A cognitive tutor for helping people overcome present bias

Falk Lieder[*]
Max Planck Institute for Intelligent Systems
Tübingen, Germany
falk.lieder@tuebingen.mpg.de

Frederick Callaway[*]
Princeton University
Princeton, NJ, USA
fredcallaway@princeton.edu

Yash Raj Jain
Max Planck Institute for Intelligent Systems
Tübingen, Germany
f20140604@hyderabad.bits-pilani.ac.in

Paul M. Krueger
Princeton University
Princeton, NJ, USA
paul.m.krueger@gmail.com

Priyam Das
UC Irvine
Irvine, CA, USA
priyam.das@uci.edu

Sayan Gul
UC Berkeley
Berkeley, CA, USA

Thomas L. Griffiths
Princeton University
Princeton, NJ, USA
tomg@princeton.edu

[*] These authors contributed equally.

Abstract

People's reliance on suboptimal heuristics gives rise to a plethora of cognitive biases in decision-making including the *present bias*, which denotes people's tendency to be overly swayed by an action's immediate costs/benefits rather than its more important long-term consequences. One approach to helping people overcome such biases is to teach them better decision strategies. But which strategies should we teach them? And how can we teach them effectively? Here, we leverage an automatic method for discovering rational heuristics and insights into how people acquire cognitive skills to develop an intelligent tutor that teaches people how to make better decisions. As a proof of concept, we derive the optimal planning strategy for a simple model of situations where people fall prey to the present bias. Our cognitive tutor teaches people this optimal planning strategy by giving them metacognitive feedback on how they plan in a 3-step sequential decision-making task. Our tutor's feedback is designed to maximally accelerate people's metacognitive reinforcement learning towards the optimal planning strategy. A series of four experiments confirmed that training with the cognitive tutor significantly reduced present bias and improved people's decision-making competency: Experiment 1 demonstrated that the cognitive tutor's feedback can help participants discover far-sighted planning strategies. Experiment 2 found that this training effect transfers to more complex environments. Experiment 3 found that these transfer effects are retained for at least 24 hours after the training. Finally, Experiment 4 found that practicing with the cognitive tutor can have additional benefits over being told the strategy in words. The results suggest that promoting metacognitive reinforcement learning with optimal feedback is a promising approach to improving the human mind.

Keywords: cognitive training; planning; metacognitive reinforcement learning; cognitive plasticity

Acknowledgements

This work was supported by grant number ONR MURI N00014-13-1-0341 and a grant from the Templeton World Charity Foundation to TLG. We thank Tania Lombrozo, Peter Dayan, Thomas Hills, and Mike Mozer for helpful comments and discussions.

1 Introduction

Research on heuristics and biases has identified many ways in which human judgment and decision-making might be sub-optimal (Tversky and Kahneman, 1974). For instance, decision-makers often partly or entirely neglect the long-term consequences of their choices while being overly swayed by their immediate costs or benefits in the present. This phenomenon is known as the *present bias* (O'Donoghue and Rabin, 1999) and gives rise to procrastination, impulsivity, and poor economic decisions, such as people's failure to save for retirement. Interventions that help people overcome present bias could thus confer significant benefits to individuals, organizations, and society.

Two of the main challenges for improving the human mind are a) discovering effective cognitive strategies, and b) teaching them effectively so that people will apply them in everyday life. Our recent work has begun to address the first problem by developing automatic methods for deriving resource-rational cognitive strategies (Lieder and Griffiths, 2019; Lieder et al., 2017). Here, we address the second problem by developing and evaluating a cognitive tutor that teaches people far-sighted planning strategies.

We start by introducing the experimental paradigm we use to observe and intervene on the decision strategies that might give rise to the present bias. We then derive a resource-rational planning strategy for overcoming the present bias. Next, we present a cognitive tutor that teaches people this cognitive strategy via a metacognitive feedback mechanism. Finally, we evaluate our cognitive tutor in a series of experiments and discuss our findings and directions for future work.

2 Measuring present bias with the Mouselab-MDP paradigm

Previous work has inferred present bias from the choices that people make. But the underlying mechanisms remain to be elucidated. Here, we use a recently developed process-tracing paradigm (Callaway et al., 2017) to measure the decision strategies that might give rise to present bias. Making future-minded decisions requires planning. Planning, like all cognitive processes, cannot be observed directly but has to be inferred from observable behavior. To be able to give people feedback on the quality of their planning strategies we therefore draw on the Mouselab-MDP paradigm (Callaway et al., 2017) illustrated in Figure 1a and 1c to make people's behavior more diagnostic of their planning strategies. This paradigm externalizes people's beliefs and planning operations as observable states and actions (Callaway et al., 2018,1). Inspired by the Mouselab paradigm (Payne et al., 1993), it uses people's mouse-clicking as a window into their planning. Participants are presented 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 of \$1.

The key property of situations in which present bias impairs human decision-making is the misalignment between immediate reward and long-term value. As an illustration of this problem, consider the choice between beginning work on a manuscript and watching a YouTube video. Staring at a blank page might make you feel anxious in the short run, but you will feel very satisfied when you submit the paper for publication months later. By contrast, the YouTube video will give you immediate joy but you might later come to regret the wasted time. To make good decisions in situations like this, people have to look beyond the salient immediate rewards, set a goal for the future, plan how to achieve it, and execute the plan. What makes this far-sighted approach worthwhile is that the range of outcomes that can be obtained by a concerted sequence of actions over an extended period of time is much larger than the range of rewards that can be attained immediately.

To capture this aspect of many real-world situations within the Mouselab paradigm, we constructed 3-step sequential decision environments where the range of rewards increases from the first step to the second step and was largest in the third step. In each episode, rewards are independently drawn from discrete uniform distributions; in the first step the possible values were $\{-4, -2, +2, +4\}$; in the second step the possible values were $\{-8, -4, +4, +8\}$; and in the third step the possible values were $\{-48, -24, +24, +48\}$. To plan their actions participants can uncover the rewards at each location by clicking on it for a fee of \$1 per click. This captures that the decision-maker's time is costly.

Recording the clicks people make in this paradigm allows us to detect whether their decisions are swayed by present bias. That is, if a participant only inspects the immediate outcomes of the first step while ignoring the outcomes of the second step and the third step, then we know that their decision was affected by present bias. Conversely, if a participant looks at the potential final outcomes in the third step while ignoring the immediate outcomes, we can be confident that they were not swayed by present bias. Our paradigm thereby allows us to i) observe the maladaptive heuristics that give rise to present bias, and to ii) trace whether and how they improve in response to interventions. To develop an effective intervention, the next section derives the optimal planning strategy for the environment modelled by this paradigm.

3 Discovering optimal planning strategies that counter present bias

Teaching clever heuristics is a promising approach to improving decision-making (Gigerenzer and Todd, 1999; Hertwig and Grüne-Yanoff, 2017). But which heuristics should be taught and how can we discover such heuristics? The theory of *resource-rationality* provides a mathematically precise definition of optimal heuristics (Lieder and Griffiths, 2019). In essence, the optimal heuristic for a decision-maker to use in a given environment achieves the best possible tradeoff between the expected utility of the resulting decision

and the expected opportunity cost of its execution. In the Mouselab-MDP paradigm, heuristics can be expressed as rules for deciding where to click given which information has already been revealed, when to stop clicking, and where to move given the information that has been uncovered. In previous work, we derived the optimal planning strategies for several Mouselab-MDP environments by solving metalevel MDPs using backward induction (Callaway et al., 2018).

In particular, Callaway et al. (2018) found that the resource-rational planning strategy for the environment described in Section 2 is to first set a goal by evaluating potential final destinations. As soon as inspecting a potential final destination uncovers the highest possible reward (+48), the optimal strategy selects the path that leads to it and terminates planning. If all potential final destinations have been inspected and one was revealed to be better than all the others, then the optimal strategy immediately decides to go there; else the optimal strategy inspects additional nodes located immediately before those most promising final destination and then chooses the path that is most promising according to the revealed information (and stops planning as soon as one path is revealed to be as good as possible). Having discovered this optimal planning strategy, we now present a cognitive tutor that teaches it to people.

4 Countering present bias with cognitive tutoring

We developed a cognitive tutor that teaches cognitive strategies by giving people metacognitive feedback. Our tutor's pedagogy is based on findings suggesting that people learn how to decide at least partly from the rewards and punishments they experience as a consequence of their decisions (Krueger et al., 2017; Lieder and Griffiths, 2017). This evidence for *metacognitive reinforcement learning* suggests that it should be possible to apply methods that have been developed to accelerate model-free reinforcement learning in robots– such as reward shaping (Ng et al., 1999)– to accelerate metacognitive learning in people. Following this line of reasoning, we used the following reward shaping method to generate optimal feedback signals for accelerating the process by which people learn how to make better decisions:

1. Model the cognitive function to be improved (i.e., planning) and the available cognitive operations (e.g., simulating the outcome of taking a certain action in a certain state) and their costs as a metalevel MDP.
2. Compute the values of the computations people might perform in different states (i.e., $Q_{\text{meta}}(b,c)$) by solving the metalevel MDP.
3. Let people practice the cognitive function to be improved and infer their computations from process tracing data.
4. Score people's inferred computations by

$$\text{score}(b,c) = \hat{Q}_{\text{meta}}(b,c) - \max_c \hat{Q}_{\text{meta}}(b,c). \tag{1}$$

5. Translate score into reinforcement and a feedback message.

We completed Step 1 and Step 2 in previous work (Callaway et al., 2018). Step 3 is accomplished by using the Mouselab-MDP paradigm to measure people's planning operations. Finally, the feedback signal computed in Step 4 is translated into a delay penalty of $2 - \text{score}$ seconds if the participant made an error or 0 seconds if their planning operation was optimal. The resulting feedback is given immediately after each click.

The cognitive tutor shown in Figure 1a integrates this feedback mechanism into the Mouselab-MDP paradigm and augments it with demonstrations of the optimal strategy described in Section 3. That is, the tutor's feedback has two components: i) a delay penalty whose duration communicates how sub-optimal the participant's planning operation was, and ii) a hint about what the optimal strategy would have done differently. Concretely, if the tutee makes an error then the planning operation that the optimal strategy would have performed instead is highlighted in blue (see Figure 1a). By contrast, when participants respond correctly, then they are told that they did a good job and can move on to the next click immediately.

5 Results

We evaluated our cognitive tutor in four online experiments that were run on Amazon Mechanical Turk. For each of these experiments we recruited about 50 participants per condition. Participants were paid performance-based bonuses.

In the control condition of each experiment, participants solved 31 different 3-step sequential decision problems in a plain version of the Mousleab-MDP paradigm shown in Figure 1a. Inspecting the recorded click sequences revealed that 38% of the participants in the control condition exhibited the present bias on the first trial. We identified three distinct planning strategies that gave rise to the present bias: i) a myopic satisficing strategy that inspects the immediate outcomes of alternative actions until it encounters a positive outcome and then immediately chooses the corresponding action, ii) a myopic maximizing strategy that inspects each action's immediate outcomes and then chooses the action with the best immediate outcome, and iii) an overly frugal myopic strategy that inspects only a single immediate outcome and nothing else.

Experiments 1-3 employed a between-subjects pre-post design comparing the effects of practicing the task with versus without the cognitive tutor. In Experiment 1, the pre-test, training, and post-test blocks all employed the same 3-step planning task shown in

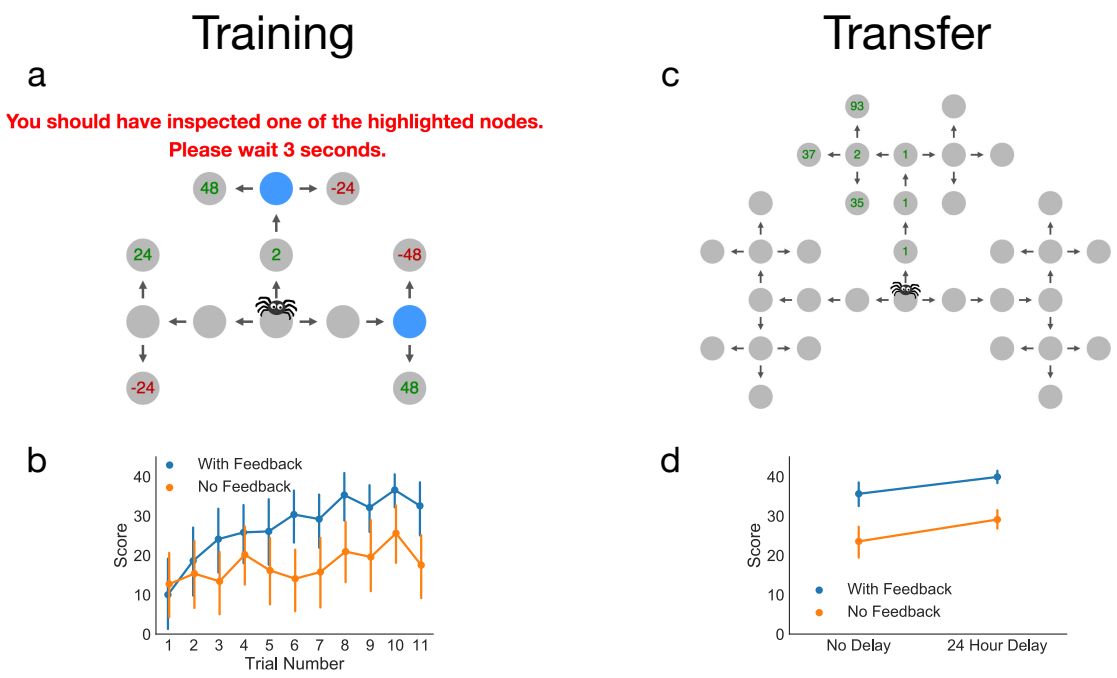

Figure 1: (a) Example feedback from the cognitive tutor in the training phase. (b) Participants learn to achieve high scores faster with the tutor's feedback. (c) A more difficult transfer problem. Feedback is not given in either condition. (d) Participants who received feedback in the training phase outperform control participants when tested immediately or after a 24 hour delay.

Figure 1a. We found that the tutor's metacognitive feedback significantly improved our participants' learning (see Figure 1b) and led to significantly higher post-test performance (36.2 \$/trial vs. 24.6 \$/trial, $t(2258) = 10.7$, $p < 0.0001$).

To elucidate how these improvements in performance were accomplished, we inspected how the prevalence of different types of strategies changed over time in the presence versus absence of optimal metacognitive feedback. We found that, over time, participants in the control condition slowly developed more adaptive planning strategies. In this process the prevalence of near-optimal goal-setting strategies increased from only 0.0% in the first trial to 26.4% after 30 trials ($\chi^2(1) = 16.1$, $p < .0001$). Conversely, the prevalence of the sub-optimal strategies giving rise to present bias dropped from 37.7% to 9.4% ($\chi^2(1) = 11.8$, $p = .0006$) and the prevalence of acting impulsively without any planning decreased from 33.9% to 26.4% ($\chi^2(1) = 0.72$, $p = .3974$). As shown in Figure 2, our tutor's optimal feedback amplified both of these changes, thereby increasing the the prevalence of near-optimal goal setting from 0.0% to 50.9% ($\chi^2(1) = 34.9$, $p < .0001$) while decreasing the prevalence of short-sighted decision strategies from 45.1% to 0.0% ($\chi^2(1) = 29.7$, $p < .0001$) and the prevalence of impulsive choices from 29.4% to 5.9% ($\chi^2(1) = 9.7$, $p = .0018$). Critically, we found that the feedback of our cognitive tutor significantly increased the proportion of people who discovered the far-sighted goal setting strategy from 26.4% to 50.9% ($\chi^2(1) = 6.63$, $p = .0100$) and significantly decreased the eventual prevalence of the maladaptive short-sighted strategies from 9.4% to 0.0% ($\chi^2(1) = 5.05$, $p = .0246$) and the prevalence of the maladaptive impulsive strategy from 26.4% to 5.9% ($\chi^2(1) = 8.01$, $p = .0046$). Furthermore, the learning curves shown in Figure 2 suggest that our tutor's feedback accelerated this transition.

In Experiment 2, the training block used a flight planning task that was structurally equivalent to task used in Experiment 1, whereas the pre-test and post-test blocks used the transfer task shown in Figure 1c). The training and the transfer task were structurally similar in that the variance of the reward distribution increased from each step to the next – being smallest for the immediate rewards and largest for the rewards attainable in the final step. But the transfer task used a different cover story (moving a money loving spider through a web of cash vs. flight planning), was more complex (5-step planning vs. 3-step planning) and involved a larger number of possible payoffs (192 vs. 10). As shown in Figure 1d, we found significant transfer effects from the relatively simple 3-step training task to the more difficult 5-step transfer task. Participants who had practiced with the cognitive tutor performed significantly better on the transfer task than participants who had practiced planning without the assistance of our cognitive tutor (37.4 \$/trial vs. 27.4 \$/trial, $t(2358) = 8.8$; $p < .0001$). This transfer effect appears to be partially mediated by people learning to plan backward – which was also beneficial in the transfer task. Concretely, participants who had practiced with the cognitive tutor were more likely to start planning by inspecting one of the final destinations than the control group (91.4% vs. 83.1%, $Z = 3.43$, $p = .0006$) and were less likely to start by inspecting one of the rewards in the first step (2.21% vs. 14.8%, $Z = -6.33$, $p < 0.0001$). In Experiment 3, we found that the transfer effect was after a delay of approximately 24 hours (39.9 \$/trial vs. 39.1 \$/trial, $t(1578) = 7.8$; $p < .0001$).

Experiment 4 compared the effectiveness of instruction plus practice with the cognitive tutor versus pure instruction and instruction plus watching a video demonstration of the optimal strategy. Participants in all three conditions read about the goal-setting principle for better decision-making discovered by Callaway et al. (2018). In the experimental conditions participants subsequently practiced applying the goal-setting principle with the cognitive tutor or saw a video demonstration of the optimal strategy. After 24 hours all participants were tested on the transfer task (Figure 1c). We found that participants who had practiced with the cognitive tutor performed significantly better on the transfer task than participants who were only told about the principle (38.0 \$/trial vs. 24.2 \$/trial, $t(83) = 10.5, p = 0.0000$). Participants who had seen a demonstration of the optimal strategy performed at the same level as participants who had practiced with the cognitive tutor (38.8 \$/trial vs. 38.0 \$/trial, $t(78) = -0.7, p = 0.49$). In either case, our cognitive tutor significantly improved people's planning by teaching them the resource-rational strategy we derived in previous work.

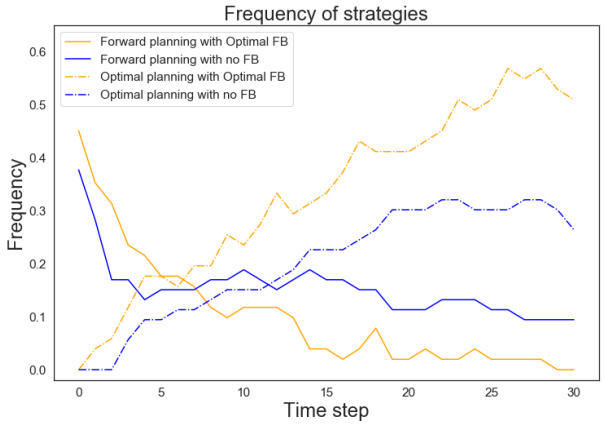

Figure 2: Prevalence of short-sighted versus far-sighted planning strategies for participants practicing with versus without out cognitive tutor.

6 Discussion

The present work illustrates how artificial intelligence can be leveraged to discover and teach rational decision-making strategies. The theoretical framework of resource-rationality allowed us to derive near-optimal planning strategies automatically, and the theory of metacognitive reinforcement learning allowed us to develop an intelligent system that can teach those rational heuristics very effectively. Our preliminary results suggest that practice with our cognitive tutor is more effective than instruction and has transferable benefits that are retained over time. Concretely, we found that participants who practiced decision-making with our cognitive tutor were significantly more likely to overcome the present bias by learning more far-sighted decision strategies than participants who practiced the same task without the tutor. This suggests that combining automatic strategy discovery with intelligent tutors is a promising approach to enhancing human rationality.

Future work will investigate how diagnostic our paradigm is of people's propensity to succumb to the present bias in everyday life and whether our approach to enhancing human rationality can be used to improve the decisions that people make in the real world.

References

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., and Griffiths, T. L. (2018). A resource-rational analysis of human planning. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, Austin, TX. Cognitive Science Society.

Callaway, F., Lieder, F., Krueger, P. M., and 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, Ann Arbor, MI*.

Gigerenzer, G. and Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press, New York, NY.

Hertwig, R. and Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6):973–986.

Krueger, P. M., Lieder, F., and 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*. Cognitive Science Society.

Lieder, F. and Griffiths, T. (2017). Strategy selection as rational metareasoning. *Psychological Review*, 124:762 –794.

Lieder, F. and Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*.

Lieder, F., Krueger, P. M., and Griffiths, T. L. (2017). An automatic method for discovering rational heuristics for risky choice. In Gunzelmann, G., Howes, A., Tenbrink, T., and Davelaar, E. J., editors, *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, pages 742–747, Austin, TX. Cognitive Science Society.

Ng, A. Y., Harada, D., and Russell, S. (1999). Policy invariance under reward transformations: Theory and application to reward shaping. In Bratko, I. and Dzeroski, S., editors, *Proceedings of the 16th Annual International Conference on Machine Learning*, pages 278–287, San Francisco, CA. Morgan Kaufmann.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

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

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.