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, levels, and badges have become an extremely popular tool to motivate, nudge, and engage people in education, business, and health. Yet, there is no theoretical account of when such *gamification* works, or how it should be designed. 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 the effectiveness of our method in two behavioral experiments. The results of the first experiment demonstrate that incentive structures designed by our method can indeed 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 the proposed 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

Many decisions require foresight, but optimal long-term planning is often intractable because the number of possible scenarios grows exponentially as you look ahead further. Consequently, decision-makers have to rely on fallible heuristics to limit the length and the number of scenarios they consider (Huys et al., 2015). When the best course of action incurs a large loss in the short-term, then these heuristics tend to favor sub-optimal actions that avoid the loss temporarily (Huys et al., 2012). This failure might manifest in procrastination (Steel, 2007) and other self-control problems. Yet, the same heuristics perform very well in decision problems that match the structure of the environment they are adapted to (Gigerenzer, 2008), for instance when actions that are good in the short run are also good in the long run.

The apparently myopic nature of human decision-making suggests that it may be possible to help people make better decisions by aligning each action's immediate rewards with the value of its long-term consequences. This could be achieved by improving our lives' reward structure through *gamification* (McGonigal, 2011). Gamification is the use of game elements such as points, levels, and badges in a non-game context (Deterding, Dixon, Khaled, & Nacke, 2011). These game elements are widely used to engage people and nudge their decisions in education, the work place, health, and business (Hamari, Koivisto, & Sarsa, 2014). Gamification has become very popular in the past five years and has inspired tools helping 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 solution to the decision problem into the incentive structure such that people will act optimally even when they think only a single step ahead. This helps people make better decisions that are less short-sighted.

The plan for this paper is as follows: We first introduce 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 by discussing 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 $\pi_M^{\star}$ maximizes the expected sum of discounted rewards,

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

and its value function satisfies the Bellman equation

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

We can therefore rewrite the optimal policy as

$$\pi_M^\star(s) = \arg\max_a \mathbb{E}\left[r(s_t,a,S_{t+1}) + \gamma \cdot V_M^\star(S_{t+1})\right], \quad (5)$$

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 framework of MDPs to model game elements like points and badges as *pseudo-rewards* $f(s,a,s')$ that are added to the MDPs reward function $r(s,a,s')$ to create a modified MDP $M' = (\mathcal{S},\mathcal{A},T,\gamma,r',P_0)$ 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 $\pi_M^\star$ of the original decision problem M into a different policy $\pi_{M'}^\star$ that is optimal for the gamified version M' but not for the original problem 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 that 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

$$f(s,a,s') = \gamma \cdot \Phi(s') - \Phi(s), \quad (6)$$

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:

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

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

If gamification is to help people achieve their goals, then the rewards added in the form of points or badges must *not* divert the best of course of action but reinforce it to make that path easier to follow. Otherwise gamification can 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 incentives structures a method to rule out incentivizing counter-productive behaviors:

1. Model the decision problem 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 precise, the next section presents a method for creating 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; Myerson & Green, 1995), the modified reward structure $r'(s,a,s')$ should be such that the best action yields the highest immediate reward, that is

$$\pi_M^\star(s) = \arg\max_a r'(s,a,s') \quad (9)$$

Here we show that this can be achieved by deriving the pseudo-rewards from the potential function

$$\Phi^\star(s) = V_M^\star(s) = \max_\pi V_M^\pi(s). \quad (10)$$

First, note that the resulting pseudo-rewards are

$$f(s,a,s') = \gamma \cdot V_M^\star(s') - V_M^\star(s), \quad (11)$$

which leads to the modified reward function

$$r'(s,a,s') = r(s,a,s') + \gamma \cdot V_M^\star(s') - V_M^\star(s). \quad (12)$$

Hence, if the agent was myopic its policy would be

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

According to Equation 5, this is the optimal policy $\pi_M^\star(s)$ of the original MDP M. Thus, if people were myopic the pseudo-rewards would make them act optimally. And even if people did optimal long-term planning in the modified MDP M' or learned its optimal solution $\pi_{M'}^\star$ through trial-and-error they would still act according to $\pi_M^\star$, because the shaping theorem (Eq. 6) guarantees that $\pi_{M'}^\star = \pi_M^\star$.

This suggests that potential-based pseudo-rewards derived from $V_M^\star$ should allow even the most short-sighted agent that only considers the immediate reward to perform optimally. In this sense, the pseudo-rewards defined in Equation 11 can be considered optimal. In addition, the optimal pseudo-rewards accelerate learning as long as 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 that may be unavailable in practical applications. 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

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

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

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

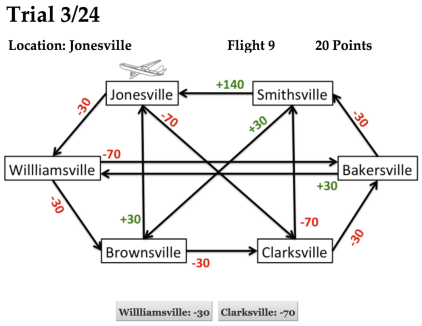

Figure 1: Interface of the control condition of Experiment 1. The map shows the unmodified rewards r.

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 \$0.5 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 was 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 take the counter-clockwise move around the circle in all states except *Williamsville* and *Brownsville*. 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). 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: the number of points participants received for taking action a in state s was changed to the modified rewards $r'(s,a,s') = r(s,a,s') + f(s,a,s')$ shown in Table 1, and the flight map was updated accordingly. This was the only change and participants were unaware of the pseudo-rewards. In the first experimental condition the optimal pseudo-rewards (Eq. 11) were added to the reward function. 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 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. After the pseudo-rewards had been determined in this way, they were mean-centered such that the average pseudo-reward was zero in all three experimental conditions. This transformation retained the guarantees of the shaping theorem as shown above (Eq. 7).

Pseudo-Rewards	Smiths-		Jones-		Williams-		Browns-		Clarks-		Bakersv.	
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.

Results and Discussion

The average completion time of the experiment was 13 min. and 37 sec. The median response time was 1.3 sec. per choice. We excluded 3 participants whose median response time was less than one third of the median response time across all subjects and 11 participants who scored lower than 95% of all participants in their group. 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

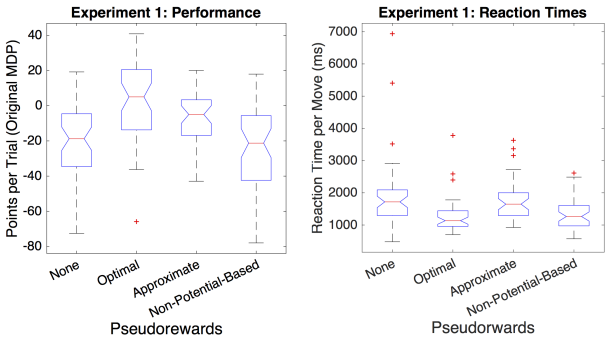

Figure 2: Performance and reaction times in Experiment 1 by condition.

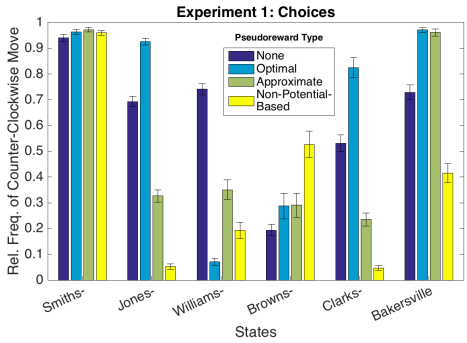

Figure 3: Choice frequencies in each state of Experiment 1 by condition. Error bars enclose 95% confidence intervals.

points/trial. A Kruskal-Wallis ANOVA revealed that the type of pseudo-rewards added to the reward function significantly affected the participants' performance in the original MDP ($H(3) = 40.35, p < 10^{-8}$) and 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 peoples performance ($Z = 0.72, p = 0.47$). By contrast, the approximate potential-based pseudo-rewards succeeded to improve people's 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 the gamification was based on the optimal pseudo-rewards instead of the approximate ones ($Z = 2.68, p = 0.0074$). Optimal pseudo-rewards led to better decisions than presenting the true reward structure ($Z = 4.76, p < 10^{-5}$), non-potential based pseudo-rewards ($Z = 5.34, p < 10^{-7}$), or approximate potential-based pseudo-rewards ($Z = 2.68, p = 0.0074$).

In addition, optimal and non-potential-based pseudo-rewards accelerated the decision process (Figure 2): optimal pseudo-rewards decreased the median response time from 1.72 to 1.14 sec. per decision ($Z = -4.19, p < 0.0001$), and non-potential-based pseudo-rewards decreased it to 1.12 sec. per decision ($Z = -3.38, p = 0.0007$). People in the condition with approximate potential-based pseudo-rewards took about the same amount of time as people in the condition without pseudo-rewards (1.65 sec.; $Z = -0.28, p = 0.78$).

Next, we inspected the effect of the pseudo-rewards on which actions our participants chose in each of the six states (see Figure 3). The optimal pseudo-rewards significantly altered our participants' choices in each of the six states. The strongest effect of optimal pseudo-rewards was to eliminate the problem that most people myopically 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 flying 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* ($\chi^2(2) = 299.8, p < 10^{-15}$). The optimal pseudo-rewards thereby increased people's propensity to follow the optimal cycle *Smithsville* $\rightarrow$ *Jonesville* $\rightarrow$ *Williamsville* $\rightarrow$ *Bakersville* $\rightarrow$ *Smithsville*. 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 participants to follow the unprofitable cycle *Jonesville* $\rightarrow$ *Clarksville* $\rightarrow$ *Smithsville* $\rightarrow$ *Jonesville* by significantly increasing the frequency of the reckless moves from *Jonesville* to *Clarksville* ($\chi^2(2) = 1578.6, p < 10^{-15}$) and from *Clarksville* to *Smithsville* ($\chi^2(2) = 813.7, p < 10^{-15}$). This explains why only potential-based pseudo-rewards had a positive net-effect on performance (Figure 2). The effect of the approximate pseudo-rewards was beneficial in *Smithsville*, *Williamsville*, and *Bakersville*, but negative in *Jonesville*, *Brownsville*, and *Clarksville* (see Figure 3).

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 contrast, 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 potential-based pseudo-rewards guided people closer towards the optimal policy, non-potential-based pseudo-rewards lured people away from it. This is consis-

tent 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 we observed for heuristic non-potential-based pseudo-rewards. Furthermore, using the optimal value function as the shaping potential lead to greatest improvement in decision-quality. In the next experiment we test whether similar improvements can be achieved when the pseudo-rewards are delivered through game elements like points and badges.

Experiment 2: Explicit Pseudo-Rewards

Methods

To assess the potential of gamification for decision-support in real life, Experiment 2 conveyed pseudo-rewards by stars with no monetary value. We modified Experiment 1 such that rewards and pseudo-rewards were presented separately: While rewards were presented as dollars earned by the pilot and determined the participants' financial bonus, pseudo-rewards were presented as stars and had no impact on their pay (see Figure 4). The number of stars collected by the player determined their character's badge. The number of stars collected and the current badge were shown at the top of the screen. A feedback message was shown whenever the character was promoted and earned a badge or was demoted and lost a badge. To further motivate our participants to attend to the pseudo-rewards we told them that the stars had been introduced to help pilots make better decisions. The instructions explained that the difference in the number of stars awarded for two alternative flights reflects the difference in the amount of money that the pilot can earn from their destinations in the long-run starting. In addition, participants were explicitly told that the flight with the higher sum of stars plus dollars is more profitable in the long run. To simplify the addition of rewards and pseudo-rewards we scaled down the rewards of Experiment 1 by a factor of 10. The pseudo-rewards were recomputed from the scaled rewards and shifted such that the smallest pseudo-reward was $+1$. The resulting reward structure is shown in Figure 4. A repeatable quiz enforced that participants understood the instructions before they could start the game. The quiz covered the difference between stars and dollars and the hint that flights with a higher sum of stars plus dollars were more profitable in the long run.

Because we had scaled down the rewards, we ran an additional control experiment that was equivalent to Experiment 1 except that the rewards were scaled down by a factor of 10. Participants were recruited on Amazon Mechanical Turk. We recruited 50 participants in the experimental condition and 51 participants in the control experiment. Participants in the experimental condition were paid \$2 for participation and could earn a performance dependent bonus of up to \$2. Participants in the control experiment were paid \$0.50 for participation and could earn a performance dependent bonus of up to \$2.

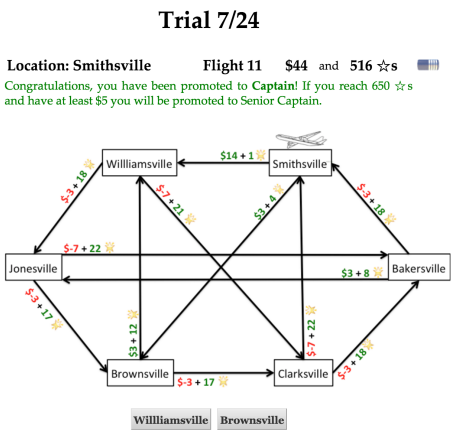

Figure 4: Screenshot from Experiment 2.

Results and Discussion

The average completion time was 23.9 min. The median response time was 1.9 sec. per choice. We excluded 16 out of 101 participants (16%), because they had either previously participated in similar experiments, responded in less than one third of their condition's median response time, or performed worse than 95% of the participants in their condition.

The addition of pseudo-rewards significantly improved people's performance from a median loss of -1.46 dollars per choice to a median loss of only -0.21 dollars per choice ($Z = 2.13, p = 0.03$). The median response time increased from 1.66 sec. per choice to 1.97 sec. per choice but this chance was not statistically significant ($Z = 1.15, p = 0.25$). Next, we analyzed the effect of the pseudo-rewards on participants' choice frequencies (see Figure 5). The addition of stars significantly increased the frequency of the correct at *Jonesville* to go to *Bakersville* from 38.4% to 66.6% despite the large immediate loss associated with this transition ($\chi^2(2) = 204.9, p < 10^{-15}$). The pseudo-rewards also increased people's propensity to go from *Bakersville* to *Smithsville* ($\chi^2(2) = 62.7, p < 10^{-13}$) and from *Williamsville* to *Jonesville* ($\chi^2(2) = 180.4, p < 10^{-15}$). The stars thereby successfully guided the players onto the optimal cycle (*Smithsville* $\rightarrow$ *Williamsville* $\rightarrow$ *Jonesville* $\rightarrow$ *Bakersville* $\rightarrow$ *Smithsville*). The only negative effect was that the pseudo-rewards increased the frequency of the move from *Clarksville* to *Smithsville* ($\chi^2(2) = 47.8, p < 10^{-10}$) but this move is only marginally worse than its alternative.

Next, we investigated learning (see Figure 5, right panel). We found that in the condition with pseudo-rewards learning allowed people to see past the immediate loss at the transition from *Jonesville* to *Bakersville* and increased the frequency of this transition from 57.5% to 75.7% ($\chi^2(2) = 17.9, p = 0.0001$). Learning also increased the frequency of the correct move from *Brownsville* to *Williamsville* ($\chi^2(2) = 25.27, p < 10^{-5}$) and from *Bakersville* to *Smithsville* ($\chi^2(2) = 26.56, p <$

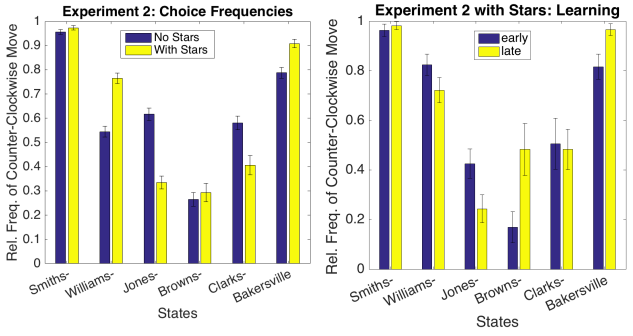

Figure 5: Results of Experiment 2: Effects of stars and badges on choice frequencies (left), and learning effects in the experimental condition (right).

10^{-5}). Only at *Williamsville* did learning decrease the frequency of the better choice ($\chi^2(2) = 8.9, p = 0.0114$).

Finally, we compared the effects observed in this experiment with the effects observed in Experiment 1. To adjust for having scaled down the rewards by a factor of 10, we divided the scores from Experiment 1 by 10. We found that while there was no significant difference in performance between the conditions without pseudo-rewards ($Z = 0.81, p = 0.42$), the performance in the condition with pseudo-rewards was lower when they were presented as stars (median reward: -0.21 dollar/trial) than when they were added directly onto the payoffs (median reward: $+0.5$ dollar/trial, $Z = -2.42, p = 0.0154$). Hence, introducing stars without monetary value was not quite as effective as changing the payoffs directly. Yet, it is remarkable that even pseudo-rewards that have no objective value can significantly improve people's decisions when they are designed and explained properly.

Conclusion

We have proposed a general method for improving incentive structures based on the theory of Markov decision processes 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 the immediate reward. The results of Experiment 1 provide a proof of principle that incentive structures designed with our method can indeed help people make better decisions. These findings demonstrate 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 effectively implemented with game elements like points and badges. In both experiments the pseudo-rewards helped people overcome their short-sighted tendency to avoid a correct but aversive action at the expense of its desirable long-term consequences in favor of immediate reward–a cognitive limitation that can manifest in procrastination and impulsivity. Our findings are consistent with the view that we can overcome the limitations of human decision-making by reshaping environments in which we fail into the ones 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

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 *Gamification in education and business* (pp. 553–568). 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 *Gamification in education and business* (pp. 417–430). 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 *47th Hawaii international conference on system sciences* (pp. 3025–3034).

Henry, A. (2014, February). *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., ONions, 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 Comput. Biol*, *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.

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the experimental analysis of behavior*, *64*(3), 263–276.

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* (Vol. 16, pp. 278–287). San Francisco, CA, USA: Morgan Kaufmann.

Steel, P. (2007). The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological bulletin*, *133*(1), 65.

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