Measuring how people learn how to plan

Yash Raj Jain
Rationality Enhancement Group, MPI for Intelligent Systems, Tübingen, Germany
Birla Institute of Technology & Science, Pilani, Hyderabad, India

Frederick Callaway
Department of Psychology, Princeton University, NJ, USA

Falk Lieder
Rationality Enhancement Group, MPI for Intelligent Systems, Tübingen, Germany
Bernstein Center for Computational Neuroscience, Tübingen, Germany

Abstract

How can people learn to make better decisions and become more far-sighted? To make the underlying learning mechanisms more accessible to scientific inquiry, we develop a computational method for measuring the time course of experience-dependent changes in people's planning strategies. We validated our method on simulated and empirical data: on simulated data its inferences were significantly more accurate than simpler approaches, and when evaluated on human data it correctly detected the plasticity-enhancing effect of performance feedback. Having validated our method, we illustrate how it can be used to gain new insights into the time course and nature of cognitive plasticity. Future work will leverage our method to i) reverse-engineer the learning mechanisms enabling people to acquire complex cognitive skills such as planning and problem-solving and ii) measure individual differences in cognitive plasticity.

Keywords: cognitive plasticity; planning; decision-making; process-tracing; statistical methods

Introduction

One of the most remarkable features of the human mind is its ability to continuously improve itself. As helpless babies develop into mature adults, their brains do not only acquire impressive perceptual and sensory-motor skills and knowledge about the world but they also learn to think, to make better decisions, to learn, and to monitor and adaptively regulate themselves. These phenomena are collectively known as *cognitive plasticity*. Just like the acquisition of perceptual skills (Hubel & Wiesel, 1970), the acquisition of cognitive skills requires specific experiences and practice (van Lehn, 1996; Ericsson, Krampe, & Tesch-Römer, 1993).

Despite initial research on how people acquire cognitive skills (van Lehn, 1996; Shrager & Siegler, 1998; Krueger, Lieder, & Griffiths, 2017), the underlying learning mechanisms are still largely unknown. Reverse-engineering how people learn how to think and how to decide is very challenging because we can neither observe people's cognitive strategies, nor how they change with experience – let alone the underlying learning mechanisms. Instead, cognitive plasticity has to be inferred from observable changes in behavior. This is difficult because each observed behavior could have been generated by many possible cognitive mechanisms. This problem is pertinent to all areas of cognition. As a first step towards a more general solution, we develop a computational method for measuring how people's planning strategies change depending on the person's experience. Initial work suggested that metacognitive reinforcement learning might play an important role in how people come to plan farther ahead (Krueger et al., 2017) and which strategies they use (Lieder & Griffiths, 2017) but the postulated mechanisms are difficult to investigate because cognitive plasticity has remained unobservable.

Our approach combines a recently developed process-tracing paradigm that renders people's behavior highly diagnostic of their planning strategies with probabilistic models of planning and learning that constrain the space of potential cognitive mechanisms and exploit temporal dependencies among subsequent planning strategies. Critically, our measurement model can be inverted to infer the sequence of people's planning strategies from the clicks they make in the process tracing paradigm. Our computational method makes it possible to observe how people's planning strategies change from each decision to the next. This sheds new light on the time course and the nature of metacognitive learning. Future work will reverse-engineer the learning mechanisms that generate the cognitive plasticity our approach is bringing to light.

The plan for this paper is as follows: we start by developing a computational method for measuring experience-dependent changes in people's planning strategies. Next, we validate it on synthetic data and human data. We then illustrate the utility of our method by measuring the time course of how people learn how to plan, characterizing the revealed learning trajectories, and testing hypotheses about cognitive plasticity. In closing, we discuss directions for future work.

Methods

Process-tracing using the Mouselab-MDP paradigm

Planning, like all cognitive processes, cannot be observed directly but has to be inferred from observable behavior. This is generally an ill-posed problem. To address this challenge, researchers have developed *process-tracing* methods that elicit and record behavioral signatures of latent cognitive processes; for instance decision strategies can be traced by recording the order in which people inspect the payoffs of different gambles (Payne, Bettman, & Johnson, 1993). While these behavioral signatures are still indirect measures of cognitive processes, they do provide additional information about

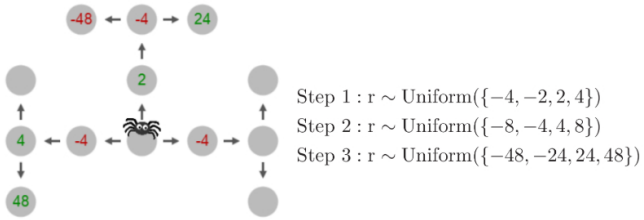

Figure 1: Illustration of the Mouselab-MDP paradigm. Rewards are revealed by clicking, prior to selecting a path with the arrow keys. The distribution of rewards underlying each node at a given step is shown on the right.

what the underlying cognitive strategy might be.

Here, we employ a process-tracing paradigm that externalizes people's beliefs and planning operations as observable states and actions (Callaway, Lieder, Krueger, & Griffiths, 2017; Callaway et al., 2018). Inspired by the Mouselab paradigm (Payne et al., 1993), the Mouselab-MDP paradigm uses people's mouse-clicking as a window into their planning.

The Mouselab-MDP paradigm illustrated in Figure 1 presents a series of route planning problems where each location (the gray circles), harbors a gain or loss. These potential gains and losses are initially occluded, corresponding to a highly uncertain belief state. The participant can reveal each location's reward by clicking on it and paying a fee. This is similar to looking at a map to plan a road trip. Clicking on a circle corresponds to thinking about a potential destination, evaluating how enjoyable it would be to go there, and adjusting one's assessment of candidate routes accordingly.

Measurement model

To develop an efficient computational method for inferring the temporal evolution of people's planning strategies, we make the simplifying assumption that the trial-by-trial sequence of peoples' cognitive strategies $(S_1, S_2, \cdots, S_{31})$ forms a Markov chain whose hidden states emit the observed process tracing data collect on each trial $(\mathbf{d}_1, \cdots, \mathbf{d}_{31})$. This hidden Markov model requires additional methodological assumptions about i) how cognitive strategies manifest in process-tracing data, ii) the space of cognitive mechanisms that can be learned, and iii) the nature and amount of cognitive plasticity that might occur. The following paragraphs detail our assumptions about each of these three components in turn.

Observation model. To plan in the Mouselab-MDP paradigm participants have to gather information by making a sequence of clicks. Our observation model thus specifies the probability of of observing a sequence of clicks $\mathbf{d_t}$ on trial t if the strategy was S_t (i.e., $P(\mathbf{d_t}|S_t)$).

To achieve this, we quantify each planning strategy's propensity to generate a click c (or stop collecting information) given the already observed rewards encoded in belief state b by a weighted sum of 29 features $(f_1(b,c), \cdots, f_{29}(b,c))$. The features describe the click c relative to this information (e.g., by the value of the largest reward that can be collected from the inspected location) and in terms of the action it gathers information about (e.g., whether it pertains to the first, second, or third step)[1]. The *depth* feature, for instance, describes each click by whether it looks 1, 2, or 3 steps into the future. The features and weights jointly determine the strategy's propensity to make click c in belief state b according to

$$P(\mathbf{d_t}|S_t) = \prod_{i=1}^{|\mathbf{d_t}|} \frac{\exp\left(\frac{1}{\tau} \cdot \sum_{k=1}^{|w^{(S)}|} w_k^{(S)} \cdot f_k^{(S)}(c_{t,i}, b_{t,i})\right)}{\sum_{c \in \mathcal{C}_{b_t}} \exp\left(\frac{1}{\tau} \cdot \sum_{k=1}^{|w^{(S)}|} w_k^{(S)} \cdot f_k^{(S)}(c, b_{t,i})\right)}, \quad (1)$$

where $d_{t,i}$ is the i$^{\text{th}}$ click the participant made on trial t (or the decision to stop clicking and take action), the decision temperature τ was set to 0.5 to match the variability of people's click sequences, and $w^{(S)}$ is the weight vector of strategy S.

Space of cognitive mechanisms. We formulated a set of 38 strategies $(\mathcal{S})$[1] to describe the process tracing data from Lieder (2018). These strategies include the optimal goal-setting strategy (Callaway et al., 2018) that starts by inspecting the possible final destinations and search-based planning algorithms such as breadth-first search, depth-first search, and best-first search (Russell & Norvig, 2016). 76.7% of the click sequences were the most likely instantiation of one of the 38 strategies. The clicks of the remaining 23.3% of the sequences were, at worst, second most likely under the best fitting strategy. These strategies differ in how much information they consider (ranging from none to all), which information they focus on, and in the order in which they collect it.

Building on the observation model in Equation 1, we represent each strategy by a weight vector $\mathbf{w} = (w_1, \cdots, w_{29})$ that specifies the strategy's preference for more vs. less planning, considering immediate vs. long-term consequences, satisficing vs. maximizing, avoiding losses (cf. Huys et al., 2012), and other desiderata. These weights span a high-dimensional continuous space with many intermediate strategies and mixtures of strategies. Cognitive plasticity could be measured by tracking how those weights change over time. But this would be a very difficult ill-defined inference problem whose solution would depend on our somewhat arbitrary choice of features. As a first approximation, our method therefore simplifies the problem of measuring cognitive plasticity to inferring a time-series of discrete strategies.

To understand what types of strategies people use, we grouped our 38 strategies using hierarchical clustering. This requires measuring the similarity between strategies. Since the strategies are probabilistic, we defined the distance metric $\Delta(s_1, s_2)$ between strategy s_1 and s_2 as the Jensen-Shannon

[1] A detailed description of the features and strategies is available at https://osf.io/y58d3/?view_only=fa2f89de3aa04d4d87af3d050bb1a64c

divergence (Lin, 1991) between the distributions of click sequences and belief states induced by strategies s_1 and s_2 respectively, that is

$$\Delta(s_1, s_2) = \mathrm{JS}\left[p(\mathbf{d}|s_1), p(\mathbf{d}|s_2)\right], \tag{2}$$

and approximate it using Monte-Carlo integration.

Applying Ward's hierarchical clustering method (Ward Jr, 1963) to the resulting distances suggested 11 types of planning strategies: acting impulsively without any planning, finding a goal and immediately moving towards it, inspecting both immediate and final outcomes (but no intermediate ones), overly frugal goal setting strategies, goal setting strategies that plan towards potential goals even when it is wasteful, exhaustive backward planning strategies that inspect all of the states, other far-sighted strategies that inspect all potential final states, forward-planning strategies similar to depth-first search, forward-planning strategies similar to best-first search, strategies similar to breadth-first search, and strategies that focus on the course of action that has received the most consideration so far.

Prior on strategy sequences. Inferring a strategy from a single click sequence could be unreliable. Our method therefore exploits temporal dependencies between subsequent strategies to smooth out its inferences. Transitions from one strategy to the next can be grouped into three types: repetitions, gradual changes, and abrupt changes. While most neuroscientific and reinforcement-learning perspectives emphasize gradual learning (e.g., Hebb, 1949; Mercado III, 2008; Lieder, Shenhav, Musslick, & Griffiths, 2018), others suggest that animals change their strategy abruptly when they detect a change in the environment (Gershman, Blei, & Niv, 2010). Symbolic models and stage theories of cognitive development also assume abrupt changes (e.g., Piaget, 1971; Shrager & Siegler, 1998), and it seems plausible that both types of mechanisms might coexist. To accommodate these different perspectives, we consider three prior distributions on participants' trial-by-trial sequence of cognitive strategies.

The *gradual learning prior* (m_{gradual} in Equation 3) assumes that strategies changes gradually, that is

$$P(S_{t+1} = s|S_t, m_{\text{gradual}}) = \frac{\exp(-\frac{1}{\tau} \cdot \Delta(s, S_t))}{\sum_{s' \in \mathcal{S}} \exp(-\frac{1}{\tau} \cdot \Delta(s', S_t))}, \tag{3}$$

where $\mathcal{S}$ is the set of strategies, $|\mathcal{S}|$ is the number of strategies, and the temperature parameter τ was set to achieve a 50% chance of a strategy change. By contrast, the *abrupt changes prior* (m_{abrupt} in Equation 4) assumes that transitions are either repetitions or jumps.

$$\begin{aligned} &P(S_{t+1} = s|S_t, m_{\text{abrupt}}) = \\ &\quad p_{\text{stay}} \cdot \mathbb{I}(S_{t+1} = S_t) + (1 - p_{\text{stay}}) \cdot \frac{\mathbb{I}(s \neq S_t)}{|\mathcal{S}|-1}, \end{aligned} \tag{4}$$

Finally, the *mixed prior* (m_{mixed} in Equation 5) assumes that both types of changes coexist.

$$\begin{aligned} &P(S_{t+1} = s|S_t, m_{\text{mixed}}) = \\ &\quad p_{\text{gradual}} \cdot P(S_{t+1} = s|S_t, m_{\text{gradual}}) \\ &\quad + (1 - p_{\text{gradual}}) \cdot P(S_{t+1} = s|S_t, m_{\text{abrupt}}). \end{aligned} \tag{5}$$

In each of these three cases, we model the probability of the first strategy as a uniform distribution over the space of decision strategies (i.e., $P(S_1) = \frac{1}{|\mathcal{S}|}$).

Together with the observation model and the strategy space described above each of these priors defines a generative model of a participant's process tracing data $\mathbf{d}$; this model has the following form:

$$P(\mathbf{d}, S_1, \cdots, S_T) = \frac{1}{|\mathcal{S}|} \cdot \prod_{t=2}^{T} P(S_t|S_{t-1}, m) \cdot P(\mathbf{d_t}|S_t). \tag{6}$$

The three measurement models differ in the identity of $m \in \{m_{\text{gradual}}, m_{\text{abrupt}}, m_{\text{mixed}}\}$. Inverting these models gives rise to a computational method for measuring an important aspect of cognitive plasticity.

Inference on cognitive plasticity

The models above describe how changes in cognitive strategies manifest in process-tracing data. To measure those cognitive changes, we have to reason backwards from the process tracing data $\mathbf{d}$ to the unobservable cognitive strategies $S_1, \cdots, S_T$ that generated it. To achieve this, we leverage the Viterbi algorithm (Forney, 1973) to compute maximum a posteriori (MAP) estimates of the hidden sequence of planning strategies $S_1, \cdots, S_T$ given the observed process tracing data $\mathbf{d}$, the measurement model m, and its parameters (p_{stay} for m_{abrupt} and p_{gradual} and p_{stay} for m_{mixed}). To estimate the model parameters we perform grid search with a resolution of 0.02 over $p_{\text{stay}} \in [0, 1]$ for m_{abrupt} and $(p_{\text{stay}}, p_{\text{gradual}}) \in [0, 1] \times [0, 1]$ for m_{mixed}.

Inferring the hidden sequence of cognitive strategies in this way lets us see otherwise unobservable aspects of cognitive plasticity through the lens of a computational microscope.

Validating the computational microscope

Validation on synthetic data

To validate our "computational microscope" for looking at cognitive plasticity, we apply it to simulated process tracing data. To avoid bias towards any one of the three measurement models, we used each of them to generate a data set with 100 simulated participants completing 31 trials each. We then combined the resulting three data sets into a single data set from 300 simulated participants.

We then inverted the three measurement models on each of the simulated trials ($\mathbf{d}$) and compared the maximum a posteriori estimate of each strategy sequence ($\hat{\mathbf{S}}$) against the ground truth (S) in terms of the proportion of correctly inferred strategies and the distance between the inferred strategies and the ground truth. To measure the distance between

two sequences of n planning strategies we define $\Delta(\mathbf{v},\mathbf{w})$ as $\frac{1}{n}\cdot\sum_{i=1}^{n}\Delta(v_i,w_i)$. For better interpretability, the relative distance $\Delta_{\text{rel}}(s_1,s_2)=\Delta(s_1,s_2)/\bar{\Delta}$ normalizes $\Delta(s_1,s_2)$ by the average distance between any strategy and its closest neighbour.

As a baseline, we evaluated the computational method that inverts the observation model in Equation 1 on each click sequences independently. This simple approach was sufficient to infer the correct strategy about 81% of the time (95% confidence interval: $[80.2\%,81.8\%]$). The average distance from the inferred strategy to the true one was only 21% of the average distance from each strategy to its closest neighbor ($\Delta_{\text{rel}}(\hat{\mathbf{s}}^{\text{baseline}},\mathbf{s})=0.215$, 95% confidence interval: $[0.20,0.23]$). This shows that the simulated click sequences were highly diagnostic of the strategies that generated them.

We found that exploiting the temporal dependencies among subsequent strategies by using either of the three measurement models significantly improved the proportion of correctly inferred strategies to 88.5%, 88.3%, and 88.5% for m_{gradual}, m_{abrupt}, and m_{mixed} respectively (all $p<0.0001$) and decreased the average distance between the inferred strategies and the ground truth by more than 40% ($\Delta_{\text{rel}}(\hat{\mathbf{s}}^{\text{gradual}},\mathbf{s})=0.124$, $\Delta_{\text{rel}}(\hat{\mathbf{s}}^{\text{mixed}},\mathbf{s})=0.124$, and $\Delta_{\text{rel}}(\hat{\mathbf{s}}^{\text{abrupt}},\mathbf{s})=0.127$, all $p<0.0001$). The minor differences between the accuracies and distances achieved with the three measurement models were not statistically significant ($\chi^2(2)=0.36$, $p=0.8373$ and $F(2,897)=0.06$, $p=0.942$ respectively). These results suggest that – under reasonable, theory-agnostic assumptions about what cognitive plasticity might be like – our computational microscopes for looking at cognitive plasticity can be expected to produce more accurate measurements than simpler methods.

Which measurement model is most suitable depends on whether the measured changes are mostly gradual, mostly abrupt, or a combination of both. This may vary across tasks and participants. We therefore invert all three measurement models on each participant's data and select the most appropriate measurement model for each participant according to the Akaike Information Criterion (Akaike, 1974). We then interpret the inferences obtained from inverting the selected model as the measurement of our computational microscope.

Validation on empirical data

To validate our computational microscope on empirical data, we applied it to the Mouselab-MDP process-tracing data from Experiments 1–3 by Lieder (2018) where 176 participants solved 31 different 3-step planning problems of the form shown in Figure 1. Concretely, we asked if our computational microscope can detect the effect of an experimental manipulation expected to promote cognitive plasticity, namely the feedback participants in the second condition of Experiment 1 received on the (sub)optimality of their chosen actions. This performance feedback stated whether the chosen move was sub-optimal and included a delay penalty whose duration was proportional to the difference between the expected returns of the optimal move versus the chose one.

Our computational microscope successfully detected the effect of this manipulation. As shown in Figure 2, the inferred learning-induced changes were significantly larger in the feedback condition than in the control condition in the first 15 trials and in trials 21–25 ($p\leq 0.012$ for each 5-trial bin) and nearly significant in trials 15–20 ($p=0.08$) and trials 25–30 ($p=0.06$). Furthermore, Figure 2 also shows that cognitive plasticity slowed down over time as participants adapted to experiment's stationary decision environment.

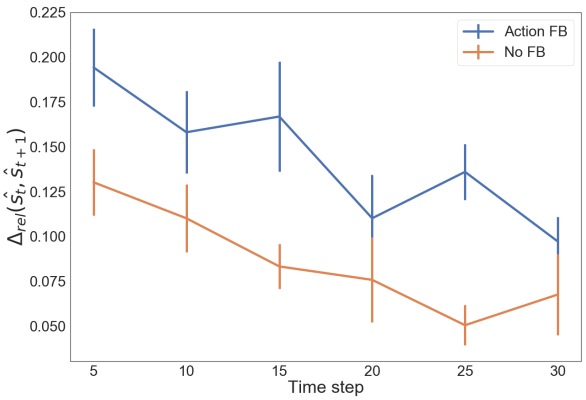

Figure 2: Feedback accelerates cognitive plasticity. This figure shows that feedback increased the amount of cognitive plasticity at the beginning of learning.

Next, we performed χ^2-tests with the Sidak correction for multiple comparisons to compare the frequencies of all possible strategy transitions (i.e., $P(S_{t+1}|S_t)$) between the experimental condition with action feedback versus the control condition. We found that action feedback selectively increased the probability of eight performance-increasing transitions from a strategy with a lower average performance (S_t) to a strategy with a higher average performance (S_{t+1}) and significantly decreased the probability of five performance-decreasing transitions and five strategy repetitions ($S_{t+1}=S_t$). By contrast, the feedback decreased the frequency of only one performance-increasing strategy-transition and increased the frequency of only two performance-decreasing strategy transitions.

Our method's ability to detect the plasticity-enhancing effects of feedback suggests that its inferences provide a valid measure of cognitive plasticity.

Shedding light on cognitive plasticity

Having validated our computational microscope on both simulated and empirical data, we now leverage it to measure how people learn how to plan by applying it to the process tracing data from the control conditions of Experiment 1 and the training phases of the control conditions of Experiments 2 and 3 from Lieder (2018). In the following, we illustrate how our computational microscope can be used to i) measure how people's propensity to use different cognitive strate-

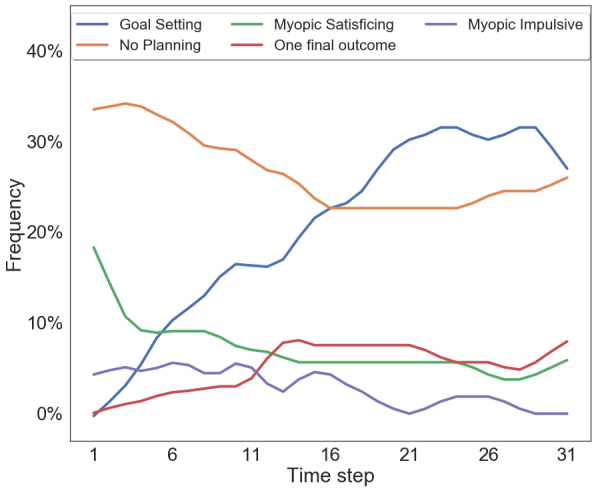

Figure 3: Time course of strategy usage frequencies of the five most common strategies.

gies evolves over time, ii) test theories of cognitive development and cognitive plasticity, and iii) characterize people's metacognitive learning trajectories.

Temporal evolution of strategy frequencies. As shown in Figure 3, we found that the most common initial strategy was to act impulsively without any planning (*No Planning*). The prevalence of this strategy decreased gradually over time from about 34% on average across the first five trials to about 25% on average across the last five trials ($\chi^2(1) = 7.95$, $p = 0.0048$).[2]. Conversely, the frequency of the near-optimal *Goal Setting* strategy increased from about 4% to 30% ($\chi^2(1) = 148.85, p < 0.0001$). The frequencies of the two maladaptive strategies that decide based on immediate rewards (*Myopic Satisficing* and *Myopic Impulsive*) dropped from about 11% and 4% respectively to about 5% ($\chi^2(1) = 11.74, p = 0.0006$) and 0.6% ($\chi^2(1) = 11.62, p = 0.0006$) respectively, whereas the frequency of the strategy *One Final Outcome* that prioritizes long-term consequences increased from about 1% to about 6% ($\chi^2(1) = 20.22, p < 0.0001$). Jointly these strategies accounted for about 53%–72% of our participants' planning across the different trials of our experiment.

Testing hypotheses about the nature of cognitive plasticity. Prominent theories of cognitive development disagree about whether it proceeds in discrete stages (Piaget, 1971) with abrupt transitions or continuous gradual change (Siegler, 1996). Inspired by these theories, we asked to which extent learning how to plan in the Mouselab-MDP paradigm proceeds through gradual changes versus abrupt transitions. Our computational microscope suggested that cognitive plasticity includes both gradual and abrupt strategy changes. We observed that the data from $63.0\% \pm 4.9\%$ of our participants was best captured by the abrupt model, while the data from $29.8\% \pm 4.6\%$ of the participants were best captured by the gradual model, and the data from $7.2\% \pm 2.6\%$ were best captured by the mixed model. A more fine-grained analysis of the individual inferred transitions revealed that the majority of strategy changes was gradual (i.e., 59.1%, $\chi^2(1) = 56.8$, $p < 0.0001$) but there was also a non-negligible percentage of abrupt changes (i.e., 40.9%). In total those different types of strategy changes constituted 22.8% of all transitions; that is 77.2% of the inferred transitions were strategy repetitions.

Siegler's overlapping waves theory (Siegler, 1996) asserts that multiple cognitive strategies are being used in parallel at each time during cognitive development. It further asserts that the relative frequencies of these strategies shift towards increasingly more adaptive strategies and that there are intermediary strategies whose frequency waxes and vanes. Under the strong assumption that the underlying plasticity mechanisms are the same as those that drive learning in the Mouselab-MDP paradigm, we predicted that the same patterns should also occur in the participants' strategy sequences. To test the first prediction, we performed χ^2-tests on the strategies' frequencies in all bins of 5 consecutive trials. In support of the hypothesis that multiple different strategies are used at each point in time throughout the learning process we found that on average 2.16 strategies were each used by significantly more than 5% of our participants in any given trial of the experiment (95% confidence interval: $[2.02, 2.30]$). Consistent with the prediction that high-performing strategies become more prevalent over time whereas low-performing strategies become less prevalent over time we found a significant rank correlation between each strategies' average performance and the change in their frequency from the first trial to the last trial (Spearman's $\rho(37) = 0.39$, $p = 0.0154$). On the population level, we did not find any evidence for intermediary strategies whose average frequency across participants initially increases and later decreases again. That is, there was no strategy whose frequency was higher in the middle two time bins than in both the first two time bins and the last two time bins. Yet, overall the measurements we obtained with our computational microscope suggest that learning in the Mouselab-MDP paradigm is better described by the overlapping waves theory than by stage theories of cognitive development.

Learning trajectories. To identify the most common learning trajectories, we categorized each inferred strategy as belonging to one of the 11 types of strategies described earlier. We then extracted the order in which different strategy types appeared in the inferred sequences. Using this analysis, we found that there were almost as many unique learning trajectories as there were learners: The 110 participants

[2] All χ^2-tests in this paragraph compare the average frequency in the first five trials against the average frequency in the last five trials.

who changed their strategy at least once displayed 94 unique learning trajectories; that is 85.4% of the learning trajectories were unique and the remaining trajectories were exhibited by only 2–4 learners each. Zooming in on the 49 participants who learned the near-optimal goal setting strategy, we found that they reached the near-optimal goal setting strategy via 38 unique learning trajectories. Consistent with the overlapping waves theory, we found that 84.2% of these learning trajectories included at least one intermediary strategy between the initial strategy and the final strategy. Most importantly, our analysis revealed three dominant gateways to optimal planning: 35% of the penultimate strategies inspected all potential final states – whereas the optimal strategy stops searching for better final states once it encounters the best possible outcome – and sometimes planned backwards from undesirable states; 27% of the penultimate strategies inspected the potential final states in a manner akin to the optimal strategy but additionally and wastefully inspected paths towards undesirable final outcomes, and 21% of the penultimate strategies inspected both immediate and final outcomes while ignoring the intermediate states. This suggests that participants discovered the optimal goal setting strategy via intermediate strategies that perform gratuitous planning. Furthermore, we found that about 42% of participants who succeeded to learn a near-optimal goal setting strategy started with strategies that inspect both immediate and final outcomes without looking at intermediate ones. In addition to the 110 participants who changed their initial strategy, 66 participants (37.5%) never changed their strategy. The majority of those participants always acted impulsively without any planning (21% of all participants). Consistent with the interpretation that those participants were less engaged in the experiment and had not paid close attention to the instructions, we found that they performed substantially worse on the four attention check questions at the end of the experiment than participants who had demonstrated learning (1.7 errors vs. 0.8 errors on average; $t(111) = -5.80, p < .0001$). In addition, 9% of all participants always inspected immediate and final outcomes while ignoring intermediate rewards, 4% always focused exclusively on final outcomes, and 3.5% used other types of strategies.

Discussion

We have successfully validated our method on both synthetic and human data. The results suggest that our computational microscope can measure cognitive plasticity in terms of the temporal evolution of people's cognitive strategies.

Our findings suggest that this method has great potential for helping cognitive scientists uncover the mechanisms of cognitive plasticity and how they are impacted by the learning environment, individual differences, time pressure, motivation, and interventions – including feedback, instructions, and reflection prompts.

We are optimistic that computational microscopes will become useful tools for reverse-engineering the learning mechanisms that enable people to acquire complex cognitive skills and shape the way we think and decide. To make this possible, we will extend the proposed measurement model to continuous strategy spaces, a wider range of tasks and strategies, and learning at the timescale of individual cognitive operations. In addition, future work will also leverage our computational microscope to elucidate individual differences in cognitive plasticity within and across psychiatric conditions and different age groups.

The tentative conclusions we obtained with our first prototype of a computational microscope for measuring cognitive plasticity should be taken with a grain of salt because more psychologically plausible distance metrics and more realistic strategy representations could lead to different conclusions about the nature of cognitive plasticity. In this first step, we determined the similarity between strategies based on their behavior. But two strategies that look very different could result from similar mechanisms. Future work will identify a low-dimensional continuous strategy space by decomposing each strategy into its Pavlovian, habitual, and model-based components (van der Meer, Kurth-Nelson, & Redish, 2012). This more realistic representation will allows us to measure the similarity between strategies by comparing the underlying neurocomputational mechanisms. In addition, we will seek to validate the robustness of our computational microscope by measuring its performance on data generated from more realistic models of cognitive plasticity (e.g., Krueger et al., 2017; Lieder et al., 2018).

The approach developed in this paper makes it possible to more directly observe the previously hidden phenomenon of cognitive plasticity in many of its facets – ranging from skill acquisition, learning to think differently, reflective learning, cognitive decline, self-improvement, changes in cognitive dispositions, and the onset, progression, and recovery from psychiatric symptoms and mental disorders. This will make it easier to reverse-engineer people's ability to discover and continuously refine their own algorithms.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., & Griffiths, T. L. (2018). A resource-rational analysis of human planning. In *Proceedings of the 40th annual conference of the cognitive science society.*

Callaway, F., Lieder, F., Krueger, P. M., & Griffiths, T. L. (2017). Mouselab-mdp: A new paradigm for tracing how people plan. In *The 3rd multidisciplinary conference on reinforcement learning and decision making.*

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, *100*(3), 363.

Forney, G. D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268–278.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological review*, *117*(1), 197.

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory.* John Wiley & Sons Inc.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *The Journal of physiology*, *206*(2), 419–436.

Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, *8*(3), e1002410.

Krueger, P. M., Lieder, F., & Griffiths, T. L. (2017). Enhancing metacognitive reinforcement learning using reward structures and feedback. In *Proceedings of the 39th annual conference of the cognitive science society.*

Lieder, F. (2018). Developing an intelligent system that teaches people optimal cognitive strategies. In F. Lieder (Ed.), *Beyond bounded rationality: Reverse-engineering and enhancing human intelligence* (chap. 8).

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762–794. doi: 10.1037/rev0000075

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS computational biology*, *14*(4), e1006043.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory*, *37*(1), 145–151.

Mercado III, E. (2008). Neural and cognitive plasticity: From maps to minds. *Psychological Bulletin*, *134*(1), 109.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge university press.

Piaget, J. (1971). *The theory of stages in cognitive development.* McGraw-Hill.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach* (3rd ed.). Harlow, UK: Pearson Education Limited.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410. doi: 10.1111/1467-9280.00076

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

van der Meer, M., Kurth-Nelson, Z., & Redish, A. D. (2012). Information processing in decision-making systems. *The Neuroscientist*, *18*(4), 342–359.

van Lehn, K. (1996). Cognitive skill acquisition. *Annual review of psychology*, *47*(1), 513–539.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236–244.