Helping people make better decisions using optimal gamification

Falk Lieder (falk.lieder@berkeley.edu)
Helen Wills Neuroscience Institute, University of California at Berkeley, CA, USA

Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California at Berkeley, CA, USA

Abstract

Game elements like points and levels are a popular tool to nudge and engage students and customers. Yet, no theory can tell us which incentive structures work and how to design them. Here we connect the practice of gamification to the theory of reward shaping in reinforcement learning. We leverage this connection to develop a method for designing effective incentive structures and delineating when gamification will succeed from when it will fail. We evaluate our method in two behavioral experiments. The results of the first experiment demonstrate that incentive structures designed by our method help people make better, less short-sighted decisions and avoid the pitfalls of less principled approaches. The results of the second experiment illustrate that such incentive structures can be effectively implemented using game elements like points and badges. These results suggest that our method provides a principled way to leverage gamification to help people make better decisions.

Keywords: Gamification; Decision-Making; Bounded Rationality; Reinforcement Learning; Decision-Support

Introduction

Most decisions have both immediate and delayed consequences. For instance, investing in a pension plan entails less fun today than buying a wide-screen TV but a higher standard of living 25 years later. Unfortunately, the immediate outcomes often dominate people's considerations because future outcomes are discounted disproportionately (Berns, Laibson, & Loewenstein, 2007). One of the reasons for this *myopia* is that optimal long-term planning is often intractable because the number of possible scenarios grows exponentially as you look ahead further. Consequently, people have to rely on fallible heuristics to limit the length and number of scenarios they consider (Huys et al., 2015). While these heuristics can make us short-sighted in some situations (Huys et al., 2012), they work very well in others (Gigerenzer, 2008). Thus, perhaps, it is not our heuristics that are broken but the decision environments in which they fail (Gigerenzer, 2008; McGonigal, 2011).

The myopic nature of human decision-making (Berns et al., 2007) suggests that decision environments can be repaired by aligning each action's immediate rewards with the value of its long-term consequences. This could be achieved through *gamification* (Deterding, Dixon, Khaled, & Nacke, 2011). Gamification is the use of game elements such as points, levels, and badges in non-game contexts like education, the work place, health, and business. This approach has become extremely popular in the past five years. It is now widely used to engage people and nudge their decisions (Hamari, Koivisto, & Sarsa, 2014), and it has also inspired tools that help people achieve their goals and improve themselves (McGonigal, 2015; Kamb, 2016; Henry, 2014).

While gamification can have positive effects on motivation, engagement, behavior, and learning outcomes (Hamari et al., 2014), it can also have unintended negative consequences (Callan, Bauer, & Landers, 2015; Devers & Gurung, 2015). Unfortunately, it is currently impossible to predict whether gamification will succeed or fail (Hamari et al., 2014; Devers & Gurung, 2015), and there is no principled way to determine exactly how many points should be awarded for a given action. Here we address these problems by connecting the practice of gamification to the theory of pseudo-rewards in reinforcement learning. We leverage this connection to offer a mathematical framework for gamification and a computational method for designing optimal incentive structures. Our method offloads the computations of long-term planning from people by building the optimal decision into the incentive structure so that foresight is no longer necessary. This helps people make better decisions that are less short-sighted.

The plan for this paper is as follows: We start by introducing the theory of pseudo-rewards from reinforcement learning. We then apply this theory to derive a method for designing optimal incentive structures. Finally, we test the effectiveness of our method in two behavioral experiments. We close with implications for decision support and gamification.

Formalizing Gamification

Each sequential decision problem can be modeled as a *Markov Decision Process* (MDP)

$$M = (\mathcal{S}, \mathcal{A}, T, \gamma, r, P_0), \tag{1}$$

where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is the set of actions, $T(s,a,s')$ is the probability that the agent will transition from state s to state s' if it takes action a, $0 \leq \gamma \leq 1$ is the discount factor, $r(s,a,s')$ is the reward generated by this transition, and P_0 is the probability distribution of the initial state S_0 (Sutton & Barto, 1998). A *policy* $\pi : \mathcal{S} \mapsto \mathcal{A}$ specifies which action to take in each of the states. The expected sum of discounted rewards that a policy π will generate in the MDP M starting from a state s is known as its *value function*

$$V_M^\pi(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \cdot r(S_t, \pi(S_t), S_{t+1})\right]. \tag{2}$$

The optimal policy $\pi^\star_M$ maximizes the expected sum of discounted rewards, that is

$$\pi^\star_M = \arg\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \cdot r(S_t, \pi(S_t), S_{t+1})\right], \tag{3}$$

and its value function satisfies the Bellman equation

$$V^\star_M(s_t) = \max_a \mathbb{E}\left[r(s_t, a, S_{t+1}) + \gamma \cdot V^\star_M(S_{t+1})\right]. \tag{4}$$

We can therefore rewrite the optimal policy as

$$\pi^\star_M(s) = \arg\max_a \mathbb{E}\left[r(s_t, a, S_{t+1}) + \gamma \cdot V^\star_M(S_{t+1})\right], \tag{5}$$

which reveals that it is myopic with respect to the sum of the immediate reward and the discounted value of the next state.

Here, we leverage the MDP framework to model game elements like points and badges as *pseudo-rewards* $f(s,a,s')$ that are added to the reward function $r(s,a,s')$ of a decision environment M to create a modified environment $M' = (\mathcal{S}, \mathcal{A}, T, \gamma, r', P_0)$ with a more benign reward function $r'(s,a,s') = r(s,a,s') + f(s,a,s')$. From this perspective, the problem with misaligned incentives is that they change the optimal policy $\pi^\star_M$ of the original decision problem M into a different policy $\pi^\star_{M'}$ that is optimal for the gamified environment M' but not for the original environment M. To avoid this problem we have to ensure that each optimal policy of M' is also an optimal policy of M.

Fortunately, research in reinforcement learning has identified the necessary and sufficient conditions pseudo-rewards have to satisfy to achieve this: according to the *shaping theorem* (Ng, Harada, & Russell, 1999) adding pseudo-rewards retains the optimal policies of any original MDP if and only if the pseudo reward function f is *potential-based*, that is if there exists a *potential function* $\Phi : \mathcal{S} \mapsto \mathbb{R}$ such that

$$f(s,a,s') = \gamma \cdot \Phi(s') - \Phi(s), \tag{6}$$

for all states s, actions a, and successor states s'.

Pseudo-rewards can be shifted and scaled without changing the optimal policy, because linear transformations of potential-based pseudo-rewards are also potential-based:

$$a \cdot f(s,a,s') + b = \gamma \cdot \Phi'(s') - \Phi'(s), \tag{7}$$

$$\text{for } \Phi'(s) = a \cdot \Phi(s) - \frac{b}{1-\gamma}. \tag{8}$$

Shifting all pseudo-rewards $f(s,a,s')$ by the rewards $r(s,a,s')$ also retains the optimal policy, because it is equivalent to multiplying all rewards by 2 and positive linear transformations of the rewards preserve the optimal policy (Ng et al., 1999).

If gamification is to help people achieve their goals, then the pseudo-rewards added in the form of points or badges must *not* divert people from the best course of action but make its path easier to follow. Otherwise gamification would lead people astray instead of guiding them to their goals. Hence, the practical significance of the shaping theorem is that it gives the architects of incentive structures a method to rule out incentivizing counter-productive behaviors:

1. Model the decision environment as a MDP.
2. Define a potential function Φ that specifies the value of each state of the MDP.
3. Assign points according to Equation 6.

This method may be useful to avoid some of the dark sides of gamification (Callan et al., 2015; Devers & Gurung, 2015). To make this proposal more concrete and actionable, the next section presents a method for designing good potential functions.

Designing Optimal Incentive Structures

While the shaping theorem constrains pseudo-rewards to be potential-based there are infinitely many potential functions that one could choose. Given that people's cognitive limitations prevent them from fully incorporating distant rewards (Huys et al., 2012; Berns et al., 2007), the modified reward structure $r'(s,a,s')$ should be such that the best action yields the highest immediate reward, that is

$$\pi^\star_M(s) = \arg\max_a r'(s,a,s'). \tag{9}$$

Here, we show that this can be achieved with our method by setting the potential function Φ to the optimal value function $V^\star_M$ of the decision environment M, that is

$$\Phi^\star(s) = V^\star_M(s) = \max_\pi V^\pi_M(s). \tag{10}$$

First, note that the resulting pseudo-rewards are

$$f(s,a,s') = \gamma \cdot V^\star_M(s') - V^\star_M(s), \tag{11}$$

which leads to the modified reward function

$$r'(s,a,s') = r(s,a,s') + \gamma \cdot V^\star_M(s') - V^\star_M(s). \tag{12}$$

Hence, if the agent was myopic its policy would be

$$\begin{aligned} \pi(s) &= \arg\max_a \mathbb{E}\left[r(s,a,s') + \gamma \cdot V^\star_M(s') - V^\star_M(s)\right] \\ &= \arg\max_a \mathbb{E}\left[r(s,a,s') + \gamma \cdot V^\star_M(s')\right]. \end{aligned} \tag{13}$$

According to Equation 5, this is the optimal policy $\pi^\star_M$ for the original decision environment M. Thus, people would act optimally even if they were completely myopic. And they should perform equally well, if they do optimal long-term planning to fully exploit the gamified environment M' or learn its optimal policy $\pi^\star_{M'}$ through trial-and-error, because the shaping theorem (Eq. 6) guarantees that the gamified environment M' has the same optimal policy, that is $\pi^\star_{M'} = \pi^\star_M$.

This suggests that potential-based pseudo-rewards derived from $V^\star_M$ should allow even the most short-sighted agent who considers only the immediate reward to perform optimally. In this sense, the pseudo-rewards defined in Equation 11 can be considered optimal. In addition, optimal pseudo-rewards accelerate learning when the agent's initial estimate of the value function is close to 0 (Ng et al., 1999).

Computing the optimal pseudo-rewards requires perfect knowledge of the decision environment and the decision-maker's preferences. This information may be unavailable in practice. Yet, even when the optimal value function $V_M^\star$ cannot be computed, it is often possible to approximate it. If so, the approximate value function $\hat{V}_M$ can be used to approximate the optimal pseudo-rewards (Eq. 11) by

$$\hat{f}(s,a,s') = \gamma \cdot \hat{V}_M(s') - \hat{V}_M(s). \tag{14}$$

For instance, you can estimate the value of a state s from its approximate distance to the goal $s^\star$:

$$\hat{V}_M(s) = \hat{V}_M(s^\star) \cdot \left(1 - \frac{\text{distance}(s,s^\star)}{\max_s \text{distance}(s,s^\star)}\right), \tag{15}$$

where $\hat{V}_M(s^\star)$ is the estimated value of achieving the goal. Based on previous simulations (Ng et al., 1999), we predict that approximate pseudo-rewards (Eq. 14) can have beneficial effects similar to those of optimal pseudo-rewards but weaker. We tested these predictions in two behavioral experiments.

Experiment 1: Modifying Rewards

Methods

We recruited 250 adult participants on Amazon Mechanical Turk. Participants were paid \$0.50 for playing the game shown in Figure 1. In this game, the player receives points for routing an airplane along profitable routes between six cities. In each of the 24 trials the initial location of the airplane was chosen uniformly at random, and the task was to earn as many points as possible. Participants were incentivized by a performance dependent bonus of up to \$2. This game is based on the planning task developed by Huys et al. (2012). Our version of this task is isomorphic to a MDP with six states, two actions, deterministic transitions, and a discount factor of $\gamma = 1 - 1/6$. The locations correspond to the states of the MDP, the two actions correspond to flying to the first or the second destination available from the current location, the routes correspond to state-transitions, and the points participants received for flying those routes are the rewards. The current state was indicated by the position of the aircraft and was updated according to the flight chosen by the participant. The number of points collected in the current trial was shown in the upper right corner of the screen. After each choice there was a 1 in 6 chance that the game would end and the experiment would advance to the next trial, and a 5 in 6 chance that the participant could choose another flight. The participants were instructed to score as high as possible, and their financial bonus was proportional to the rank of their score relative to the scores of all participants in the same condition. The optimal policy in this MDP is to move counter-clockwise around the circle in all states except *Williamsville* and *Brownsville* (see Figure 1). Importantly, at *Williamsville* the optimal policy incurs a large immediate loss, and no other policy achieves a positive reward rate.

Participants were randomly assigned to one of four conditions: In the control group, there were no pseudo-rewards

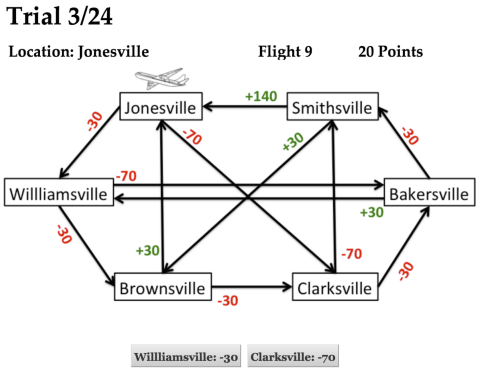

Figure 1: Interface of the control condition of Experiment 1. The map shows the unmodified rewards r.

Pseudo-Rewards	Smiths-		Jones-		Williams-		Browns-		Clarks-		Bakers-	
None	140	30	−30	−70	−30	−70	−30	30	−30	−70	−30	−70
Optimal	2	−76	2	−5	−12	2	−4	2	2	0	2	−42
Approximate	8	−102	−22	−4	−22	−4	36	38	−34	−16	24	−32
Non-Potential-Based	119	9	−51	−41	−51	−41	−1	9	−51	41	−1	9

Table 1: Rewards in Experiment 1. The first entry of each cell is the (modified) reward of the counter-clockwise move and the second one is the (modified) reward of the other move.

(Figure 1). In this condition finding the optimal path required planning 4 steps ahead. In the three experimental conditions the rewards were modified by adding pseudo-rewards. To keep the average reward constant, the pseudo-rewards were mean-centered by subtracting their average; since mean-centering is a linear transformation this retained the guarantees of the shaping theorem (see Eq. 7). Next, the mean-centered pseudo-rewards were added to the rewards of the control condition (see Figure 1) yielding the modified rewards shown in Table 1, and the flight map was updated accordingly. In all other respects the interfaces of the experimental conditions were exactly the same as the interface of the control condition (see Figure 1). All participants were thus unaware of the existence of pseudo-rewards. In the first experimental condition the pseudo-rewards were derived from the optimal value function according to the shaping theorem (Eq. 11). In this condition, looking only 1 step ahead was sufficient to find the optimal path. The second experimental condition used the approximate potential-based pseudo-rewards defined in Equation 14 with the distance-based heuristic value function defined in Equation 15 where $s^\star$ was *Smithsville*, $\Phi(s^\star)$ was its highest immediate reward (i.e., $+140$), and $\text{distance}(a,b)$ was the minimum number of actions needed to get from state a to state b. The resulting pseudo-rewards simplified planning but not as much as the optimal pseudo-rewards. Finding the optimal path required planning 2-3 steps ahead and the immediate losses were smaller. In the third experimental condition, the pseudo-rewards violated the shaping theorem: the pseudo-reward was $+50$ for each transition that reduced the distance to the most valuable state (i.e. *Smithsville*) but there was no penalty for moving away from it.

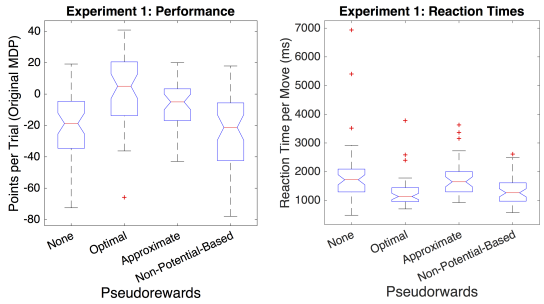

Figure 2: Performance and reaction times in Experiment 1.

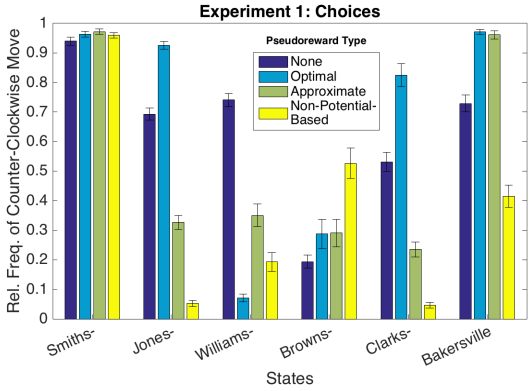

Figure 3: Choice frequencies in each state of Experiment 1 by condition. Error bars enclose 95% confidence intervals.

Results and Discussion

The average completion time of the experiment was 13:37 min. The median response time was 1.3 sec. per choice. We excluded 3 participants who invested less than one third of the median response time of their condition and 11 participants who scored lower than 95% of all participants in their condition. The boxplots in Figure 2 summarize the median performance and reaction times of participants in the four conditions. The median performer of the control group *lost* 18.75 points per trial. By contrast, the majority of the group with optimal pseudo-rewards achieved a net *gain* in the unmodified MDP (median performance: $+5.00$ points/trial). The median performance in the group with approximate potential-based pseudo-rewards was -5.00 points per trial, and in the group with non-potential-based pseudo-rewards the median performance was -21.25 points per trial. A Kruskal-Wallis ANOVA revealed that the type of pseudo-rewards added to the reward function significantly affected people's performance in the original MDP ($H(3) = 40.35, p < 10^{-8}$) as well as their reaction times ($H(3) = 29.96, p < 10^{-5}$). Given that the pseudo-reward type had a significant effect, we performed pairwise Wilcoxon rank sum tests to compare the medians of the four conditions. The non-potential-based pseudo-rewards failed to improve people's performance ($Z = 0.72, p = 0.47$). By contrast, the approximate potential-based pseudo-rewards succeeded to improve their performance ($Z = 2.86, p = 0.0042$) and led to better performance than the heuristic pseudo-rewards that violated the shaping theorem ($Z = 3.61, p = 0.0003$). People performed even better when gamification was based on the optimal pseudo-rewards: adding optimal pseudo-rewards led to better decisions than adding the approximate potential-based pseudo-rewards ($Z = 2.68, p = 0.0074$), presenting the true reward structure ($Z = 4.76, p < 10^{-5}$), or adding the non-potential-based pseudo-rewards ($Z = 5.34, p < 10^{-7}$).

In addition, some pseudo-rewards accelerated the decision process (Figure 2): optimal pseudo-rewards decreased the median response time from 1.72 to 1.14 sec/decision ($Z = -4.19, p < 0.0001$), and non-potential-based pseudo-rewards decreased it to 1.12 sec/decision ($Z = -3.38, p = 0.0007$). But approximate potential-based pseudo-rewards had no significant effect on response time (1.65 sec/decision; $Z = -0.28, p = 0.78$).

Next, we inspected how the pseudo-rewards affected the choices our participants made in each of the six states (see Figure 3). The optimal pseudo-rewards significantly changed the choice frequencies in each of the six states and successfully nudged participants to follow the optimal cycle *Smithsville* $\rightarrow$ *Jonesville* $\rightarrow$ *Williamsville* $\rightarrow$ *Bakersville* $\rightarrow$ *Smithsville* (see Figure 1). Their strongest effect was to eliminate the problem that most people would avoid the large loss associated with the correct move from *Williamsville* to *Bakersville* ($\chi^2(2) = 1393.8, p < 10^{-15}$). The optimal pseudo-rewards also increased the frequency of all other correct choices along the optimal cycle, that is the decisions to fly from *Bakersville* to *Smithsville* ($\chi^2(2) = 326.5, p < 10^{-15}$), from *Smithsville* to *Jonesville* ($\chi^2(2) = 7.9, p = 0.0191$), and from *Jonesville* to *Williamsville* ($\chi^2(2) = 299.8, p < 10^{-15}$). In addition, the optimal pseudo-rewards increased the frequency of the correct move from *Clarksville* to *Bakersville* ($\chi^2(2) = 92.0, p < 10^{-15}$). The only negative effect of the optimal pseudo-rewards was to slightly increase the frequency of the suboptimal move from *Brownsville* to *Clarksville* ($\chi^2(2) = 13.2, p = 0.0013$). By contrast, the non-potential-based pseudo-rewards misled our participants to follow the unprofitable cycle *Jonesville* $\rightarrow$ *Clarksville* $\rightarrow$ *Smithsville* $\rightarrow$ *Jonesville* by raising the frequency of the reckless moves from *Jonesville* to *Clarksville* ($\chi^2(2) = 1578.6, p < 10^{-15}$) and from *Clarksville* to *Smithsville* ($\chi^2(2) = 813.7, p < 10^{-15}$). The effect of the approximate pseudo-rewards was beneficial in *Smithsville*, *Williamsville*, and *Bakersville*, but negative in *Jonesville*, *Brownsville*, and *Clarksville* (see Figure 3). This explains why only potential-based pseudo-rewards had a positive net-effect on performance (Figure 2).

Finally, we investigated learning effects by comparing the average choice frequencies in the first five trials versus the last five trials. While people's decisions improved with learning in the conditions with potential-based pseudo-rewards, learning had a negative effect when the pseudo-rewards violated the shaping theorem: In the condition with non-potential-based pseudo-rewards learning reduced the frequency of the correct choice in *Jonesville* ($\chi^2(2) = 9.22, p = 0.01$). By con-

trast, in the condition with optimal pseudo-rewards learning improved people's choices in *Smithsville* ($\chi^2(2) = 13.02, p = 0.0015$), and in the condition with approximate potential-based pseudo-rewards learning improved people's choices in *Jonesville* ($\chi^2(2) = 11.44, p = 0.0033$). In the control condition, learning made people more likely to take the correct action in *Williamsville* ($\chi^2(2) = 24.16, p < 0.0001$) and *Bakersville* ($\chi^2(2) = 22.74, p < 0.0001$) but less likely to take the correct action in *Clarksville* ($\chi^2(2) = 8.80, p = 0.0123$). In summary, while learning with potential-based pseudo-rewards guided people closer towards the optimal policy, non-potential-based pseudo-rewards lured them away from it. This is consistent with the shaping theorem's assertion that pseudo-rewards have to be potential-based to always retain the optimal policy.

In summary, we found that pseudo-rewards can help people make better decisions–but only when they are designed well. The results support the proposed method for designing incentive structures: Assigning pseudo-rewards according to the shaping theorem avoided the negative effects of non-potential-based pseudo-rewards. Furthermore, using the optimal value function as the shaping potential lead to the greatest improvement in decision-quality.

Experiment 2: Explicit Pseudo-Rewards

To assess the potential of gamification for decision-support in real life, Experiment 2 conveyed pseudo-rewards by points without monetary value.

Methods

We recruited 400 participants on Amazon Mechanical Turk. Participants were paid \$2.50 for about 20 min of work plus a performance dependent bonus of up to \$2 that averaged at \$1. We inspected the data from the 355 participants who had not participated in earlier versions of the experiment and excluded 19 participants who responded in less than one third of the median response time of all participants or performed worse than 95% of the participants in their condition.

The task from Experiment 1 was modified by adding stars in two new experimental conditions (see Figure 4). To make it easy for participants to add rewards plus pseudo-rewards, the rewards were scaled down by a factor of 10 and the optimal pseudo-rewards were recomputed according to Equation 11. The pseudo-rewards awarded for each action were then shifted by the expected return of the optimal policy (\$0.90) so that they predict how much money the player will earn in the long-run if they choose that action and act optimally afterwards. The experiment had four conditions: In the control condition no pseudo-rewards were presented. Three experimental conditions presented the optimal pseudo-rewards in three different formats: condition 2 embedded the pseudo-rewards into the reward structure as in Experiment 1; condition 3 presented them separately in the form of stars; and in condition 4 the number of stars communicated the sum of pseudo-reward plus immediate reward. All numbers were rounded to one significant digit. In the conditions with stars

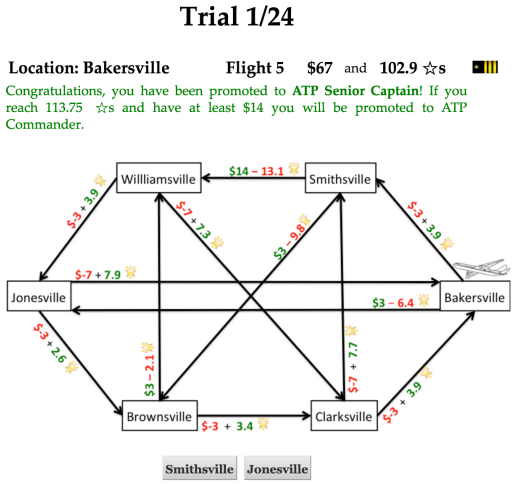

Figure 4: Screenshot of Experiment 2.

the instructions stated that the stars were designed to help the pilots make better decisions and explained their meaning. In addition, the instructions included the tip that the better flight is the one with the higher sum of stars plus dollars (condition 3) or that it is the one awarded more stars (condition 4). The stars had no monetary value, but they determined whether and when the character played by the participant would be promoted. The character could rise from *Trainee* to *ATP senior captain* via 15 intermediate levels. The number of points required to reach the next level increased according to the difficulty curve proposed by Bostan and Öğüt (2009). Participants were told how many stars and dollars were required to reach the next level (see Figure 4). The current score and the shoulder batch corresponding to the current level were shown at the top of the screen, and a feedback message appeared whenever the character was promoted. The player started the game with +\$50 so that their balance would remain positive as they learned to play the game. In all conditions a quiz ensured the participants understood the task and its incentives before they could start the game. This quiz comprised three questions on how the participant's financial bonus would be determined and three questions testing they understood the mechanics of the task.

Results and Discussion

Overall, presenting pseudo-rewards in one of the three formats significantly improved people's performance ($Z = 3.43$, $p = 0.0006$). Most importantly, adding points (i.e., stars) without monetary value can be just as effective as directly modifying the reward structure of the environment: Integrated pseudo-rewards significantly increased people's performance from -0.73 dollars/trial to $+0.17$ dollars/trial ($Z = 3.69, p = 0.0002$). The resulting level of performance was not significantly different from the performance in the condition with embedded pseudo-rewards ($+0.42$ dollars/trial, $Z = 0.52, p = 0.62$). By contrast, presenting pseudo-rewards separately failed to significantly increase people's performance

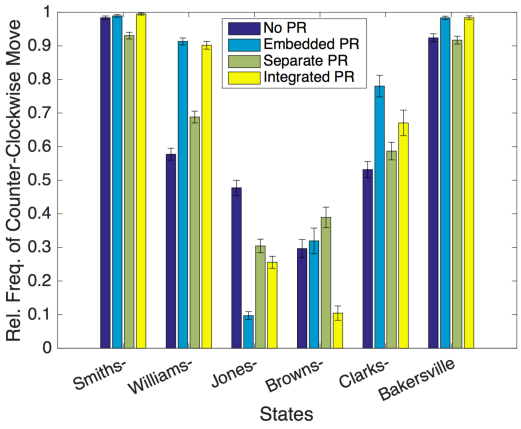

Figure 5: Choice Frequencies in Experiment 2: Effects of stars and badges on performance

(median performance: -0.5 dollars/trial; $Z = 0.22, p = 0.83$).

Inspecting the choice frequencies (Figure 5) confirmed that the three presentation formats had significantly different effects: Embedded pseudo-rewards and integrated pseudo-rewards were more beneficial than separately presented pseudo-rewards in all 6 states (all $p \leq 0.0218$). Embedded pseudo-rewards were more beneficial than integrated pseudo-rewards in Jonesville and Clarksville (both $p \leq 0.0001$), but integrated pseudo-rewards were more beneficial than embedded pseudo-rewards in Brownsville ($p < 10^{-9}$). Furthermore, participants were significantly faster when pseudo-rewards were embedded in the decision environment than when they were presented separately ($Z = -4.06, p < 0.0001$) or in the integrated format ($Z = -2.78, p = 0.0053$).

In conclusion, adding points that convey the sum of optimal pseudo-rewards plus immediate reward can be as effective as changing the reward structure itself.

Conclusion

We have proposed a general method for improving incentive structures based on the theory of MDPs and the shaping theorem. Its basic idea is to offload the computation necessary for long-term planning into the reward structure of the environment such that people will act optimally even when they consider only immediate rewards. The results of Experiment 1 provide a proof of principle that our approach can help people make better sequential decisions. Our findings suggest that the shaping theorem can be used to delineate when gamification will succeed from when it will fail and to design incentive structures that avoid the perils of less-principled approaches to gamification. Experiment 2 illustrated that the incentive structures designed with our method can be implemented with game elements like points and badges. In both experiments the pseudo-rewards helped people overcome their short-sighted tendency to avoid an aversive action with desirable long-term consequences in favor of immediate reward–a cognitive limitation that can manifest in procrastination and impulsivity. Therefore, our method might be useful for improving inter-temporal choice. Our findings are consistent with the view that the limitations of human decision-making can be overcome by reshaping incentive structures that make us prone to fail into ones that our heuristics were designed for. Our method achieves this by solving people's planning problems for them and restructuring their incentives accordingly. The program providing the pseudo-rewards can be seen as a cognitive prosthesis because it compensates for people's cognitive limitations without restricting their freedom. In conclusion, optimal gamification may provide a principled way to help people achieve their goals and procrastinate less.

Acknowledgments. This work was supported by grant number ONR MURI N00014-13-1-0341. We thank Rika Antonova, Ellie Kon, Paul Krueger, Mike Pacer, Daniel Reichman, Stuart Russell, Jordan Suchow, and the CoCoSci lab for feedback and discussions.

References

Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice–toward an integrative framework. *Trends in Cognitive Sciences*, *11*(11), 482–488.

Bostan, B., & Öğüt, S. (2009). Game challenges and difficulty levels: lessons learned from RPGs. In G. K. Yeo & Y. Cai (Eds.), *Learn to Game, Game to Learn: Proceedings of the 40th Conference of the International Simulation and Gaming Association.*

Callan, R. C., Bauer, K. N., & Landers, R. N. (2015). How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 553–568). Cham: Springer.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15).

Devers, C. J., & Gurung, R. A. (2015). Critical perspective on gamification in education. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 417–430). Cham: Springer.

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford: Oxford University Press.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?–A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 3025–3034).

Henry, A. (2014). *The best tools to (productively) gamify every aspect of your life.* Retrieved from `http://lifehacker.com/the-best-tools-to-productively-gamify-every-aspect-of-1531404316`

Huys, Q. J. M., 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.

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098–3103.

Kamb, S. (2016). *Level up your life: How to unlock adventure and happiness by becoming the hero of your own story.* Emmaus: Rodale Books.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world.* New York: Penguin.

McGonigal, J. (2015). *SuperBetter: A revolutionary approach to getting stronger, happier, braver and more resilient–powered by the science of games.* London, UK: Penguin Press.

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

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA, USA: MIT press.