


Behavior engineering using quantitative reinforcement learning models

Received: 1 May 2023

Accepted: 3 April 2025

Published online: 02 May 2025

 Check for updatesOhad Dan¹✉, Ori Plonsky² & Yonatan Loewenstein^{3,4,5,6}✉

Effectively shaping human and animal behavior is of great practical and theoretical importance. Here we ask whether quantitative models of choice can be used to achieve this goal more effectively than qualitative psychological principles. We term this approach, which is motivated by the effectiveness of engineering in the natural sciences, ‘choice engineering’. To address this question, we launched an academic competition, in which teams of academic competitors used either quantitative models or qualitative principles to design reward schedules that would maximally bias the choices of experimental participants in a repeated, two-alternative task. We found that a choice engineering approach is the most successful method for shaping behavior in our task. This is a proof of concept that quantitative models are ripe to be used in order to engineer behavior. Finally, we show that choice engineering can be effectively used to compare models in the cognitive sciences, thus providing an alternative to the standard statistical methods of model comparison that are based on likelihood or explained variance.

For generations, people have been using folk physics, a qualitative, intuitive understanding of the laws of nature, in order to construct tools and structures. Over the past several centuries, the physical sciences have reached a level of maturity that allows us to accurately describe physical processes using mathematical equations. These equations serve as the foundation for myriad engineering marvels, spanning from the mass production of computers to the towering presence of skyscrapers that dominate our urban landscapes¹.

Shaping the behavior of others, from educating our children to persuading our friends, has always been a prime goal in our society. To this end, we all use intuition and folk wisdom when interacting with fellow human beings². More sophisticated salesmen and politicians often utilize established psychological principles (e.g., anchoring, primacy, recency, etc.) to more successfully sell their products^{3,4}. The use of a qualitative understanding of cognitive processes to shape the behavior of others is analogous to the use of a qualitative understanding of the laws of nature to manipulate the world around us. Motivated by the success of quantitative modeling in modern

engineering, we asked whether quantitative models in the cognitive sciences can be used to efficiently engineer behavior. Quantitative modeling is prevalent when studying operant learning, the learning of association between actions and their consequences⁵, and therefore, we set out to address this question in the framework of an operant learning task.

In a typical operant learning task, a participant is instructed to repeatedly choose between two alternatives and is rewarded according to their choices (Fig. 1a)⁶. Over trials, the participant will bias their choices in favor of the alternative that they deem more rewarding⁷. The resultant sequence of choices made by the participant will depend on the reward schedule, the rule that determines which actions are rewarded and when, as well as the participant’s learning and decision-making strategies. Consider a choice engineer whose goal is to influence the participant’s sequence of choices by constructing an appropriate reward schedule. We expect that an engineer who has a deeper understanding of the strategies employed by the participant will be more effective in constructing the reward schedule that has a desired impact on the participant’s choices.

¹Department of Comparative Medicine, Yale University, New Haven, CT, USA. ²Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel. ³The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, Jerusalem, Israel. ⁴Department of Cognitive and Brain Sciences, The Hebrew University, Jerusalem, Israel. ⁵The Alexander Silberman Institute of Life Sciences, The Hebrew University, Jerusalem, Israel. ⁶The Federmann Center for the Study of Rationality, The Hebrew University, Jerusalem, Israel. ✉e-mail: ohad.dan@yale.edu; yonatan.loewenstein@mail.huji.ac.il

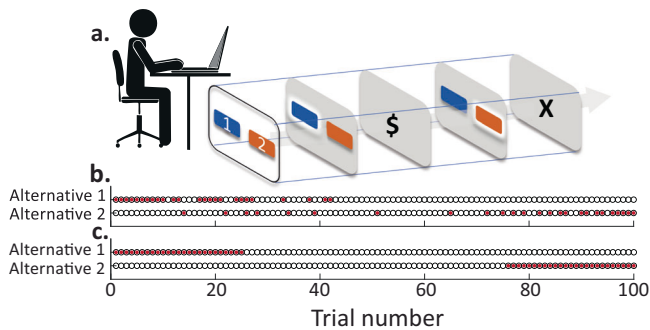


Fig. 1 | Choice engineering assignment. **a** In the experimental task, the participant repeatedly chooses between two alternatives (1 and 2). Following each choice, the participant is rewarded or not rewarded (\$ or X, respectively) in accordance with a predefined binary reward schedule. Unbeknownst to the participant, each of the alternatives was associated with exactly 25 rewards (red-filled circles). **b** With the objective of maximizing bias in favor of alternative 1, a choice engineer can use a model of the participant's learning strategy in order to construct an effective reward schedule. The schedule depicted here is the competition winner, a reward schedule optimized for the behavioral model CATIE. **c** Alternatively, a choice architect may use the principle of primacy and favor allocating as many rewards as possible to alternative 1 at the beginning of the task. Similarly, primacy may dictate deferring all rewards allocated to Alternative 2 to the end of the task.

We demonstrate this idea using a specific engineering assignment: the engineer's goal is to construct a binary reward schedule (allocate binary rewards to choices) that will maximally bias the participant's choices in favor of choosing a predefined alternative, defined here as Alternative 1. Common sense intuition, qualitative psychological principles, and different quantitative reinforcement learning models dictate the same optimal reward schedule: the participant will be maximally biased in favor of Alternative 1 if they are rewarded whenever they choose Alternative 1 but never when they choose the other alternative, Alternative 2. In this case, an accurate model of the participant's strategies does not seem necessary for achieving the engineer's goal. The assignment is more challenging if constraints are added to the allowed reward schedules. For concreteness, we consider the assignment of finding a reward schedule that maximally biases the participant's choices in favor of Alternative 1 in a 100-trial session, such that the schedule assigns a binary reward to each of the alternatives in exactly 25 trials. An engineer that possesses an accurate model of the participant's cognitive strategies can use their model to simulate the expected sequence of choices in response to different reward schedules and use these simulations to select the most effective schedule (Fig. 1b). Possessing a quantitative model, however, is not essential for this task, and one can apply qualitative psychological principles to construct an effective reward schedule. For example, primacy dictates that many rewards should be associated with Alternative 1 at the beginning of the session, whereas the association of rewards with Alternative 2 should be restricted to the end of the session (Fig. 1c).

Sunstein and Thaler coined the term choice architecture in the framework of 'nudges', behavioral interventions that "alter people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives"⁸. Sunstein and Thaler's architecture approach is commonly guided by a qualitative and not quantitative understanding of behavior, and the effectiveness of the interventions can only be tested empirically. Therefore, we adopt the term choice architecture here to describe the use of qualitative psychological principles in order to influence choices without changing the alternatives' objective values. This is in contrast to an approach that is based on quantitative models, which we term choice engineering.

In order to effectively engineer behavior, an accurate model of the participant's strategies is needed. This necessity mirrors the

requirement in the field of civil engineering for precise models of the mechanics of materials when designing tall buildings and long bridges. However, in the field of operant learning, there is no uniquely accepted model of human learning and decision-making⁷. Rather, different models, characterized by different functional forms and parameters, are used in different publications to explain learning behavior in operant tasks. Therefore, rather than committing to a specific quantitative model, we tested the effectiveness of choice engineering by launching the Choice Engineering Competition⁹. In this competition, academics were invited to propose reward schedules (with the constraints described above) with the objective of biasing the average choices of a population of participants in favor of Alternative 1. The reward schedules could be either based on quantitative models (choice engineering) or qualitative models (choice architecture).

The effectiveness of the proposed schedules depends on the accuracy of the model that describes the participants' learning. Therefore, the ability to engineer an effective reward schedule can be used as a measure of the accuracy of the underlying model. Engineering can thus be used as a way of comparing the 'goodness' of models. Moreover, because qualitative models, as well as intuition, can also be used to construct reward schedules, this approach allowed us to compare, using the same scale, qualitative and quantitative models, a feat that cannot be achieved using standard approaches for model comparison that are based on likelihood or explained variance.

We tested the proposed reward schedules with human participants, and here, we report the results of the competition. We found that a reward schedule that was based on choice engineering was comparable to or better than schedules based on qualitative principles, demonstrating the potential effectiveness of choice engineering. This winning schedule was based on CATIE (Contingent Average, Trend, Inertia, and Exploration)¹⁰, a phenomenological model of operant learning. Somewhat surprisingly, schedules that were based on the much more popular Q-learning (QL) model were significantly less effective in this task.

Results

Choice engineering in the competition

In the competition, we tested the effectiveness of 11 different reward schedules on 3386 human participants, each tested on a single schedule. Of the 11 schedules tested, 7 were based on quantitative models, and 4 were based on qualitative principles (see Supplementary Data 1 and Fig. S1). The fraction of trials in which Alternative 1 was chosen in a reward schedule, averaged over all trials and over all participants, is depicted in Fig. 2. We found that there was a significant difference in the efficacy of the different schedules. While the least effective schedule was not statistically significantly different from chance ($50.7\% \pm 1.8\%$, Wilcoxon Test $W(85) = 1841$, $p = 0.953$, $r = 0.006$, bootstrap 95% CI for the median: [45.00, 54.00]), the best-performing reward schedule yielded a bias of $64.3\% \pm 0.5\%$ that was significantly different than chance (Wilcoxon Test $W(591) = 170416.5$, $p < 0.001$, $r = 0.822$, bootstrap 95% CI for the median: [62.00, 64.00]). This reward schedule was submitted by a choice engineer who optimized the reward schedule based on CATIE.

The CATIE model describes an agent whose choices are the result of four principles¹¹. First, similar to other RL models, some choices are driven by exploitation. However, rather than assuming that each alternative is associated with a single 'value', the model learns many 'values', which depend on different histories of actions. Second, the model detects and utilizes trends in the delivery of rewards. Third, choices in the CATIE model are also driven by inertia (repetition of the action chosen in the previous trial). Finally, choices are also driven by exploration. However, rather than using a fixed probability of exploration, the probability of exploration depends on the reward prediction error: the larger the error, the higher the chances that the participant will explore¹².

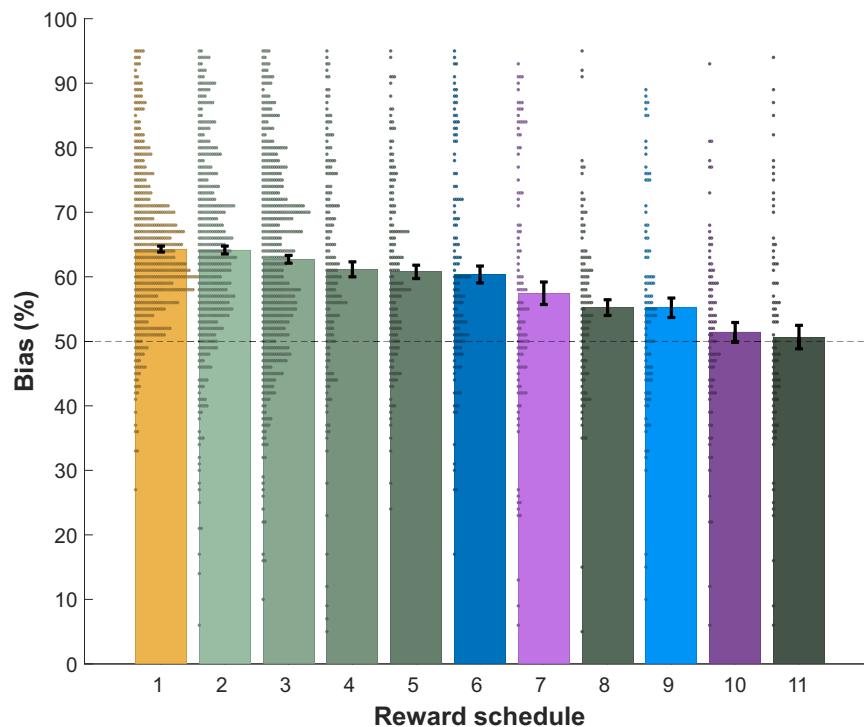


Fig. 2 | Competition results. The bias, average proportion of choices of Alternative 1 for the different reward schedules (see schedules and their description in Fig. S1 and Supplementary Data 1). The winner of the competition (schedule 1, orange) was designed by a choice engineer who utilized the CATIE model, achieving an average bias of 64.3%. Noteworthy are also the results of schedules optimized using a QL model with different sets of parameters (schedules 6, 7, 9, and 10; same color, blue or purple, represent schedules that utilized parameters from the same dataset, see

Supplementary Data 1). The number of participants (data points) per schedule is different (see “Methods” section), ranging from $n = 87$ for schedule 11 to $n = 595$ for Schedule 1. Error bars are the standard error of the mean, centered around the mean. Bias of individual participants is represented by single data points, with identical values stacked from left to right. Data of schedules 2–5, 8, and 11 are arbitrarily represented in green, with brighter shades representing better performance.

Second came a reward schedule that was submitted by a choice architect that combined primacy with strategic considerations to design the reward schedule (Supplementary Data 1). Specifically, they created short streaks of trials in which only Alternative 1 is rewarded, followed by short streaks in which both alternatives were rewarded, in the hope that participants will learn to exclusively choose Alternative 1 by missing the rewards associated with Alternative 2. The average bias of that schedule was $64.1\% \pm 0.6\%$, not significantly different from that of the first place (permutation test, $p = 0.842$, $d = 0.012$, 95% CI $[-1.362, 1.641]$).

The schedule that came third, also constructed by a choice architect, attempted to dynamically manipulate the level of exploration. It achieved a bias of $62.7\% \pm 0.6\%$, significantly lower than the winning schedule (permutation test, $p = 0.037$, $d = 0.121$, 95% CI $[0.078, 3.062]$). All other schedules achieved a significantly lower bias compared to the winning schedule (all $p < 0.004$, see Supplementary Data 1).

Specifically, four different reward schedules were engineered based on the popular QL model. The QL model is a reinforcement learning algorithm that posits that the participant learns the expected average reward associated with each of the alternatives (also known as their values). In a trial, the participant typically chooses the alternative associated with the larger value (exploitation). Occasionally, however, the participant chooses the alternative associated with the lower value (exploration). The models used an ϵ -softmax function to determine the tradeoff between exploration and exploitation¹³. The QL model is characterized by different parameters, and the different schedules were based on variants of the model that utilized different sets of parameters. All four schedules fared considerably worse than the CATIE-based schedule.

The winning schedule induced a bias of $64.3\% \pm 0.5\%$. However, how effective is it relative to the maximally attainable bias? The latter depends on the learning strategies utilized by humans. If, for example, humans employ random exploration¹² then their probability of choosing Alternative 2 will not vanish even under the optimal reward schedule. Heterogeneity between participants is another limiting factor for the maximally attainable bias^{14,15}. This is because a schedule that is optimal for one participant may not be that effective for another participant.

Schedule tuning

We cannot find the optimal schedule or calculate the maximally attainable bias, but it is possible to use a data-driven approach in order to improve a reward schedule (constructed by an engineer, an architect, or even chosen randomly). Such an approach involves iteratively testing various schedules on human participants and refining the schedules based on their performance. The ability of AlphaZero to achieve superhuman performance in games such as chess and Go without incorporating any explicit human knowledge about these games is a testament to the strength of this approach^{16,17}. With sufficient resources, this approach can be successful even without using any domain knowledge about the computational principles or algorithms underlying operant learning in humans. We, therefore, used it to construct a benchmark, relative to which the performance of the competing schedules can be compared.

Specifically, we started by testing participants on the naive reward schedule of Fig. 1c. We then calculated the empirical per-trial probability of choosing each of the alternatives. With these probabilities, we used our intuition to reallocate some of the rewards, with a goal of maximizing the harvest of rewards from Alternative 1 and minimizing

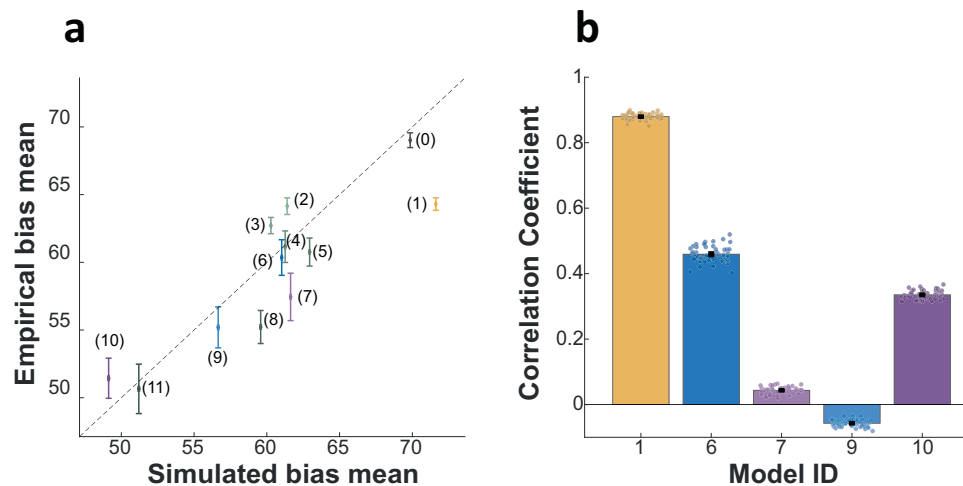


Fig. 3 | Simulated versus experimentally measured bias. **a** The mean bias for each of the reward schedules, as predicted by the CATIE model versus the mean experimentally measured bias (see mean bias and number of participants in Supplementary Data 1, and full bias distribution in Fig. S2). Dashed line is the diagonal. Error bars are the standard error of the means, centered around the mean bias. **b** The high correlation between CATIE's predicted bias and the empirical bias shown in (a) is quantified by the orange bar and compared with the correlation

between the empirical bias and the bias predicted by other models used for engineering (see also Fig. S3). Error bars represent the standard error of the mean, centered around the mean correlation coefficient. Individual points show the correlation coefficients in single simulations. In both panels, the simulated standard error was computed by averaging over 1000 simulations, each with the same number of participants as in the competition. In both panels, the color scheme is identical to Fig. 2.

the harvest of rewards from Alternative 2. We repeated this process three times until the changes that we made to the reward schedule were detrimental to performance (bias in favor of Alternative 1). This process resulted in Schedule 0, which is depicted in Fig. S1 (top). Notably, we executed this iterative process prior to the competition so that we would be able to test this schedule at the time of the competition. We found that across participants, the elicited bias was $69.0\% \pm 0.5\%$. This result is a lower bound of the maximally attainable bias. Given that a random schedule induces a bias of 50%, the effectiveness of the winning schedule is lower than $\frac{64.3\% - 50\%}{69.0\% - 50\%} = 75\%$, but we speculate that it is not much lower.

The CATIE model and the maximally attainable bias

To understand why the schedule that was based on the CATIE model has reached a significantly lower bias, 64.3%, compared to our lower bound for the maximally attainable bias, 69.0% (permutation test, $p < 0.001$, $d = 0.394$, 95% CI [3.310, 6.169]), we considered several hypotheses:

One possibility is that the winning engineer failed to find an optimal schedule for the CATIE model. A corollary of this hypothesis is that when tested on the CATIE model, Schedule 0 would fare better than Schedule 1. This, however, is not the case. Schedule 1 induced a bias of $71.61\% \pm 0.01\%$ in the CATIE model, significantly higher than the bias Schedule 0 induced in that model, $69.85\% \pm 0.02\%$ (permutation test, $p < 0.001$, $d = 0.131$, 95% CI [1.213, 1.870]). This result indicates that the failure of Schedule 1 relative to Schedule 0 is not due to the optimization procedure.

Another possibility for the failure of the CATIE model is that humans may be substantially more heterogeneous in their learning strategies^{18–22}. To test this possibility, we used the fact that a large number of participants were tested on the same reward schedule, and we can quantify their heterogeneity. For example, while averaged over participants, Alternative 1 was chosen in $64.3\% \pm 0.5\%$ of the trials when tested using the winning schedule, some participants chose it in more than 90% of the trials, whereas others chose it in less than 50% of the trials (see Fig. 2 and Fig. S2). The standard deviation of the bias, $11.2\% \pm 0.3\%$, is a measure of the heterogeneity in their performance. This heterogeneity, however, does not necessarily imply that

participants employed different strategies. The reason is that the decision-making strategies can incorporate stochasticity, which can account for variability in their choices (for example, random exploration is a component of both QL and CATIE models). Therefore, we compared the standard deviation of the bias measured over the participants to that predicted by the CATIE model. We found that for the winning schedule, the experimentally observed standard deviation ($11.2\% \pm 0.3\%$) was greater than predicted by the CATIE model ($10.0\% \pm 0.2\%$) by only 1.2% (permutation test, $p < 0.001$, $d = 3.842$, 95% CI [1.15, 1.33]). This result shows that most of the heterogeneity between participants' choices can be accounted for by the CATIE model and not as the result of additional heterogeneity between participants' strategies.

Thus, we conclude that the failure of the CATIE-induced schedule to achieve maximal bias is due to deviations of the CATIE model from the behavior of the experimental participants.

Choice engineering and model comparison

In the competition, a schedule that was based on CATIE fared better than schedules that were based on QL. This raises the question of whether CATIE also describes or predicts participants' choices better than QL. To address this question, we first focused on the goal of the competition, to bias choices in favor of Alternative 1, and asked whether CATIE is more effective in predicting the bias from the schedule than alternative models. To this end, we simulated the four versions of the QL model and the CATIE model and compared, for each of the tested reward schedules, the predicted and experimentally measured biases. This is depicted in Fig. 3a for the CATIE model (see Fig. S3 for the other models). One measure of the agreement between the model and the experimental results is the correlation coefficient r of the predicted bias with the experimentally measured bias. We found that the correlation coefficient of the CATIE model was $r = 0.88$, significantly larger than that of all other models ($p < 0.001$, simulation-based hypothesis testing, see "Methods" section; Fig. 3b).

The CATIE and the QL models can also be compared by their ability to predict individual choices of individual participants. Therefore, another way of quantifying the accuracy of a model is to compare the probabilities p that it assigns to the participants' choices

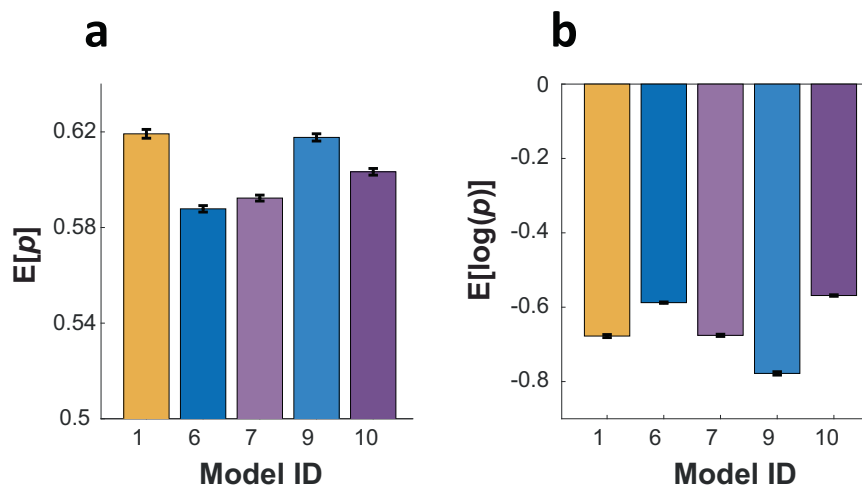


Fig. 4 | Model comparison. The 333,200 decision probabilities (collapsed across all trials and participants) as predicted by 5 models. **a** Average choice probability, $E[p]$ and **b** Log-likelihood, $E[\log(p)]$. While CATIE (orange bar) was the best model with respect to average choice probability, a QL model (schedule 10, dark purple) was

superior with respect to log-likelihood. Error bars are the standard error of the mean. Bar color scheme is identical to Fig. 2. See full distributions in Fig. S4. Complete statistical tests can be found in Tables S1 and S2.

(“Methods” section). The larger p , averaged over trials and participants ($E[p]$), the more accurate the model. Because the parameters of the models were determined independently prior to the competition, there was no need to correct for the number of parameters. As depicted in Fig. 4a, CATIE outperformed the other models in this measure with $E[p] = 61.9\%$, a number that is slightly higher than $E[p] = 61.8\%$, the QL model of schedule 7 that came second ($p = 0.041$, see full statistical comparison in Table S1). Another standard way of comparing models is based on the likelihood of the data given the model. Ranking is done by comparing $E[\log(p)]$, where the larger the number (closer to zero), the more accurate the model (Fig. 4b). In this analysis, the QL model of schedule 10 significantly obtained the highest likelihood, whereas the CATIE model came only third (all $p < 0.001$, see full statistical comparison in Table S2). The reason why CATIE outperformed the other models with respect to $E[p]$ but not with respect to $E[\log(p)]$ is that in the CATIE model, there was a relatively large fraction of actions to which the model assigned a relatively small probability (Fig. S4). A more nuanced picture emerges when comparing the predictive power of the models per schedule (Fig. S5). CATIE outperformed the other models in some schedules, but this was no longer true for some of the other schedules. Taken together, while CATIE was better than QL in predicting the overall bias of the participants (Fig. 3), all models were comparable in their ability to predict individual choices. The ‘winning’ model with respect to the predictions of individual choices depended on the objective function ($E[p]$ vs. $E[\log(p)]$) (Fig. 4), as well as the specifics of the reward schedule (Fig. S5).

Worthwhile noting that while 11 schedules were tested in the competition, only five models were compared with respect to their abilities to describe the data. This is because while qualitative models can be used to construct successful reward schedules (‘architect’), only quantitative models can be used to compare models by their ability to quantitatively describe behavior.

Discussion

In this paper, we used an academic competition to compare the abilities of qualitative principles on one side and quantitative models on the other side to influence human choices by designing a reward schedule. The design principles underlying the two most effective schedules were radically different. The winning schedule was engineered using a quantitative model of choice. The runner-up, who was

almost equally effective, was based on an intuitive understanding of human nature. The fact that the outcome of an automated optimization approach is as effective as the best intuition of academics is proof of principle of the potential of choice engineering.

Faced with an objective function and a quantitative model of human learning strategy, the winning choice engineer successfully designed an effective reward schedule using standard optimization techniques. This is an approach that can be automatically applied to other objective functions and constraints. By contrast, the choice architect approach that is based on human expertise and intuition cannot be easily automated, and any change in the objectives and constraints would require the laborious design of a new schedule.

Operant learning

In the last decades, reinforcement learning algorithms have dominated the field of operant learning²³. These algorithms have been used to model human and animal learning behaviors, and effort has been made to identify the brain regions, neurons, and neurotransmitters that implement them²⁴. This development has been motivated in part by the proven ability of these algorithms to learn complicated tasks, from motor control tasks to playing games without prior knowledge²⁵. The QL algorithm is a most successful example. In machine learning, it converges (under certain conditions) to the optimal solution²⁶. In the cognitive sciences, it has been widely used to model the behavior of animals and humans. In neuroscience, its variables have been linked to neural activity in specific neurons in specific brain regions^{27,28}. In view of these, the success of CATIE is surprising. Particularly remarkable is the fact that the parameters of this model were not fitted to this particular competition. Rather, they were taken from previous experiments. The CATIE model is a phenomenological model that has not been derived from first principles and has not been proven effective in the field of machine learning. Its success, therefore, may indicate that caution should be exercised when using reinforcement learning algorithms to model human learning strategies.

Model comparison

Model comparison is becoming increasingly popular in the cognitive sciences. Methods that are based on maximum likelihood are popular because they have a Bayesian interpretation²⁹. If we know that behavior was generated by one of the models and a-priori they are equally likely,

the model associated with the maximum likelihood is the one most likely to be the correct one. However, if none of the models is exact (as is always the case), there is no clear interpretation of a higher likelihood of one model compared to another³⁰. Therefore, there is no unique method for model comparison (see also Fig. 4 and Fig. S5). The ability of a model to engineer behavior is an alternative approach to model comparison. This approach is also not unique, as there is no general rule that dictates the proper task for engineering. In fact, in the academic competition⁹ we offered two tracks. In the ‘static’ track, described in this paper, academic participants proposed a full reward schedule, as described in this paper. In the ‘dynamic’ track, they could use a more sophisticated approach and dynamically allocate, in each trial, the reward(s) of the next trial. Regrettably, only choice architects participated in the dynamic track, and therefore, we were unable to use it in order to compare choice architecture to choice engineering (see Supplementary Data 2).

Engineering competitions vs. prediction competitions

There is a long tradition of choice prediction competitions, in which academic participants compete in their ability to predict behavior in different settings^{31–33}. A clear advantage of a prediction competition over an engineering competition is that in the latter, different experiments (here, different schedules) must be assigned to each participant. By contrast, in the former, the same experiment can be used for all participants. In this sense, a prediction competition is much more efficient than an engineering one. Engineering competitions, however, may have their own advantages, especially in their ability to compare qualitative and quantitative approaches.

In prediction competitions, participants are typically required to make predictions without predetermining all the parameters of the test. This is done by forcing the participants to submit a computer code, which computes the predictions as a function of the parameters³³. The requirement of a code can be a barrier to qualitative-thinking scholars, hampering the comparison of the performance of qualitative and quantitative approaches (architects vs. engineers). This, however, is a technicality, and it is easy to envision a prediction competition in which the participants would be requested to submit a set of numerical predictions (e.g., by predetermining the parameters) without the need to quantitatively justify them. A more important advantage of the engineering approach is that making quantitative predictions about our fellow humans may be less important ecologically than finding ways of biasing their choices, and therefore also less intuitive. For example, effectively convincing your offspring to drive more carefully is, perhaps, ecologically more important than quantifying the probability that they will actually do it. Along the same lines, it may be more intuitive to rank order the effectiveness of different schedules than to predict the actual biases that they will induce. We hypothesize that humans are experts not in quantitatively predicting behavior in any arbitrary scenario, but rather, in generating situations in which the behavior of fellow humans is malleable and then taking actions that would drive that behavior in a desired direction. This is the regime in which an engineering competition takes place.

Implications and limitations

Influencing humans’ preferences and choices is a major objective in our society, and scientific research has been harnessed to achieve this goal³⁴. For over a century, psychological principles have been used to construct effective advertising and political campaigns³⁵. More recently, large data and statistical methods are used, especially in online advertising, to achieve similar goals^{36,37}. Choice engineering can be used as an additional tool in this toolkit. Its main advantage is that it does not need large data. Given a model, the objectives can be achieved almost automatically by solving an optimization problem.

The generalizability of our findings and, more broadly, of the engineering approach, is constrained by the specific experimental task

used in the competition. We designed the task with the intention that the primary behavioral drivers of participants would be captured by operant learning models, a field that has been extensively studied using quantitative approaches. Not all behavioral domains, however, are as extensively studied in this way, limiting the application potential of behavioral engineering approaches. For example, it is not clear how to apply the engineering approach to biasing humans towards buying one brand of soft drink over another or voting for one political candidate and not another.

The competition is inherently limited in its ability to assess the absolute goodness of the winning model. As a criterion for model comparison, the better performance of a model in the competition could suggest that the model better represents human behavior. However, by design, the competition compared only those models that were used to optimize the schedules submitted to the competition. There may very well be other models that would have produced schedules with an even greater mean bias compared to all those observed in our dataset. While we attempted to bound the maximal attainable bias in the competition (Results), this bound does not substantially inform the specifics of a behavioral model that would have yielded such performance.

In terms of the winning model, generalizing our conclusions may be limited by the constraints we imposed on the reward schedule design, namely associating binary rewards with exactly 25 trials per alternative. It is possible that under different conditions, participants use different behavioral strategies^{38,39}. In that case, under different experimental conditions, for example, a different number of trials or different reward distributions, the same models compared in the competition may have ended up with different performance rankings. Finally, considering the winning model, one limitation of CATIE, compared with QL, is the generalization to other operant problems. While QL can be directly applied to a large family of decision problems known as Markov Decision Problems, the generalization of CATIE to this family of decision problems is unclear. Despite these limitations, the competition is a proof of concept for the efficiency and effectiveness of choice engineering, which could become a common tool in the coming years. Therefore, the social and ethical implications of choice engineering should be considered⁴⁰.

In this work, we explore an applicative aspect of behavioral models. Of interest to practitioners, our results demonstrate that behavioral models in the cognitive sciences can be used not only to describe or predict behavior but also, akin to engineering approaches in the natural sciences, to shape and engineer behavior. Furthermore, we show that employing quantitative cognitive models to engineer behavior is both efficient, as they can be integrated into automated optimization procedures, and empirically effective. For cognitive scientists, we emphasize the conceptual challenges in selecting methods for model comparison. We propose the engineering approach as an additional tool for future modelers. Overall, the success of the behavior-engineering approach highlights the importance of advancing quantitative models in the cognitive sciences and the potential of applying them to enhance individual and societal well-being.

Methods

Experimental task

The study was approved by the Hebrew University Committee for the Use of Human Subjects in Research, and all participants provided informed consent. Recruitment was based on the online labor market, Amazon Mechanical Turk. Data for the static track were collected from 3386 participants. Participants received a fixed participation fee of \$0.4 and a bonus of an additional 1¢ for every obtained reward. In the experiment, participants were presented with two choice alternatives, displayed as rectangles on the left and right sides of the screen, one colored in red and the other in blue in random order. For each

participant, one of the two rectangles was randomly set as Alternative 1, and the other as Alternative 2.

Participants were instructed to make sequential choices between these alternatives and were informed that some choices might yield rewards. The number of remaining trials was shown on the screen, starting at 100 and decreasing by 1 after each choice. Rewards were displayed as smiley face icons that accumulated in a dedicated section of the screen, representing the total rewards earned so far. Choices that did not result in a reward were marked by a sad smiley icon, which did not accumulate. The experiment concluded after 100 trials.

We considered all the participants who completed the experiment (100 choices). We excluded from the analysis 54 participants (1.6%) who chose one of the alternatives less than 5 times and are therefore likely to have ignored the reward schedule. Adding them to the analysis does not change the ranking of the first five schedules. We did not collect any demographic information about the participants.

The competition

Competition rules are described in detail in Dan & Loewenstein⁹ and are available on the competition website at <http://decision-making-lab.com/competition/index.html>. Briefly, we published an open call for participation in the competition⁹ and accepted all valid submissions. The goal of the participants in the competition was to generate the reward schedule that is the most effective in biasing experimental participants' choices in favor of Alternative 1, as described in the Introduction. We also invited participants to contribute a dynamic reward schedule in which, in each trial, the rewards are allocated based on past choices (see "Discussion" section and Supplementary Data 2).

To identify the most effective schedule, we tested the schedules on human participants. As described by Dan & Loewenstein⁹, we utilized an adaptive statistical method that allocates a larger number of participants to schedules with better performance (see the "Sample size" section).

Behavioral models

For the model comparison of Fig. 4, we predicted the choice of each participant in each trial t based on all the participant's choices in preceding trials 1 to $t-1$ and the model. In heterogeneous models (see Supplementary Data 1), we fit a parameter distribution over the population and weighted predictions based on each set's likelihood of generating the observed choices. Specifically, for each participant and for each trial t , we calculated the likelihood that choices in trials 1 to $t-1$ were generated by each set of parameters, and then used this probability to weight the prediction of that set for the overall choice predictions in trial t .

The QL and CATIE models are described in Dan & Loewenstein (2019)⁹. For completeness, we provide these descriptions here.

The Q-learning model is a reinforcement learning algorithm that has been widely used to model sequential decision-making behavior in humans and animals^{7,41,42}. Applied to the experimental task (formally, a one-state MDP), Q-learning describes how the value (expected average reward) associated with each action changes in response to that trial's chosen action and the resultant reward. Formally, at each trial t , the agent selects an action $a_t \in \{1, 2\}$ and receives a reward $r_t \in \{0, 1\}$. The updated value of the selected action $Q_{t+1}(a_t)$, is a weighted average of the previous value $Q_t(a_t)$ and the received reward:

$$Q_{t+1}(a_t) = Q_t(a_t) + \eta(r_t - Q_t(a_t)) \quad (1)$$

where $0 < \eta \leq 1$ is the learning rate. The value of the nonchosen action remains unchanged $Q_{t+1}(a \neq a_t) = Q_t(a \neq a_t)$. The difference Eq. (1) is not complete without specifying the initial conditions. Following experimental findings, we posit that in the first two trials, the two alternatives are chosen in random order, and the initial value of each of the two actions is the reward received the first time this action was chosen¹³.

Equation (1) describes how the action-values adapt over trials but does not specify how these action-values are used to select actions. The mapping between action-values and actions is given by an action-selection rule. Motivated by experimental findings¹³, we consider an ϵ -softmax action-selection rule such that:

$$\Pr[a = a_i] = (1 - \epsilon) \frac{e^{\beta Q_t(a_i)}}{\sum_a e^{\beta Q_t(a)}} + \epsilon \quad (2)$$

where $0 \leq \epsilon \leq 1$ and $\beta > 0$ are exploration parameters.

The contingent average, trend, inertia, and exploration (CATIE) model is a phenomenological model developed to explain human behavior in tasks of sequential choices with partial feedback, similar to the experimental task of the competition¹⁰. The model describes an agent that, in each trial, chooses according to one of four modes: exploration, exploitation, inertia, and a simple heuristic. In the first two trials, the agent samples the two alternatives in random order. In each consecutive trial, one of the four modes is selected according to the following probabilistic rules:

1. If the agent chose the same alternative in the past two consecutive trials and their outcomes differed, with a probability τ in the subsequent trial, the selected mode is 'trend' (also called 'heuristic' mode).
2. If the trend mode was not chosen, with a probability p_{explore} (see below) the selected mode is 'explore'.
3. If neither the trend nor the explore modes were selected, the agent selects the inertia mode with a probability ϕ .
4. If none of the above modes were chosen, the agent selects the contingent average mode.

In the trend mode, the agent would choose the same alternative as in the previous trial if the trend of the last two outcomes was positive, or switch if it was negative:

$$(a_{t+1} | \text{trend mode}; a_t, a_{t+1} = a^i) = \begin{cases} a^i & r_t > r_{t-1} \\ a^j & r_t < r_{t-1} \end{cases}$$

Where $a^i, i = \{1, 2\}$ are the available alternatives in the task, a_t is the action at trial t (i.e., the choice of a^1 or a^2), and $r_t \in \{0, 1\}$ is the reward at trial t . In the context of the competition (binary rewards), trends are manifested by either a sequence of reward followed by no reward (negative trend), or no reward followed by a reward (positive trend).

In the explore mode, the agent chooses between the alternatives randomly, with an equal probability. The probability of entering the explore mode is a function of the sequence of surprises experienced by the agent up until the current trial. Surprise _{t} $\in [0, 1]$, is calculated at trial t as follows:

$$\text{Surprise}_t = \begin{cases} \frac{|\text{Exp}_t^i - V_t^i|}{\text{ObsSD}^i + |\text{Exp}_t^i - V_t^i|}, & \text{if } \text{ObsSD}^i > 0 \\ 0, & \text{otherwise} \end{cases}$$

Where Exp_t^i is the reward the agent expects to receive at trial t after choosing an alternative i at trial t (the alternative's 'contingent average', see below), V_t^i is the actual reward obtained by choosing the alternative i at trial t , and ObsSD^i is the standard deviation of all pay-offs observed from the alternative i at trials 1 through t . Note that at each trial, the surprise is calculated relative to the single alternative chosen by the agent. Given the sequence of previous surprises, the probability of entering the explore mode, conditioned on not choosing the heuristic mode, is:

$$p_{\text{explore}} = p(\text{Explore} | \text{No_Heuristic})_{t+1} = \epsilon * \frac{1 + \text{Surprise}_t + \text{Mean_Surprise}}{3}$$

Where ϵ is a free parameter capturing a basic exploration tendency, and Mean.Surprise is the average surprise in trials 1 through t . Thus, the probability of exploration is minimal (equals $\frac{\epsilon}{2}$) when the surprise is minimal (both Surprise_t and Mean.Surprise equal 0), and maximal (approaches ϵ) when the surprise is large.

The inertia mode simply repeats the same choice as the previous trial.

In the contingent average (CA) mode, the agent chooses the alternative associated with the higher k -CA, defined as the average payoff observed in all previous trials which followed the same sequence of k outcomes (the same ‘contingency’ of length k). At its initialization, the agent uniformly draws an integer k from the discrete range $\{0, \dots, K\}$ ($k \sim u(0, K)$, K is a free parameter of the model). Each time the agent enters the CA mode, it calculates k -CA for the two alternatives. For each alternative i , k -CA is the average reward obtained in trials that followed a contingency of length k which is identical to that of the most recent k -contingency and was followed by a choice of i . In particular, $k=0$ implies averaging all the past outcomes of an alternative. To illustrate this process, consider the following sequence of 10 choices in alternative 1 (“A”) and 2 (“B”): [A, A, A, A, A, A, A, B, B, B], implying the agent is currently at the 11th trial, and assume that these choices yielded the following reward sequence [1, 0, 0, 0, 1, 0, 0, 1, 0, 0]. An agent characterized by a $k=2$ would imply that the current contingency is [0, 0] (the sequence of the last two rewards). Previous indices where the same reward sequence was observed are 3, 4, and 7; the choices that followed these sequences are A, A, and B, respectively; and the rewards that followed these contingencies are 0, 1, and 1 (the rewards of trials 4, 5, and 8). This would imply that the contingent averages of alternatives A and B are 0.5 and 1, respectively. Hence, in the current trial, alternative B, which has the higher contingent average, would be chosen. If for alternative i there are no past k -contingencies matching the most recent contingency, the value of a different random k -contingency replaces the contingent average for that alternative. A random k -contingency value is obtained by considering the set of all reward sequences of length k which were followed by the choice of i , randomly choosing one of them with a uniform probability, and calculating its contingent average value. Note that choosing uniformly from past trials implies that recurring contingencies have a higher probability of being chosen. If there are no contingencies of length k that are followed by a choice of i , a k smaller by 1 is used iteratively until at least one contingency exists. Since the model begins by sampling both alternatives in the first two trials, for $k=0$ it is guaranteed that both alternatives would have a 0-contingency value and hence that the CA mode is well defined.

As described above, the model is characterized by four free parameters (one for each mode: τ, ϵ, ϕ, K). The parameters of the model were estimated from behavior¹⁰, and these estimations were used in all simulations: $\tau = 0.29$, $\epsilon = 0.30$, $\phi = 0.71$, $K = 2$.

Schedule optimization

The schedule optimization algorithm we utilized is described by Dan & Loewenstein⁹. For completeness, we provide its description here. Note that this method is not guaranteed to find a global optimum, nor to converge to the same solution in independent applications.

To find a reward schedule that attains a high bias given a specific model, we employed an iterative optimization. In the first iteration, an initial sequence S_1 is generated by randomly shuffling, independently for the two alternatives, a sequence that complies with the sequence constraints (rewards are assigned to 25 of the experiment’s 100 trials). In the k ’th iteration, a partially reshuffled sequence S_k is generated from the current sequence S_k by choosing $t \sim u\{2, T_k\}$ indices, with uniform probability without replacement, independently for the two alternatives, and randomly reshuffling the rewards assigned to the trials of these indices. The resulting schedule S_k is tested N times

(equivalent to N independent participants) and compared to the bias previously obtained for the unshuffled schedule. This procedure is repeated m times and the schedule used for the subsequent iteration is the one that yields the largest bias out of the $m+1$ tested schedules (m shuffled schedules in addition to the original schedule). The variable T_k , which represents the ‘temperature’ of the algorithm, progressively decreased to zero, such that $T_k = \phi(k)$, where $\phi(k) = \lceil \frac{A-k}{B} \rceil$ where A, B are parameters. The parameters we used for this procedure are $A = 10,000$, $B = 100$, $N = 500$, $m = 10$. Finally, we note that the reward of the last trial cannot change behavior. Hence, it is never advantageous to allocate a reward to the last trial of Alternative 1. If the stochastic optimization converges to a solution that does not meet this criterion, the reward schedule of Alternative 1 is modified to conform with this heuristic by reallocating the reward of the last trial to a random trial currently not associated with a reward. Similarly, if a reward is not assigned to the last trial of Alternative 2, one of the rewards associated with that alternative is randomly chosen and reallocated to the last trial. Due to the stochastic nature of this optimization, we repeated the complete procedure $n = 20$ times and chose the schedule that yielded the maximal bias.

Statistical analyses

Analysis was conducted using Matlab 2023A. All reported statistical tests are 2-tailed. To assess the statistical difference of a specific bias distribution from chance, we used Matlab’s signrank function to perform the Wilcoxon signed-rank test. The reported effect size for this test (r) was calculated as the z-statistic normalized by the square root of the sample size, excluding elements, e , equal to chance level ($r = z / \sqrt{N - \sum_{i=1}^N \delta_{e_i, 50\%}}$). Confidence intervals for this test were calculated using bootstrapping. Specifically, given a distribution D containing N elements, we generated bootstrap samples by randomly sampling N elements with replacement from D . For each bootstrap sample, the median was calculated. This procedure was repeated 10,000 times, and the 95% confidence interval was defined as the range between the 2.5th and 97.5th percentiles of the resulting distribution of medians.

To compare the statistical difference in the bias distribution of two schedules, we used the following permutation test: given two distributions D_1 and D_2 , with n_1 and n_2 elements respectively, we first calculated the true difference between the two distributions’ means. Next, we pooled D_1 and D_2 and, defining a single permutation, we randomly drew, without replacement, n_1 and n_2 elements from the joint distribution, calculating the difference between the means of both groups. Repeating this process 10,000 times, we calculated the proportion of differences from the permuted groups that exceeded the true difference between the means of the original groups and reported it as a p -value. The reported effect size for these tests is the true difference between the means, normalized by the two groups’ pooled variance⁴³. Confidence intervals were calculated using the following bootstrap procedure: n_1 and n_2 elements were randomly sampled with replacement from D_1 and D_2 , respectively. This procedure was repeated 10,000 times, and for each iteration, the difference between the means of the two resampled distributions was calculated. The 2.5th and 97.5th percentiles of the resulting distribution of mean differences were used to define the 95% confidence interval.

To compare the correlation coefficients between the actual bias observed in the 12 schedules and the bias predicted by each model (Fig. 3b), we simulated a choice sequence for each schedule for every model. For each simulation, we calculated the correlation coefficient between the actual bias and the simulated bias. Repeating this process 10,000 times, we report as p -value the proportion of simulations in which the correlation coefficient of one model was greater than another.

Sample size

To identify the most effective schedule, we utilized the method of Successive Rejects⁴⁴. According to this method, more participants (more “samples”) are allocated to better-performing reward schedules. The method of Successive Rejects operates iteratively. In the first iteration, all reward schedules are allocated an equal number of participants. Then, the worst performing schedule (the schedule associated with the minimal empirical bias) is “rejected” in the sense that no more participants are allocated to it. The process continues iteratively until just two schedules remain. The schedule that yields the highest bias over all tested schedules is considered the most effective one. According to this process, the number of participants that should be used in each iteration depends on (1) the total number of schedules tested, (2) the (unknown) distribution of their “true” biases, and (3) the desired confidence level for identifying the most effective one. We pre-determined the number of participants per iteration, assuming the true bias of the schedules is evenly distributed between 50% (chance level) and 70%. One minor complication when running online experiments is that not all the participants who start the task also end up completing it. We randomly allocated participants to schedules in advance, which resulted in some variability in the number of participants who completed the task per schedule. To make sure that each schedule is tested on no fewer than the number of participants required by the Successive Reject method, based on pilot data, we pre-allocated 30% more participants than required to each schedule.

The data presented in the paper were collected in three phases:

Phase 1: In the formal part of the competition, we used the method of Successive Rejects with a confidence level of 95%, and tested 10 of the schedules (0–6, 8–9, and 11, see below) on 941 participants. We found that schedule 0 was the most effective schedule, and schedule 1 was the second most effective schedule. We declared schedule 1 to be the winner of the competition.

Phase 2: After concluding the competition, we decided to test two additional schedules that were not included in the formal part of the competition (schedules 7 and 10). To do that, we ran a pseudo-competition over all 12 schedules using the same Successive Reject method, this time requiring a confidence level of 99%. A higher confidence level entails a larger number of participants per iteration, and a larger number of schedules entails a larger number of iterations. We repeated the process of Successive Rejects for the pseudo-competition. The initial participants used in the earlier iterations were those collected in the formal part of the competition, and when needed, we collected additional participants. We concluded this pseudo-competition with 2633 participants (941 of those collected in the formal competition). The results of the pseudo-competition were similar to those of the competition: schedule 0 was the most effective schedule, followed by schedule 1.

Phase 3: By the end of phase 2, the bias induced by schedules 1–3 was rather similar. We were wondering if by increasing the number of participants, we would be able to better distinguish between them statistically. We added participants to equalize to number of participants per schedule (as before, we allocated more participants than necessary, by a factor of 1.3, because of dropouts). This increased the total number of participants to 3332. The ranking of the schedules was not different from that of Phase 2.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data files can be found on the competition website: <https://sites.google.com/view/cec19/home> as well as the competition's repository⁴⁵: <https://doi.org/10.5281/zenodo.14548839>.

Code availability

Supporting code may be found on the competition website: <https://sites.google.com/view/cec19/home> as well as the competition's repository⁴⁵: <https://doi.org/10.5281/zenodo.14548839>.

References

1. National Research Council, Committee on the Mathematical Sciences in 2025. *Fueling Innovation and Discovery: The Mathematical Sciences in the 21st Century* (National Academies Press, 2012).
2. Hutto, D. D. & Ratcliffe, M. (eds.) *Folk Psychology Re-Assessed* (Springer, 2007).
3. Alderson, W. Psychology for marketing and economics. *J. Mark.* **17**, 119–135 (1952).
4. Donthu, N., Kumar, S., Pattnaik, D. & Lim, W. M. A bibliometric retrospective of marketing from the lens of psychology: insights from Psychology & Marketing. *Psychol. Mark.* **38**, 834–865 (2021).
5. Staddon, J. E. R. & Cerutti, D. T. Operant conditioning. *Annu. Rev. Psychol.* **54**, 115–144 (2003).
6. Thyer, B. A. (ed.) *Comprehensive Handbook of Social Work and Social Welfare: Human Behavior in the Social Environment* Vol. 2 (John Wiley & Sons, 2008).
7. Mongillo, G., Shteingart, H. & Loewenstein, Y. The misbehavior of reinforcement Learning. *Proc. IEEE* **102**, 528–541 (2014).
8. Thaler, R. H. & Sunstein, C. R. *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale Univ. Press, 2008).
9. Dan, O. & Loewenstein, Y. From choice architecture to choice engineering. *Nat. Commun.* **10**, 2808 (2019).
10. Plonsky, O. & Erev, I. Learning in settings with partial feedback and the wavy recency effect of rare events. *Cogn. Psychol.* **93**, 18–43 (2017).
11. Plonsky, O., Teodorescu, K. & Erev, I. Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychol. Rev.* **122**, 621–647 (2015).
12. Fox, L., Dan, O., Elber-Dorozko, L. & Loewenstein, Y. Exploration: from machines to humans. *Curr. Opin. Behav. Sci.* **35**, 104–111 (2020).
13. Shteingart, H., Neiman, T. & Loewenstein, Y. The role of first impression in operant learning. *J. Exp. Psychol. Gen.* **142**, 476–488 (2013).
14. Ratcliff, R., Voskuilen, C. & McKoon, G. Internal and external sources of variability in perceptual decision-making. *Psychol. Rev.* **125**, 33–46 (2018).
15. Renart, A. & Machens, C. K. Variability in neural activity and behavior. *Curr. Opin. Neurobiol.* **25**, 211–220 (2014).
16. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
17. Hassabis, D. Artificial intelligence: chess match of the century. *Nature* **544**, 413–414 (2017).
18. Lebovich, L., Darshan, R., Lavi, Y., Hansel, D. & Loewenstein, Y. Idiosyncratic choice bias naturally emerges from intrinsic stochasticity in neuronal dynamics. *Nat. Hum. Behav.* **3**, 1190–1202 (2019).
19. Berry, A. S., Jagust, W. J. & Hsu, M. Age-related variability in decision-making: Insights from neurochemistry. *Cogn. Affect. Behav. Neurosci.* **19**, 415–434 (2018).
20. Peters, J. & Büchel, C. The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn. Sci.* **15**, 227–239 (2011).
21. Wyart, V. & Koehlin, E. Choice variability and suboptimality in uncertain environments. *Curr. Opin. Behav. Sci.* **11**, 109–115 (2016).
22. Laquitaine, S., Piron, C., Abellanas, D., Loewenstein, Y. & Boraud, T. Complex population response of dorsal putamen neurons predicts the ability to learn. *PLoS ONE* **8**, e80683 (2013).
23. Shteingart, H. & Loewenstein, Y. Reinforcement learning and human behavior. *Curr. Opin. Neurobiol.* **25**, 93–98 (2014).
24. Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J. & Kurth-Nelson, Z. Deep Reinforcement Learning and Its Neuroscientific Implications. *Neuron* **107**, 603–616 (2020).

25. Mousavi, S. S., Schukat, M. & Howley, E. Deep reinforcement learning: an overview. *Lect. Notes Netw. Syst.* **16**, 426–440 (2018).
26. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 2018).
27. Niv, Y. Reinforcement learning in the brain. *J. Math. Psychol.* **53**, 139–154 (2009).
28. Dayan, P. & Daw, N. D. Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* **8**, 429–453 (2008).
29. Pratt, J. W. Bayesian interpretation of standard inference statements. *J. R. Stat. Soc. B* **27**, 169–203 (1965).
30. Le Cam, L. Maximum likelihood: an introduction. *Int. Stat. Rev.* **58**, 153–171 (1990).
31. Erev, I., Ert, E., Plonsky, O., Cohen, D. & Cohen, O. From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychol. Rev.* **124**, 369 (2017).
32. Plonsky, O. et al. Predicting human decisions with behavioral theories and machine learning. Preprint at <https://arxiv.org/abs/1904.06866> (2019).
33. Erev, I. et al. A choice prediction competition: choices from experience and from description. *J. Behav. Decis. Mak.* **23**, 15–47 (2010).
34. Fennis, B. M. & Stroebe, W. *The Psychology of Advertising* (Routledge, 2020).
35. Poffenberger, A. T. *Psychology in Advertising* (A. W. Shaw Co., 1925).
36. Hofacker, C. F., Malthouse, E. C. & Sultan, F. Big data and consumer behavior: imminent opportunities. *J. Consum. Mark.* **33**, 89–97 (2016).
37. Lee, H. & Cho, C.-H. Digital advertising: present and future prospects. *Int. J. Advert.* **39**, 332–341 (2020).
38. Soltani, A. & Koehlin, E. Computational models of adaptive behavior and prefrontal cortex. *Neuropsychopharmacology* **47**, 58–71 (2022).
39. Ferguson, T. D., Fyshe, A., White, A. & Krigolson, O. E. Humans adopt different exploration strategies depending on the environment. *Comput. Brain Behav.* **6**, 671–696 (2023).
40. Schmidt, A. T. & Engelen, B. The ethics of nudging: an overview. *Philos. Compass* **15**, e12658 (2020).
41. Barto, A. G., Sutton, R. S. & Watkins, C. J. C. H. *Learning and Sequential Decision Making* (Univ. Massachusetts, 1989).
42. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
43. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 2013).
44. Audibert, J.-Y. & Bubeck, S. Best arm identification in multi-armed bandits. In *COLT - 23th Conference on Learning Theory - 2010*, 13 (Haifa, Israel, 2010).
45. Dan, O. & Loewenstein, Y. Behavior engineering using quantitative reinforcement learning models. *GitHub repository; ohaddan/competition_analysis* <https://doi.org/10.5281/ZENODO.14548839> (2024).

Acknowledgements

We would like to thank Lea Kaplan for technical assistance, Peter Dayan, Amir Dezfouli, Ido Erev, Yosef Rinott, and Jon Roiser for helpful discussions, and all the participants in the competition. This work was supported by the Israel Science Foundation (ISF) Grant No. 757/16 (Y.L.) and the Gatsby Charitable Foundation (Y.L.). Y.L. is incumbent of the David and Inez Myers Chair in Neural Computation.

Author contributions

O.D. and Y.L. analyzed the data and wrote the manuscript. O.P. contributed the winning schedule.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58888-y>.

Correspondence and requests for materials should be addressed to Ohad Dan or Yonatan Loewenstein.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025