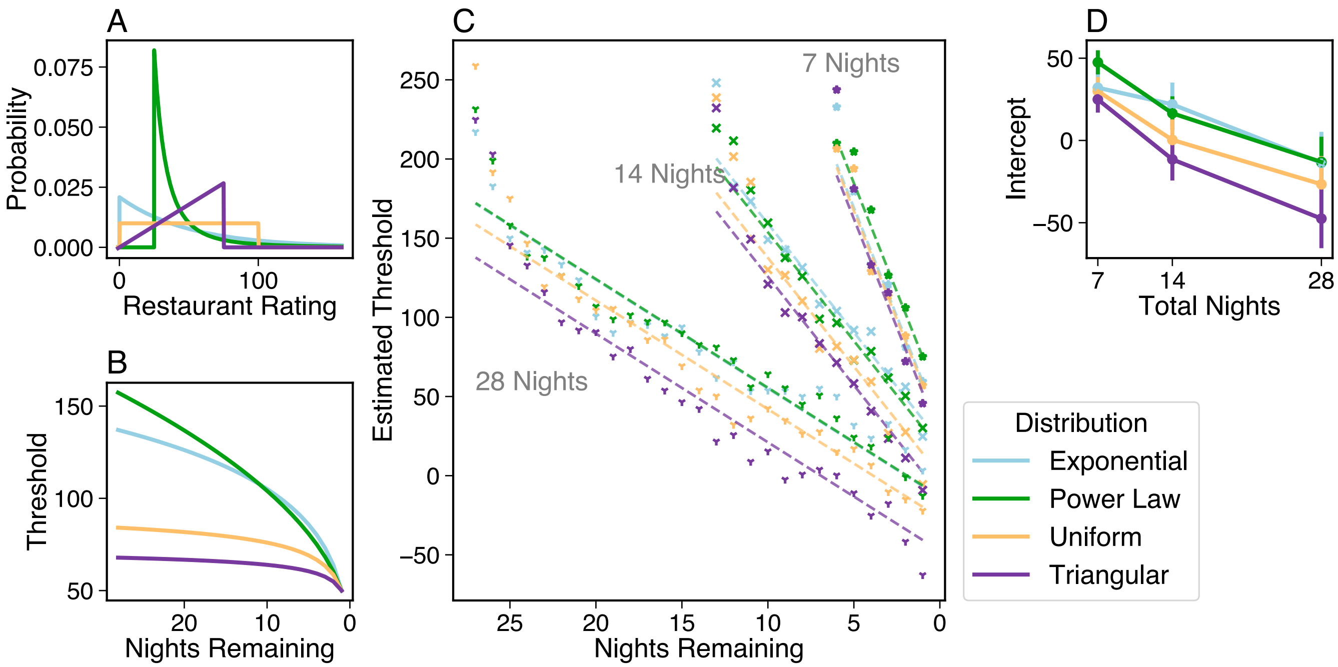

Fig. 2. Participants use thresholds that are linear but reflect the optimal policy. (*A*) Distributions of restaurant ratings. In our experiments, the Uniform distribution was taken over values from 0 to 100 and the values produced from the other distributions were multiplied by a constant to make sure that they all have a mean of 50. We plot the distributions on that scale for comparison with the experimental results. (*B*) Optimal thresholds for each distribution. The optimal threshold for the final night is always the mean of the distribution, so rescaling the distributions to have a mean of 50 results in convergence of those final thresholds to the same value and puts them on the same scale. (*C*) Estimated decision thresholds decrease linearly with proportion of nights remaining. Markers indicate maximum-likelihood estimates of the threshold $t_{n,T,D}$ from the logistic model (Eq. **5**), which estimates a separate threshold for each combination of nights remaining, total nights, and distribution. Lines show estimated thresholds from a linear model in which slope is constant in proportion of nights remaining across conditions but intercept varies with total nights and distribution. (*D*) Fitted intercepts from linear model. Intercepts vary with distribution in a manner similar to the optimal policy, with the greatest thresholds for Exponential and Power Law, then Uniform, then Triangular. Intercepts also decrease with total nights. Error bars show one bootstrap-estimated SE.

We show that this results in the optimal threshold

$$t_n = \sqrt{n} + 1 \ . \qquad [3]$$

The Triangular distribution assumes $p(x) = 2x$ for $x \in [0, 1]$. We show that this results in the optimal threshold

$$t_n = \begin{cases} \frac{2}{3}, & n = 1, \\ 2\sqrt{\frac{n}{n-1}} \cos\left(\frac{\pi}{3} + \frac{1}{3} \arcsin \frac{1}{\sqrt{n}}\right), & n > 1. \end{cases} \qquad [4]$$

The specific constants used in defining these distributions were selected to simplify the analytic expressions for their thresholds—we use rescaled values of these distributions in our experiment as shown in Fig. 2 and explained below. Derivations of the optimal thresholds for all distributions are provided in *SI Appendix*.

Across all four distributions the threshold t_n decreases over time (as the number of nights remaining n decreases), with the form of the function depending on the distribution (Fig. 2*B*). Indeed, the distribution has a substantial effect on the resulting threshold: the heavy tail of the Power Law distribution means that it starts out with higher thresholds than any of the other distributions, followed by the Exponential which also has the capacity to produce high values, albeit more rarely. By contrast, the Triangular distribution has lower maximum values than the Uniform and as a consequence has lower optimal thresholds. A decision-maker who is sensitive to the distribution of values should thus adapt their threshold to reflect these differences.

Identifying Human Policies. Recovering this problem and its solutions under multiple distributions led us to a natural question: what do people do when actually faced with the Feynman restaurant problem? In fact, psychologists have explored cases that are very close to Feynman's original formulation: Song et al. (29) considered a problem in which the options are drawn from a distribution that is Gaussian rather than Uniform, and Sang et al. (30) defined a problem where the options are drawn from a Uniform distribution over a discrete set of values (and mentioned a version of Feynman's problem as motivation). Song et al. relied on numerical solutions, while Sang et al. used backward induction to derive a solution in their discrete setting that we were able to show is equivalent to Feynman's in the continuous limit (see *SI Appendix* for proof). Evidence from these and other experiments on related optimal stopping problems (26–28) suggests that people do not follow the optimal strategy, and instead tend to use thresholds that decrease linearly. While use of linearly decreasing thresholds would seem to be a suboptimal heuristic for these tasks, it is possible that this heuristic is a close approximation of the optimal policy at minimal computational cost (31, 32). Using a large-scale experimental design and the mathematical results obtained above, we definitively show that people do indeed use linearly decreasing thresholds for versions of the Feynman problem defined with different distributions over options, and that these linear functions share an identical slope but vary their intercepts in a manner consistent with the optimal solutions. However, we also show that there are meaningful deviations from these linear thresholds early in decision-making, consistent with an early bias toward exploration.

In our experiment, we introduced participants ($N = 2{,}520$) to the problem of choosing which restaurant to eat at each night, explaining that restaurants differed in quality but the quality could only be observed by visiting the restaurant. On each night, participants would thus decide between trying a new restaurant whose value would be sampled from a specified distribution (“exploring”) or returning to the best restaurant they had previously encountered (“exploiting”). Restaurant values were sampled from one of the four distributions introduced above, which varied between participants. These distributions were multiplicatively rescaled so that their means were equal to the mean of a Uniform distribution over values from 0 to 100 (i.e., 50) prior to generating the values seen by the participants. This choice was motivated by the fact that the optimal threshold on the final night is simply the mean of the distribution, so using identical means results in the same threshold for all four distributions on the final night and puts them on the same scale (Fig. 2*B*). Before beginning the main experiment, where they would have the chance to visit restaurants of their choice for some number of nights, participants observed 84 values sampled from the distribution that they would encounter in the town they were visiting. The forms of the distributions are quite different (Fig. 2*A*), and a control experiment determined that people could distinguish between these four distributions based on observing 84 sample values (*SI Appendix*), meaning that people had enough information to be able to modify their strategy to match the underlying distribution. The total number of nights was also varied across participants (7, 14, or 28). Finally, in order to have enough relevant decisions to estimate participants’ thresholds late in a trial, participants were assigned to a hidden clamping condition that limited the maximum value of new restaurants until a fixed proportion of the total nights had passed (*Materials and Methods*). Each participant only made one sequence of decisions, so they did not have an opportunity to adjust their strategy based on its success or failure.

After scaling to match the thresholds on the final night, the magnitude of the optimal thresholds for the different distributions across the remaining nights varies in the following order: (Exponential, Power Law) > Uniform > Triangular, where Exponential and Power Law are not consistently ordered (Power Law is higher earlier, Exponential is higher later). To determine whether the heuristic strategies employed by participants show characteristics of the optimal policy, we examined whether participants used thresholds that vary in a similar manner across distributions.

Following a preregistered analysis plan, we examined the thresholds used by participants by fitting a logistic model that assumed participants decide by comparing the best reward seen thus far b to a threshold $t_{n,T,D}$ that depends on the number of nights remaining n, total nights T, and distribution D:

$$P(a = \text{exploit}) = \frac{1}{1 + e^{-\beta(b - t_{n,T,D})}}, \qquad [5]$$

where β is the gain of the logistic and indicates how deterministic people’s choices are around the threshold.[*] Here $t_{n,T,D}$ is an imputed logistic threshold (the model’s 50% indifference point), which need not lie within the support of the rating distribution. Values above the maximum possible rating simply indicate that participants continued to explore even at the top of the scale, and negative values indicate that participants are broadly likely to exploit regardless of how low the best score seen by that point.

In line with the optimal policy, the maximum-likelihood estimates of the thresholds $t_{n,T,D}$ decayed with nights remaining (Fig. 2*C*, markers). However, unlike the optimal policy, and in line with prior reports (26, 29), they decayed linearly and also exhibited a dependence on total-nights condition. Finally, thresholds also varied with distribution in a manner consistent with the optimal policy: within each total-nights condition, intercepts were offset in the same order (Exponential, Power Law > Uniform > Triangular) as the optimal thresholds (Fig. 2*D*).

Effects of Distribution on Linear Thresholds. Our preregistered prediction that thresholds would differ depending on distribution was confirmed by a factorial model comparison (33) with three factors; furthermore this model comparison allowed us to disambiguate the nature of the linear relationship. All models specified that thresholds vary according to a linear function, $t_{n,T,D} = a + mx$. The first factor specified what task variable the linear function was defined over: nights remaining ($x = n$), proportion of nights remaining ($x = n/T$), or nights elapsed ($x = T - n$). The second factor concerned whether linear parameters (a and m) varied or were fixed over different conditions of total nights T and distribution D. The third factor specified which linear parameters should vary across these task conditions (a, m, both, or neither). We found that for the best-performing model, thresholds decreased linearly with proportion of nights remaining (Fig. 2*C*, lines; see *SI Appendix*, Fig. S2 for statistical results), using intercepts that varied freely but a single slope parameter shared across all twelve combinations of total-nights and distribution conditions. This model fits the inferred thresholds remarkably well, although we note it does not capture a tendency to explore in the first few nights (we investigate this in more detail below).

The use of a single slope across all twelve total nights × distribution conditions suggests a striking parsimony, and the fitted intercepts revealed a correspondence with the optimal policy. The predicted ordering of thresholds by distribution was well matched: intercepts for Exponential and Power Law distributions were greater than intercepts for the Uniform distribution, which were greater than intercepts for the Triangular distribution (Fig. 2*D*). We assessed the significance of these differences by fitting an additional null model for each pair of distributions requiring their two intercepts to be the same (Table 1); for all distributions, these models produced worse fits, and the significance of an unconstrained model was

Table 1. Performance of constrained models, for which pairs of distributions (left column) share intercept parameters, relative to unconstrained model with separate intercept parameters for each distribution

Model	ΔBIC	ΔAIC	$\chi^2(3)$	P
Exponential = Power Law	−11.1	1.7	9.5	0.023
Exponential = Uniform	11.3	24.1	54.3	<0.001
Exponential = Triangular	85.3	98.1	202.2	<0.001
Power Law = Uniform	13.8	26.6	59.2	<0.001
Power Law = Triangular	113.8	126.7	259.3	<0.001
Triangular = Uniform	26.4	39.2	84.3	<0.001

ΔBIC and ΔAIC denote difference in performance between constrained models and the unconstrained model (positive values favor the unconstrained model). The final two columns provide statistics for likelihood-ratio tests measuring whether the additional parameters of the unconstrained model are justified.

---

[*] We also confirmed that this linear threshold model outperforms other simple heuristics suggested in ref. 30 as well as reinforcement learning algorithms that have previously been used to explain human learning in multiarmed bandit tasks when applied to our problem (*SI Appendix*).

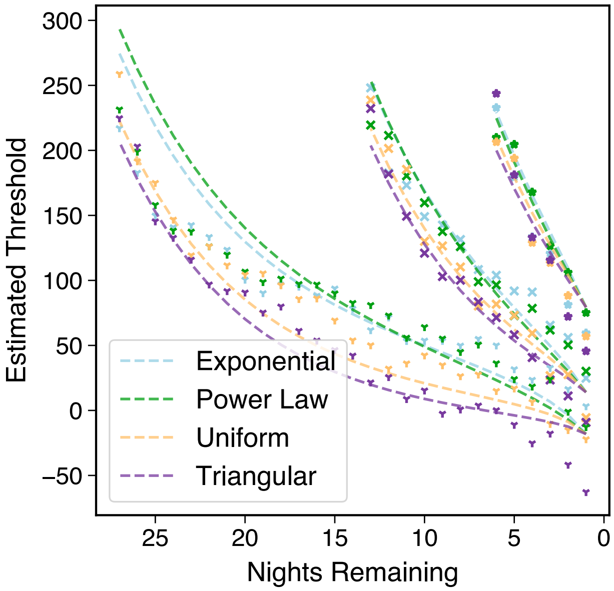

Fig. 3. Participants' decisions reflect an early exploration bonus. Markers indicate maximum-likelihood estimates of the threshold $t_{n,T,D}$ from the logistic model (Eq. **5**), which estimates a separate threshold for each combination of nights remaining, total nights, and distribution. Lines show predicted thresholds from the best-fitting linear model in which slope is constant in proportion of nights remaining across conditions but intercept varies with total nights and distribution, augmented with an exponentially decaying bonus.

supported by likelihood-ratio tests. This supports the conclusion that participants' thresholds were sensitive to differences in distribution in a manner qualitatively prescribed by the optimal policy. Finally, in addition to the effect of distribution, we found that intercepts decreased with increasing total nights. Hierarchical models that allow for between-participant variation support the same conclusions (*SI Appendix*).

A Bias Toward Early Exploration. Although the best-fitting linear model provides a generally good account of how participants' thresholds change over task conditions, Fig. 2*C* demonstrates that the linear model underestimates thresholds in the first few nights of each total-nights condition. We found that this overexploration for the first few decisions could be captured by a nonstationary bonus to the linear threshold that decays exponentially with nights elapsed:

$$\mathrm{bonus}(T-n;\delta,h) \;=\; \delta\, 2^{-(T-n)/h}\,, \qquad [6]$$

where δ is the first-night threshold bump and $h > 0$ is a half-life measured in nights elapsed, $T-n$. The effective threshold entering the logistic rule is

$$t_{n,T,D} \;=\; t_{\mathrm{linear}}(n,T,D) + \mathrm{bonus}(T-n;\delta,h)\,, \qquad [7]$$

where t_{linear} are thresholds from the linear model. We found that adding an early-exploration bonus to participants' thresholds improved the fit to participants' exploration decisions ($\Delta\mathrm{BIC} = 79.7$ for the best-fitting linear model) and resulted in a close correspondence to participants' higher thresholds in the first few nights (Fig. 3). Furthermore, the inclusion of early explore bonuses did not alter the ordering of intercepts under that model, nor did it alter which linear model provided the best fit (*SI Appendix*, Fig. S3).

Linear Thresholds Achieve Near-Optimal Performance. Although participants' strategies deviated from the optimal policy overall, they were quite well suited to the task. By computing the expected value of the policies that they implemented and performing numerical optimization over linear policies, we found that participants achieved scores comparable to those of the optimal (unconstrained) policies, using linear thresholds that were close to the optimal linear policies (Fig. 4). People thus seem to follow a simple strategy that can be easily modified to accommodate differences in both total nights and distribution, allowing them to come close to optimal performance while minimizing cognitive effort (32).

Discussion

Our work revisits a problem originally posed by Richard Feynman, providing both a mathematical foundation and an empirical investigation into how humans approach an optimal stopping task under uncertainty. We found that Feynman's original analysis—largely lost to history—yields an elegant closed-form policy for maximizing reward through sequential exploration and exploitation. This problem, though formulated in terms of choosing restaurants, captures a broader class of real-world decisions where individuals must balance the potential value of new options against the known value of past experiences.

By building on Feynman's result, we were able to derive closed-form solutions for the optimal policy when restaurant quality varies according to four different distributions. This allowed us to evaluate how human behavior on this optimal stopping task compares to the optimal policy as the underlying distribution is varied. Our behavioral results contribute to a growing body of evidence suggesting that people do not rely on fully optimal strategies in such environments but instead employ simple heuristics that approximate optimal performance. Specifically, participants used linear threshold rules, where the decision threshold declined linearly with the proportion of trials remaining. This structure provides computational efficiency and adaptability across conditions. Remarkably, even though these linear strategies deviate from the optimal policy, they perform nearly as well, especially when noise is removed. The result offers a compelling case study in resource-rational decision-making, where people deploy simple yet effective strategies under cognitive constraints.

We also found that people modulate their thresholds depending on the distribution of outcomes—a behavior that mirrors changes in the true optimal policy. These findings suggest that human cognition reflects a nuanced sensitivity to structural differences in the environment, even in one-shot decisions. By using a theoretically well-specified task, grounded in closed-form solutions, and by employing an experimental design that minimizes learning and memory confounds, we were able to isolate the structure of human decision policies in a clean and interpretable way.

Our results also identified a meaningful deviation from the linear threshold strategy, consistent with a bias toward exploration early in the task. Understanding the source of this bias, and how people adapt their decision strategies, are important questions for future work. In the remainder of the paper we consider how our results relate to previous findings and highlight some of the limitations of the current work and the other future directions that they suggest.

Previous Analyses of Optimal Stopping. Previous studies have examined human behavior in related optimal stopping problems,

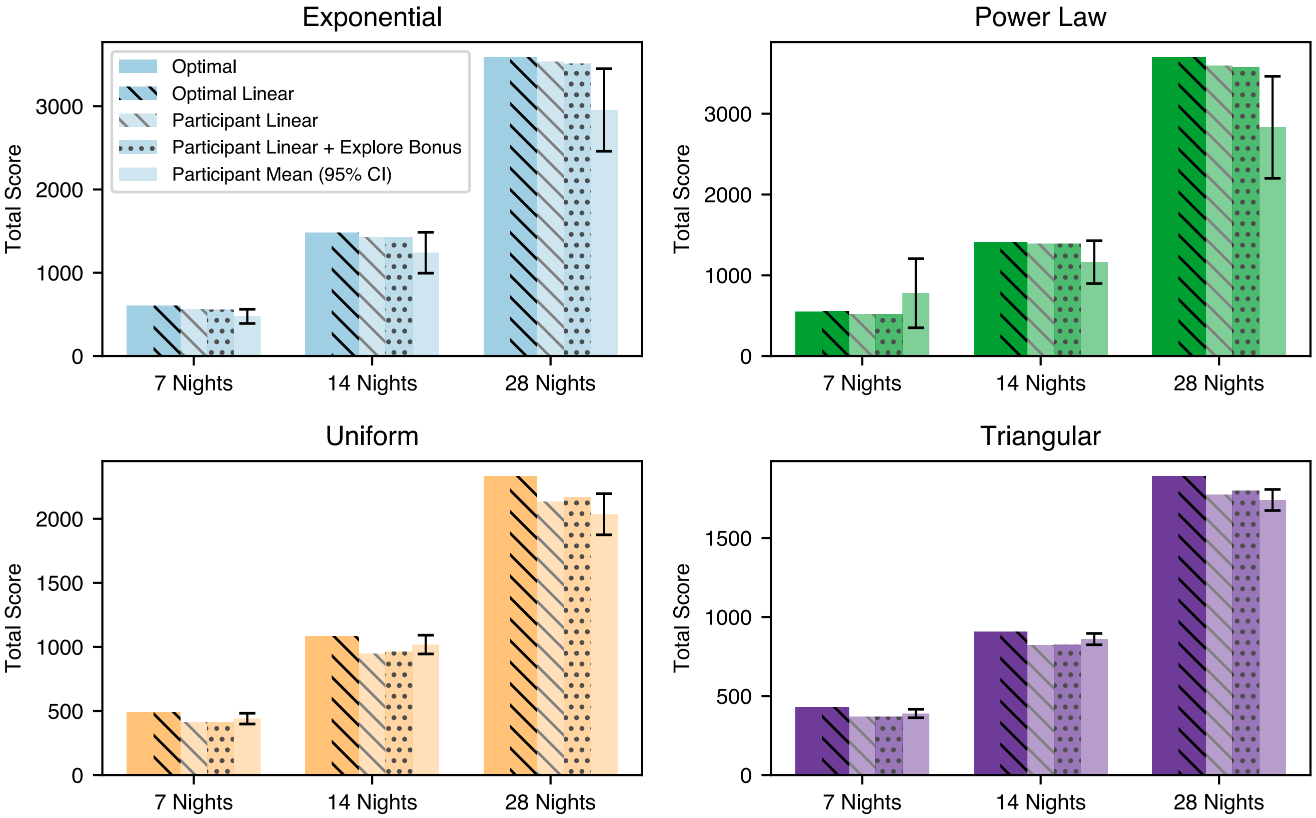

Fig. 4. Linear thresholds achieve near-optimal performance. Across all four distributions, optimal linear models achieve results extraordinarily close to the optimal solutions. However, the expected value from linear models fitted to participant data (with and without an exploration bonus), as well as the mean scores actually achieved by our participants, are nearly as high as those from the optimal (and optimal linear) strategies.

some of which have been instances of the restaurant problem, and attempted to model participants' strategies. Here we briefly review these studies and highlight the novel components of the approach we took in this paper. Compared to all the prior studies that we review, three features of our current study are completely unique. First, we provide a general solution for this class of problems and obtain analytic results for multiple distributions. Second, each participant in our study performed the task only once, allowing us to examine individual behavior without the influence of in-task learning. In contrast, participants in previous studies discussed performed each task between 20 and 200 times, thereby enabling substantial learning. Third, our study also measured how thresholds change with both total nights and distribution conditions. While some previous studies have examined effects of one of these variables in isolation, only by varying both of them could we identify that a) participants utilized thresholds defined with a decreasing linear function, b) this function was defined over the proportion of nights remaining, c) the slope of this function is constant across total night and distribution conditions, and d) the intercepts vary with distribution in a manner consistent with the optimal thresholds. We discuss specific differences with each study in turn. The set of optimal stopping problems considered in these studies are summarized in *SI Appendix*, Table S1.

Bhatia et al. (27), Wilson et al. (28), and Baumann et al. (26), studied how humans behave in optimal stopping tasks that are similar, but not identical, to Feynman's restaurant problem. Bhatia et al. (27) explored optimal stopping problems where, rather than having a fixed number of decisions (specified by total nights), participants could explore for as long as they wanted, but paid a fixed cost for each new option explored. In contrast to the restaurant problem, because there is no explicit time horizon, but rather a fixed per-decision cost, the optimal solution is to use a fixed threshold invariant to the number of decisions remaining. Results indicated that differences between participants' behavior and an optimal model could be accounted for by risk aversion, additional subjective search costs, and also an assumption of Gaussian errors distributed around the value of an option.

Wilson et al. (28) studied a related task in which, rather than having an explicit time horizon, the game had a fixed probability of ending after each decision. This formulation also yields an optimal policy with a fixed threshold. Participants' behavior was well described by a sampling framework: taking more samples led to discovering the optimal fixed threshold and low-noise decisions, whereas taking fewer samples produced lower thresholds and noisier decisions.

Baumann et al. (26) asked participants to accept or reject a series of ten airline tickets whose prices were normally distributed, with their reward corresponding to the cost (or savings) of that ticket. Unlike Feynman's restaurant problem, accepting a ticket ended the game, thus requiring participants to maximize the value of the single chosen ticket as opposed to maximizing a sum over multiple choices. Furthermore, participants could not return to prior offers they had encountered. Despite these differences, similar to Feynman's restaurant problem, the optimal solution was to utilize a nonlinear decreasing acceptance threshold. Similar to our findings, participants were best modeled as using a linearly decreasing acceptance threshold. Additionally, similar to our findings, participants' linear parameters were sensitive to the distribution from which tickets were drawn, consistent with optimal behavior. However, since the number of tickets in the sequence was not varied, it could not be determined which variable (e.g., nights remaining vs. proportion of nights remaining vs. nights elapsed) the linear function was taken over. In addition to using Feynman's restaurant problem, our findings add the result that thresholds are defined as a linear function over proportion of nights remaining with constant slope, but differing intercepts over total nights and distribution.

As discussed in *Identifying Human Policies*, both Song et al. (29) and Sang et al. (30) used tasks that are variants on Feynman's restaurant problem. However, in contrast to the study presented here, Sang et al. only tested a single setting of total nights, and both studies only tested a single distribution (Gaussian and Discrete Uniform, respectively) and thus could not examine effects of distribution on thresholds. Both studies found that participants used thresholds that decayed linearly with total nights, and Sang et al. also showed that the observed linear policy was close to the optimal policy they computed for their discrete setting. By varying total nights, Song et al. identified a dependence on total nights consistent with the linear function being taken over the proportion of nights remaining. Our results build on these findings by i) providing closed-form solutions for several versions of the problem, ii) using only a single iteration of the task for each participant and hence removing potential confounds related to learning, and iii) determining that while the slope of people's threshold is constant across total nights, the intercept varies with both total nights and the distribution of values.

The use of a single iteration of the task might help to explain the early bias toward exploration that we observed in our experiment. Because participants only completed the task once, they may have been more reluctant to commit to the first restaurant they encountered. The thresholds reported by Song et al. and Sang et al. did not show this early exploration bias, which might reflect an increasing comfort with early commitment that grows with experience. We would also hypothesize that more opportunities to adapt the learned strategy to the task might result in greater differentiation between the thresholds used for different distributions. Our focus here was on the immediate intuitions that people had about how to perform the task rather than their ability to adapt their strategies, but understanding this adaptation process is an important direction for future research as we discuss in more detail below.

Limitations and Future Directions. While our findings offer insights into both the optimal policies for this class of optimal stopping problem and human behavior, several limitations warrant discussion.

First, the task setting—while elegant—abstracts away from many complexities of real-world decision-making. For instance, restaurant values were fixed and deterministic, and participants faced no real-world costs (e.g., travel time, money, or social factors). Future work could explore how the introduction of noise, costs, or time-varying rewards impacts the structure of decision policies and their alignment with optimal solutions. This would introduce elements that have been explored in other optimal stopping problems: adding noise returns us to the domain of the multiarmed bandit, and explicit costs for searching have been investigated in the setting of the secretary problem (16). Exploring the impact of these dimensions has the potential to more clearly define the space of optimal stopping problems and how variation along each dimension influences the kinds of strategies that people adopt.

Second, as discussed above, the linear threshold heuristic was inferred from a single sequence of decisions per participant. While this avoids contamination from learning, it also means we cannot assess how strategies evolve over time. Relatedly, our derivations assumed full knowledge of the value distributions, but people had to estimate these distributions from only a limited sample. Real-world decisions often involve uncertainty about the environment itself. A natural extension would be to examine meta-learning: how people simultaneously learn about and adapt to unknown distributions over time. This requires being explicit about the set of distributions under consideration. Similar analyses have been performed for other optimal stopping problems, such as the secretary problem (34). Our experiment was deliberately constructed to identify the strategy that people adopt without experience with this task, but both Song et al. (29) and Sang et al. (30) used designs in which participants had the opportunity to perform the task multiple times and observed improvements in performance over time. Such a design would make it possible to study the learning process and compare it to formal accounts based on meta-learning.

Third, we modeled participant behavior using deterministic thresholds and logistic noise. However, people may also exhibit dynamic variability in decision criteria across trials (e.g., due to mood, attention, or prior experience). Richer modeling approaches (e.g., hierarchical Bayesian models) could uncover whether such variability contributes systematically to deviations from optimality. This kind of approach has been pursued productively in optimal stopping tasks that have been studied more extensively, such as the multiarmed bandit task (8).

Finally, our analysis of the correspondence between the linear thresholds and the optimal solutions raises an interesting question: to what extent does the linear strategy that people adopt generalize beyond these distributions, and does it remain an effective approximation to the optimal strategy for other distributions? Knowing the set of distributions that an agent might face in the world would provide an opportunity to construct a strategy that is robust to those distributions, suggesting that it might be fruitful to explore the natural statistics of optimal stopping problems in human environments [as has previously been done for memory (35) and other decision problems (31)].

Conclusion. Richard Feynman notoriously had a good instinct for elegant solutions to problems that provided insights into the physical world. His analysis of how one should choose a meal suggests the same instinct guided him toward problems that offer insight into human behavior. The Feynman restaurant problem has the rare attributes of being analytically tractable while capturing a fundamental tension in human decision-making—the trade-off between exploration and exploitation. Its simplicity also means that we can definitively identify the strategy people use to solve this problem: adopting thresholds that decrease linearly in the proportion of opportunities that remain, but adjusting these thresholds up and down for different distributions. More than four decades after Feynman sketched some notes in a Thai restaurant, we have resolved the mathematical puzzle he left behind, and revealed how people naturally approach such sequential decision problems. We hope that these results help to put Feynman's restaurant problem in its appropriate place alongside the secretary problem and the multiarmed bandit as fundamental paradigms for how people approach problems that involve the challenge of knowing when to stop, providing the foundation for creating a more complete characterization of the space of these problems and the strategies that people use to solve them.

Materials and Methods

Experimental Procedure. The experimental procedures and planned analysis were preregistered at https://osf.io/9e5g7. Participants were 2,520 individuals in the United States, recruited via the Prolific online platform. Ages ranged from 18 to 94 y ($M = 40.4$, $SD = 13.6$) with 53.2% identifying as male, 46.6% as female, and 0.2% preferring not to say. Participants were eligible if they had completed between 100 and 10,000 previous studies on Prolific with a $\geq$95% approval rate, and they needed to pass a short quiz to demonstrate they understood the task in order to proceed to the task itself. The study was approved by Princeton IRB, and all participants provided informed consent. Participants were randomly assigned to one of four rating distributions (Triangular, Uniform, Exponential, and Power Law), one of three numbers of nights (7, 14, and 28), and one of seven different "clamping" conditions, described below. Each condition received a total of 30 participants, resulting in a total of $4 \times 3 \times 7 \times 30 = 2{,}520$ participants. (In the process of randomizing participant assignments, some conditions received more than 30 participants, resulting in a total of 2,530 submissions; any participant data beyond the first 30 in each condition were ignored.)

The experimental procedure was as follows: after providing consent, participants were asked to imagine that they are going to be living for some period of time in a city where restaurants vary in quality, that their quality scores are distributed randomly with mean of 50, and that they are initially unknown but do not change once discovered. Participants were instructed that they will earn a bonus proportional to the total score achieved. They were then presented with a tutorial, showing a 4-by-7 grid of restaurants with unknown ratings that are revealed when "visited" (clicked). Following successful completion of the tutorial, they took a short quiz to ensure they understood the nature of the task. The restaurant values in the tutorial were all drawn from the distribution that the experimental trial itself would use. Participants then saw an additional three screens of fully revealed 4-by-7 grids of random restaurant values (84 values in total) to help them learn the distribution. Following this, the experiment began, with the participant attempting to maximize total score within a new, unexplored 4-by-7 grid of restaurants over a fixed total number of nights. *SI Appendix*, Fig. S1 depicts the user interface of the experiment, showing the grid of restaurants, some of which have been visited and have their scores visible.

In the task, each time a participant selected a new restaurant, this choice was categorized as Exploring. When participants selected a previously explored restaurant that was the best restaurant they had thus-far experienced, this choice was categorized as Exploiting. Occasionally, but rarely, participants returned to a restaurant they had previously encountered whose value was below the best encountered previously. Because this sort of choice cannot be clearly categorized as either an "Explore" or "Exploit," it is labeled as a "Mistake," and such choices are excluded from our main analysis (Fig. 2). However, because such mistakes may reflect either careless behavior or a failure in understanding the task, one could instead remove all participants who make at least one mistake; we show in *SI Appendix* that our results are robust to stricter exclusion criteria.

One of the challenges with collecting human data for Feynman's restaurant problem is that it is very unlikely that participants will need to explore late in a trial. For instance, the optimal exploration threshold for the final night is 0.5, and so the chance of a participant in the 7-night condition not already having found a restaurant with a value greater than 50 by their last night is $(0.5)^6$, or one in every 64 trials. The chance of a participant in the 28-night condition not having already found a restaurant with a value greater than 50 by their last night is $(0.5)^{27}$, or less than once per hundred million trials. In order to help gather more useful data about human exploration thresholds, particularly later into trials, we "clamped" the random values of new restaurants to the range $[0, t_n)$, where t_n is the optimal exploration threshold with n nights remaining. (Thus, it is never optimal to immediately exploit such a restaurant; however, because t_n decreases as the participant proceeds through the trial, it may become optimal to exploit that restaurant later.) There were seven different conditions used, which would apply this clamp to the values of new restaurants for the first 0 to 6 sevenths of their trial. For instance, a participant in clamp condition 0 would not experience any clamping behavior at all. A participant in clamp condition 1, in the 14-nights condition, would have the values of new restaurants clamped for the first 1/7th (2 nights) of their trial, and unclamped afterward. A participant in clamp condition 6, in the 28-nights condition, would have the values of new restaurants clamped until their final four nights. This design allowed us to collect richer and more instructive participant data, using fewer participants, relative to leaving the values totally unclamped for all participants. (Due to a numerical error in calculating thresholds for the Exponential condition, a slightly higher value was used for clamping, but this only reduced the efficiency of data collection in that condition.)

Behavioral Analysis. We compared the ability of a variety of models to account for participants' decisions in the task. All models assume that participants decide by comparing the best seen reward thus far b to a decision threshold $t_{n,T,D}$, with the probability of exploiting the current best option given by the logistic function defined in Eq. **5**.

For all models, the gain of the logistic β is a free parameter that modulates how noisily the model followed this threshold rule. Models differ with respect to how that threshold changes across nights remaining n, total nights T, and distribution D.

For the optimal model, thresholds are provided by the optimal threshold, $t_n = t^*_{n,D}$ (*SI Appendix*, Fig. S2*B*), which is dependent on n and D, but not on T. We additionally tested a model where a free-parameter intercept a could be added to these optimal thresholds, $t_n = t^*_{n,D} + a$ (optimal w/ intercept).

Following previous work, we additionally compared the ability of a variety of linear models to describe the data. All linear models specify that thresholds be formed as a linear function, $t_n = a + mx$. We tested all combinations of linear models, varying across three factors. The first factor specified the input to the linear function. Input was either nights remaining, $x = n$, proportion of nights remaining, $x = n/T$ or nights elapsed, $x = T - n$. The second factor varied which task conditions the linear parameters should vary over: total nights, distribution, both, or neither. The final factor varied which linear parameters should vary across task conditions (intercept a, slope m, "Both," or "Neither").

To provide a concrete example, for a model where the first factor is set to nights remaining, the second factor specifies that variation occurs across total nights, and the third factor specifies that slope is the linear parameter that varies, thresholds are defined as $t_{n,D,T} = a + m_T n$, where a is a free parameter and m_T is three free parameters, one for each total nights condition.

Combinatorially varying these three factors in principle defines a space of $3 \times 4 \times 4 = 48$ possible combinations. However, if either the second factor or third factor is set to Neither, the other must also be set to this, reducing the number of possible models to 30. Finally, a subset of these models are algebraically equivalent under transformation of the slope parameters m. Specifically, if the slope parameter varies across total nights, then models for which the second two factors are the same, but the first factor varies, are all the same (and the fitted parameters are transformations of one another). For display purposes, in *SI Appendix*, Fig. S2, we plot model comparison statistics for these models separately, however mark them with a pattern across bars to denote that they are the same across varying the function input. Accounting for these algebraically identical models sets the number of independent linear models to 22. The simplest of these models (where both the second factors are set to neither) have three parameters: β, a, and m. The most complicated model, where the second and third factors are each set to Both, has 19 parameters: β and $a_{T,D}$ and $m_{T,D}$ for each combination of total nights T and distribution D. For purposes of visualizing thresholds used by participants in a model-agnostic manner, we also fit a maximally flexible model for which $t_{n,T,D}$ itself consisted of a separate free parameter for each combination of nights remaining, distribution condition, and total nights condition (Fig. 2*C*, markers). This model was only used to visualize thresholds and thus was not included in our model comparison.

For each model, we estimated the free parameters by maximizing the likelihood of all participants' choices using custom code written in the Julia programming language (version 1.9.1). For each model, Julia functions specifying the negative log likelihood were minimized by using the package Optim (version 1.31.3), and additionally providing this method with a Jacobian function computed using Julia package ForwardDiff (version 0.10.36). To avoid identifying local minima, for each model, `optimize` was called from 30 different randomly sampled start points. Models were compared by computing both Bayesian information criterion (BIC) and Akaike information criterion (AIC; *SI Appendix*, Fig. S2).

SI Appendix, Fig. S2 shows BIC (*A*–*C*) and AIC (*D*–*F*) scores for all models compared. All linear models explained the data better than the optimal model. For BIC scores, the best performing model was a linear function where the input corresponded to the proportion of nights remaining, the slope parameter was constant, and the intercept varied across both total-nights and distribution conditions. This model outperformed a model where the slope also varied across total-nights and distribution conditions ($\Delta\text{BIC} = 26.2$). In examining AIC scores, these models change places ($\Delta\text{AIC} = 20.7$). Both models greatly outperformed the next-best model where intercepts varied across total-nights but not distribution conditions ($\Delta\text{BIC} = 113.7$, $\Delta\text{AIC} = 152.2$ compared to the best model), thus supporting that participants' thresholds are sensitive to distribution condition.

SI Appendix, Fig. S2 shows maximum likelihood intercept values from the best-fitting model. In addition to demonstrating a correspondence with the optimal policy's ordering across distributions, we found that intercepts decreased with increasing total-night conditions.

To further test that participants' thresholds are sensitive to the distribution condition, we additionally performed a likelihood ratio test which verified that a model where intercepts vary according to both distribution and total nights outperformed a model where they vary according to total nights alone ($\chi^2(9) = 322.3, P < 0.001$). To examine the effects of distribution on intercept parameters in a more fine-grained manner, we fit a null model for each pair of distributions, which constrained each of those distributions' three intercept parameters to be the same (Table 1). Comparing the fit of each of these constrained models to the unconstrained model, allowing each distribution's intercept parameters to vary, allowed for a quantification of evidence for whether the intercept parameters between distributions should be different. Comparison of AIC between models as well as performance of the likelihood ratio tests provided support for different intercept parameters between each pair of distributions. Comparison of BIC between models did not support a difference between intercept parameters of Exponential and Power Law distributions. These results are qualitatively in line with effects of distribution on predictions of the optimal policy, which specifies differences in thresholds between all distributions, however with the direction of these differences not being fixed for Exponential and Power Law.

Human Performance Compared to Optimal. Having characterized our participants as using thresholds that descend in a linear fashion as a proportion of nights remaining, we can ask the question of how the specific linear thresholds we model our participants as using compare against the optimal linear-in-proportion thresholds.

We can use dynamic programming to calculate the expected value of a given set of thresholds, given the distribution type and the total number of nights. *SI Appendix*, Fig. S8 shows the results of using dynamic programming over various combinations of slope and intercept.

Overall, we might ask how the optimal linear thresholds compare to the optimal (unconstrained) thresholds in terms of their expected value across the task. As we can see from Fig. 4, the linear constraint comes with an extremely small penalty, and the optimal linear policies are very close to the absolute optimal policies in expectation.

Furthermore, we can ask how our participants did relative to optimal performance in each scenario. Here we select only the $N = 30$ participants in each condition whose data was not clamped at all, as other participants received slightly lower scores in expectation. We see from Fig. 4 that participants were generally quite competitive with the optimal. (Owing to the heavy-tailed nature of the Exponential and Power Law distributions, there was even a condition in which this particular group of participants managed to outperform the optimal strategy, on account of one particularly lucky participant.) Participants' lower scores compared to optimal is due not only to their threshold strategy, but also due to noise in their decision policy, which may be out of their control (36). To determine how effective participants' strategy was, independent of noise, we also computed the expected value of the linear model used by participants (at their best fit parameters), however implemented with no noise. This model performs comparably to optimal performance (Fig. 4).

Data, Materials, and Software Availability. Experimental data, model-fitting code, and a reproducibility notebook that regenerates all key statistics and figures from this paper are available at ref. 37.

ACKNOWLEDGMENTS. We thank Michael A. Gottlieb for sharing Feynman's original notes with us and for providing comments on our proof. His original formulation of the Feynman restaurant problem, a facsimile and partial transcription of the notes, and his response to our results appear on his website: https://www.feynmanlectures.caltech.edu/info/other/Feynmans_Restaurant_Problem_Revealed.html. We thank Ralph Leighton for sharing his recollections with us, Christopher Summerfield, Max Shron, and Rob Goldstone for helpful discussions, and our reviewers for providing clear and insightful comments that improved this paper. This work was made possible with the support of the NOMIS Foundation and the Clarendon Fund.

1. P. Rozin, "The selection of foods by rats, humans, and other animals" in *Advances in the Study of Behavior*, J. S. Rosenblatt, R. A. Hinde, E. Shaw, C. Beer, Eds. (Elsevier, 1976), vol. 6, pp. 21–76.
2. P. Rozin, P. M. Todd, "The evolutionary psychology of food intake and choice" in *The Handbook of Evolutionary Psychology*, D. M. Buss, Ed. (Wiley Online Library, 2005), pp. 183–205.
3. B. Scheibehenne, L. Miesler, P. M. Todd, Fast and frugal food choices: Uncovering individual decision heuristics. *Appetite* **49**, 578–589 (2007).
4. B. J. Rolls, E. T. Rolls, E. A. Rowe, K. Sweeney, Sensory specific satiety in man. *Physiol. Behav.* **27**, 137–142 (1981).
5. R. K. Ratner, B. E. Kahn, D. Kahneman, Choosing less-preferred experiences for the sake of variety. *J. Consum. Res.* **26**, 1–15 (1999).
6. B. McFerran, D. W. Dahl, G. J. Fitzsimons, A. C. Morales, I'll have what she's having: Effects of social influence and body type on the food choices of others. *J. Consum. Res.* **36**, 915–929 (2010).
7. T. S. Ferguson, Optimal stopping and applications (2008). https://www.math.ucla.edu/~tom/Stopping/Contents.html. Accessed 31 May 2024.
8. M. D. Lee, A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cogn. Sci.* **30**, 1–26 (2006).
9. M. Gardner, Mathematical games. *Sci. Am.* **202**, 152 (1960).
10. H. Robbins, Some aspects of the sequential design of experiments. *Bull. Am. Math. Soc.* **58**, 527–535 (1952).
11. D. A. Berry, B. Fristed, *Bandit Problems: Sequential Allocation of Experiments* (Chapman and Hall, New York, NY, 1985).
12. B. Christian, T. Griffiths, *Algorithms to Live By: The Computer Science of Human Decisions* (Henry Holt, 2016).
13. J. P. Kahan, A. Rapoport, L. V. Jones, Decision making in a sequential search task. *Percept. Psychophys.* **2**, 374–376 (1967).
14. A. Rapoport, A. Tversky, Choice behavior in an optional stopping task. *Organ. Behav. Hum. Perform.* **5**, 105–120 (1970).
15. C. A. Kogut, Consumer search behavior and sunk costs. *J. Econ. Behav. Org.* **14**, 381–392 (1990).
16. D. A. Seale, A. Rapoport, Sequential decision making with relative ranks: An experimental investigation of the "secretary problem". *Organ. Behav. Hum. Decis. Process.* **69**, 221–236 (1997).
17. D. A. Seale, A. Rapoport, Optimal stopping behavior with relative ranks: The secretary problem with unknown population size. *J. Behav. Decis. Mak.* **13**, 391–411 (2000).
18. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
19. M. Steyvers, M. D. Lee, E. J. Wagenmakers, A Bayesian analysis of human decision-making on bandit problems. *J. Math. Psychol.* **53**, 168–179 (2009).
20. M. D. Lee, S. Zhang, M. Munro, M. Steyvers, Psychological models of human and optimal performance in bandit problems. *Cogn. Syst. Res.* **12**, 164–174 (2011).
21. R. C. Wilson, E. Bonawitz, V. D. Costa, R. B. Ebitz, Balancing exploration and exploitation with information and randomization. *Curr. Opin. Behav. Sci.* **38**, 49–56 (2021).
22. S. Zhang, A. J. Yu, Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting. *Adv. Neural Inf. Process. Syst.* **26**, 2607–2615 (2013).
23. M. Speekenbrink, E. Konstantinidis, Uncertainty and exploration in a restless bandit problem. *Top. Cogn. Sci.* **7**, 351–367 (2015).
24. A. S. Rich, T. M. Gureckis, Exploratory choice reflects the future value of information. *Decision* **5**, 177 (2018).
25. E. Schulz, S. J. Gershman, The algorithmic architecture of exploration in the human brain. *Curr. Opin. Neurobiol.* **55**, 7–14 (2019).
26. C. Baumann, H. Singmann, S. J. Gershman, B. Von Helversen, A linear threshold model for optimal stopping behavior. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12750–12755 (2020).
27. S. Bhatia, L. He, W. J. Zhao, P. P. Analytis, Cognitive models of optimal sequential search with recall. *Cognition* **210**, 104595 (2021).
28. R. C. Wilson, S. Wang, H. Sadeghiyeh, J. D. Cohen, Deep exploration as a unifying account of explore-exploit behavior. PsyArXiv [Preprint] (2020). (Accessed 31 May 2024). https://doi.org/10.31234/osf.io/uj85c.
29. M. Song, Z. Bnaya, W. J. Ma, Sources of suboptimality in a minimalistic explore-exploit task. *Nat. Hum. Behav.* **3**, 361–368 (2019).
30. K. Sang, P. M. Todd, R. L. Goldstone, T. T. Hills, Simple threshold rules solve explore/exploit trade-offs in a resource accumulation search task. *Cogn. Sci.* **44**, e12817 (2020).
31. G. Gigerenzer, P. M. Todd, *Simple Heuristics that Make Us Smart* (Oxford University Press, 1999).

32. F. Lieder, T. L. Griffiths, Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* **43**, e1 (2020).
33. R. Van den Berg, E. Awh, W. J. Ma, Factorial comparison of working memory models. *Psychol. Rev.* **121**, 124 (2014).
34. T. J. Stewart, Optimal selection from a random sequence with learning of the underlying distribution. *J. Am. Stat. Assoc.* **73**, 775–780 (1978).
35. J. R. Anderson, L. J. Schooler, Reflections of the environment in memory. *Psychol. Sci.* **2**, 396–408 (1991).
36. A. Renart, C. K. Machens, Variability in neural activity and behavior. *Curr. Opin. Neurobiol.* **25**, 211–220 (2014).
37. B. Christian, E. M. Russek, T. L. Griffiths, Data from "Resolving Feynman's restaurant problem reveals optimal solutions and human strategies." OSF. https://doi.org/10.17605/OSF.IO/DA5HW. Deposited 25 March 2026.