# A Bounded Rationality Account of Wishful Thinking

**Rebecca Neumann (beckieneumann@gmail.com)**
Cognitive Science Program, University of California, Berkeley, CA 94720 USA

**Anna N. Rafferty (rafferty@cs.berkeley.edu)**
Computer Science Division, University of California, Berkeley, CA 94720 USA

**Thomas L. Griffiths (tom_griffiths@berkeley.edu)**
Department of Psychology, University of California, Berkeley, CA 94720 USA

**Abstract**

People tend towards wishful thinking, in which they overestimate the probability of favorable outcomes and underestimate the probability of unfavorable outcomes. Many explanations for this phenomenon focus on its irrationality. We explore whether wishful thinking could actually help people make better decisions given that they have limited cognitive resources. We consider a situation in which multiple decisions must be made over a period of time, where the consequences of these decisions are not fully determined. We model this situation as a Markov decision process, and incorporate limited cognitive resources by varying the amount of time in the future that the agent considers the consequences of its decisions. Through simulations, we show that with limited cognitive resources, this model can exhibit better performance by incorporating a bias towards wishful thinking. This advantage occurs across a range of decision-making environments, suggesting that the same effect could be applicable to many real life scenarios.

**Keywords:** rational process models; Markov decision processes

## Introduction

People tend to overestimate the probability that their preferred outcomes will occur and to underestimate the probability of non-preferred outcomes (e.g., Camerer & Lovallo, 1999; Larwood & Whittaker, 1977; Lyles & Thomas, 1988; Svenson, 1981; Weinstein, 1980). This "wishful thinking" or optimistic bias occurs in a variety of situations, from estimating the likelihood of a desired candidate winning an election (Babad, 1997) to predicting one's future salary (Weinstein, 1980). This phenomenon seems irrational: people have unrealistic expectations, and these expectations could lead to risky choices. Most explanations of wishful thinking have focused on it as an irrational cognitive bias.

While wishful thinking appears to be detrimental, we explore whether this bias could be a rational strategy given people's limited cognitive resources. Determining the best decision in a given situation requires considering all possible outcomes and their long range consequences. Yet, considering all of these possibilities is computationally intractable. Wishful thinking might help to compensate for the fact that people cannot fully evaluate the long term consequences of their actions. Reinforcement learning models for how artificial agents should make decisions have shown that employing optimistic beliefs when faced with actions with unknown consequences can lead to improved performance (Kaelbling, 1993). Similarly, using optimistic confidence bounds when exploring a search tree to choose actions can lead to improved agent performance in practice (e.g. Munos & Teytaud, 2006), although theoretical bounds show that this effect is not guaranteed (Coquelin & Munos, 2007).

We use Markov decision processes to determine whether these advantages for optimism in exploring actions with unknown consequences might also hold for the problem of choosing actions when one has limited cognitive resources. Using a simple decision making situation in which an agent must make choices about whether to keep or quit its job over a period of time, we examine how an optimistic bias affects the agent's performance when it can only consider the consequences of its actions for a limited period of time into the future. The limited horizon that agents consider mimics people's limited computational resources. In our simulations, we find that agents with moderate optimism perform better than agents with no bias when the horizon is relatively small compared to the true time period over which agents can act.

We begin by describing existing theories for the wishful thinking bias, and background about the Markov decision processes that we use to model human action planning. We then demonstrate how we model limited computational resources by restricting the horizon that agents consider when choosing actions. Through four simulations, we determine under what environmental and computational constraints an optimistic bias improves performance.

## Background

We briefly review existing theories about the wishful thinking bias and provide an overview of Markov decision processes.

### Wishful Thinking

Several theories of wishful thinking have been proposed that aim to explain why this bias occurs and how prevalent it is, although there remain significant gaps in the theories accounting for the phenomenon (Krizan & Windschitl, 2007). Ego-utility theory suggests that the bias protects the agent's self image, and thus should only occur in situations that affect self image. For example, most people estimate that they are better drivers than the average person (Svenson, 1981), and students believe they will have fewer health problems than their peers (Weinstein, 1980). In contrast to this theory, the strategic theory asserts that the bias should occur in a broader range of situations, and that it can be tempered by incentives for accuracy (Akerlof & Dickens, 1982). However,

both of these theories conflict with some experimental evidence showing that wishful thinking occurs in decisions not involving self image and that incentives for accuracy have limited effects on people's beliefs (Babad, 1997; Mayraz, 2011). Wishful thinking has also been explained as emerging from other cognitive biases, such as confirmation bias, cognitive dissonance, and failure to correct for information asymmetries rather than from a causal link between preferences and beliefs (Kahneman, Slovic, & Tversky, 1982; Knox & Inkster, 1968). For instance, Knox and Inkster (1968) explain observations that individuals' perceptions of a horse's likelihood of winning a race increased after betting on the horse as instances of reducing post-decision cognitive dissonance. While experimental evidence supports the existence of wishful thinking, none of these theories suggests it plays a part in improving decision making.

### Markov Decision Processes

A Markov decision process (MDP) is a decision-theoretic model of sequential decision making that naturally incorporates actions with uncertain effects and takes into account both immediate and future consequences (see Sutton & Barto, 1998, for an overview). MDPs model both the environment in which actions are being taken and the effect of the agent's actions on this environment. Formally, MDPs are defined by a tuple $\langle S, A, T, R, \beta \rangle$. At each time $t$, the environment is in some state $s_i \in S$. The agent chooses an action $a \in A$, and the transition model $T$ encodes the conditional probabilities $p(s_i^{(t+1)}|a, s_j^{(t)})$ that the state at time $t+1$ will be $s_i$ given that the current state is $s_j$ and the action chosen is $a$. MDPs may have either finite or infinite horizons. In a finite horizon MDP, which is the type we consider, the number of time steps in which an agent acts is limited to some $N$. Finite horizon MDPs must take into account both the current state and the amount of time remaining for the agent to act.

The incentive structure or goals are encoded in the reward model $R$. For each state-action pair, $R(s,a)$ is equal to the immediate reward of choosing action $a$ in state $s$. Usually, an agent is trying to choose actions that will result in large rewards over the period of time in which it can act. Thus, the agent must determine the value of each action for each state and timestep, taking into account both immediate and long term rewards. This value is known as the $Q$-value, and the optimal values $Q_t^*$ are defined as those achieved by always choosing the action with highest expected value for the current state $s$ and timestep $t$:

$$Q_t^*(s,a) = R(s,a) + \beta \sum_{s' \in S} p(s'|a,s) \max_{a'} Q_{t+1}^*(s',a'), \quad (1)$$

where $\beta$ is a discount factor that weights the value of immediate versus future rewards. The optimal $Q$-function can be calculated using dynamic programming (Bellman, 1957).

## Approximating Optimal Planning

Finite horizon MDPs provide a framework for calculating the best action in any state and with any amount of time remaining. Given unlimited computational resources, an agent with accurate beliefs about the transition and reward models will perform at least as well as an agent with biased beliefs. However, with limited resources, calculating the optimal policy may not be possible. Instead, approximations, such as limiting the horizon that the agent considers when planning its actions, are necessary. The complexity of solving for the optimal policy scales exponentially with the number of years in the future that are considered. Thus, limiting this horizon of consideration significantly reduces the complexity of the problem that the agent must solve. We explore whether a wishful thinking bias could prove advantageous when the planning horizon is small. We first consider this scenario in a specific decision-making environment, defined by the transition and reward models, and then conduct additional simulations that consider a broader set of environments to determine when the wishful thinking bias results in higher rewards.

To explore possible computational benefits of wishful thinking, we set up a Markov decision process in which various degrees of bias are expressed as inflated and deflated transition probabilities to high and low value states in agents' perception of the decision-making problem. Because agents' beliefs about the transition matrices affect their beliefs about the value of taking an action in a given state, differences in beliefs result in different action policies. For all of our simulations, we consider the problem of an agent making job-related choices and measure performance as the total earned rewards (salary) over the course of a fixed number of years.

## Simulation 1: A Simple Example

We first set transition and reward matrices reflecting typical trends for salaries and ease of acquiring a particular job.

### Methods

The simulation covers the problem of an agent choosing to *keep* or *quit* its current job over the course of $N = 40$ years. The states $S$ correspond to five possible jobs, which we label 'unemployed,' 'waiter,' 'police officer,' 'banker,' and 'movie star.' Since the MDP has a finite horizon, the policy defines which action to take at each time and in each state. Choosing *keep* means that the agent retains its current job for the next year and earns the full annual salary of that job. This reward is constant: there is no increase in salary over time. Choosing *quit* means that the agent takes a new job in the next year and earns a reward of half the annual salary of the current job.

The reward model reflects the fact that unemployment is typically the least remunerative of the jobs and movie star is the most: unemployment earns \$0, waiter earns \$30,000, police officer earns \$50,000, banker earns \$150,000, and movie star earns \$1,000,000. The transition matrix for choosing *quit* follows a realistic ordering of the prevalence and difficulty of achieving each profession (Figure 1(a)) The four agents we consider are shown in Figure 1: *realistic* agents have accurate beliefs that are equal to the true transition matrix, *optimistic* agents believe that higher valued states are more likely and lower valued states are less likely, *highly optimistic* agents

**(a)**

| | Unemployed | Waiter | Police | Banker | Movie Star |
|---|---|---|---|---|---|
| Unemployed | 0.135 | 0.650 | 0.200 | 0.010 | 0.005 |
| Waiter | 0.785 | 0 | 0.200 | 0.010 | 0.005 |
| Police | 0.335 | 0.650 | 0 | 0.010 | 0.005 |
| Banker | 0.145 | 0.650 | 0.200 | 0 | 0.005 |
| Movie Star | 0.140 | 0.650 | 0.200 | 0.010 | 0 |

**(b)**

| | Unemployed | Waiter | Police | Banker | Movie Star |
|---|---|---|---|---|---|
| Unemployed | 0.070 | 0.600 | 0.300 | 0.0200 | 0.010 |
| Waiter | 0.670 | 0 | 0.300 | 0.0200 | 0.010 |
| Police | 0.370 | 0.600 | 0 | 0.020 | 0.010 |
| Banker | 0.090 | 0.600 | 0.300 | 0 | 0.010 |
| Movie Star | 0.080 | 0.600 | 0.300 | 0.020 | 0 |

**(c)**

| | Unemployed | Waiter | Police | Banker | Movie Star |
|---|---|---|---|---|---|
| Unemployed | 0 | 0.050 | 0.050 | 0.100 | 0.800 |
| Waiter | 0 | 0 | 0.100 | 0.100 | 0.800 |
| Police | 0 | 0.100 | 0 | 0.100 | 0.800 |
| Banker | 0 | 0.150 | 0.050 | 0 | 0.800 |
| Movie Star | 0 | 0.850 | 0.050 | 0.100 | 0 |

**(d)**

| | Unemployed | Waiter | Police | Banker | Movie Star |
|---|---|---|---|---|---|
| Unemployed | 0.240 | 0.650 | 0.100 | 0.010 | 0 |
| Waiter | 0.890 | 0 | 0.100 | 0.010 | 0 |
| Police | 0.340 | 0.650 | 0 | 0.010 | 0 |
| Banker | 0.250 | 0.650 | 0.100 | 0 | 0 |
| Movie Star | 0.240 | 0.650 | 0.100 | 0.010 | 0 |

Figure 1: (a) True transition matrix for the action *quit* in Simulation 1. (b) Transition matrix considered by the optimistic agent. (c) Transition matrix considered by the highly optimistic agent. (d) Transition matrix considered by the pessimistic agent

exaggerate the beliefs of optimistic agents, and *pessimistic* agents believe higher valued states are less likely and lower valued states are more likely.

We simulated $100,000$ episodes for each agent and possible horizon. We considered eight possible horizons over which the agent could plan: 1, 2, 4, 5, 8, 10, 20, or 40 years. At a horizon of one year, only the immediate value of the action is considered, and thus agents always keep their initial jobs. A horizon of 40 years is equivalent to an agent with no computational limitations on planning. We model planning with a horizon $h$ as an agent finding an optimal policy for $h$ years, carrying out this contingent policy, and then planing for $h$ additional years, repeating this process until the full time has elapsed. This pattern might be thought of as analogous to a "five year plan": decisions are made to maximize the reward from the next $h$ years, and while the five year plan is contingent upon the effects of each action in the plan, consequences after the length of the plan are not considered.[1] Each agent follows an optimal policy for their beliefs and the planning horizon, with a small probability $\varepsilon$ of deviating from this policy at each timestep. This probability reflects the fact that human decision making is noisy; for all simulations, $\varepsilon = 0.05$. We set the discount factor $\beta = 1$, resulting in future rewards having the same value as immediate rewards.

## Results

As shown in Figure 2, the optimistic and highly optimistic agents earned more money on average than the realistic agent for small planning horizons. Both of these policies are more likely to take the risk of quitting a low paying job than the realistic policy. For small horizons, this risk taking is advantageous as it helps to compensate for the limited amount of time that the agent is considering. For larger horizons, the realistic agent can better estimate the value of risk taking, and thus outperforms the other policies. Across all horizons, there is no advantage for the pessimistic agent.

[1] An alternative possibility for incorporating the constraint of planning over a limited number of years into an MDP is to have the agent use the $Q_1$-values for the first action until the $N-h-1$th year, and then use the remaining $Q_t$ values for the final $h-1$ years. All simulations in this paper have also been conducted with this version of the policy, and results are very similar to using the "$h$ year plan" version of the policy.

# Simulation 2: Sampled Rewards

The results of Simulation 1 demonstrate that it is possible for an optimistic agent to outperform a realistic agent when the agent considers the effects of a decision over only a limited amount of time. However, these results do not illustrate whether this advantage holds for a variety of different types of reward and transition matrices. To explore how far these results generalize, we next consider a more general set of possible reward matrices, sampled from different distributions.

## Methods

Simulation 2 was conducted in the same way as Simulation 1, except that the reward matrix was varied for each episode. We consider sampling annual salaries from three distributions: an exponential distribution, a power law distribution, and a uniform distribution. For each distribution, we set the mean $\mu = 100,000$, and for the uniform distribution, we set the allowed range of rewards to $2\mu$. The exponential distribution produces the most skewed distribution, favoring small values, while the power law distribution is also skewed but has a heavier tail. To maintain the structure of higher salaries for harder to acquire jobs, we sort the sampled salaries such that the ordering matches Simulation 1: the highest salary goes to the *movie star* job and the lowest to *unemployed*.

We simulated $10,000$ episodes for each horizon, agent, and reward distribution. For each episode, we sampled a reward matrix, then generated episodes for each agent with that reward matrix. To compare earnings across episodes, we record the proportion of possible earnings that were earned in a given episode, where the possible earnings are the number of years in the episode ($N = 40$) multiplied by the maximum salary.

## Results

As shown in Figure 2, the average proportion of earnings acquired varies across the three reward distributions, but the agents show similar trends in earnings relative to one another. Unlike in Simulation 1, we do not see an advantage for the highly optimistic agent, even at short horizons; instead, this agent underperforms all other agents. However, a small advantage for the optimistic agent persists: at planning horizons of two, four, and five years, this agent has higher earnings than the realistic agent.

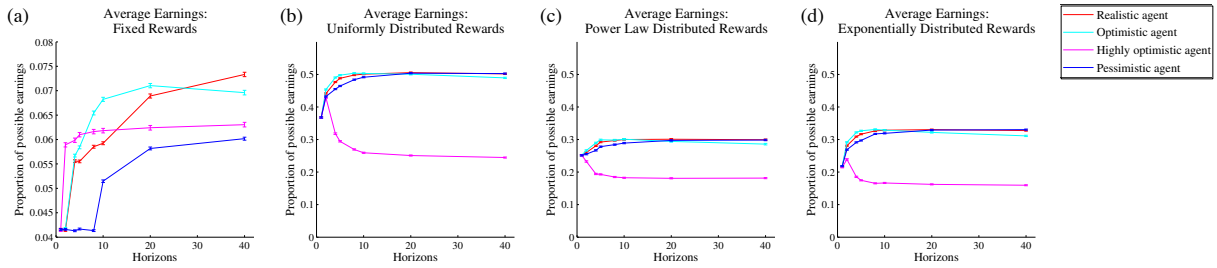


Figure 2: Average proportion of possible earnings acquired by each agent, Simulations 1 and 2. Error bars reflect 1.96 standard errors. (a) Simulation 1: At small horizons, the optimistic and highly optimistic agents earn more than the realistic agent. (b-d) Simulation 2: The proportion of rewards earned varies across sampling distributions, but the relative performance of the strategies remains the same. There is a slight advantage at very small horizons for the optimistic agent, although the highly optimistic agent performs poorly.

## Simulation 3: Sorted Transition Matrices

The previous simulation demonstrated that there is not always an advantage at short horizons for highly optimistic agents, but suggested that some advantage for an optimistic (wishful thinking) bias may exist more generally than in the specific conditions in Simulation 1. One of the difficulties in generalizing from these simulations is that they do not quantify the differences between agents: each of the biases was implemented by hand by altering a specific transition matrix. To quantify different levels of bias and develop a better understanding of how much optimism is useful in what situations, we now generalize the simulations such that both the transition matrix governing the likelihood of attaining particular jobs and the salaries for these jobs are sampled. As in Simulations 1 and 2, we constrain these matrices to pair jobs that are hard to achieve with higher salaries. To implement different levels of bias, we transform the transition matrix such that the agents' beliefs are skewed to be optimistic or pessimistic. By including a parameter in the transformation representing the desired degree of bias, we can determine whether there are advantages to moderate levels of optimism outside of the specific scenario explored in Simulations 1 and 2.

### Methods

The basic structure of Simulation 3 mirrored previous simulations, with forty years for the agent to act and five possible jobs. Both the salary and transition matrices were sampled in this simulation. The sampling of the salaries was the same as in Simulation 2. To construct the transition matrix, we sampled each distribution from a symmetric Dirichlet distribution with parameter $\alpha$. Larger values of $\alpha$ favor uniform distributions, while smaller values favor sparse distributions. We let $\alpha = 0.01, 0.1, 1, 10$. To ensure that harder to achieve jobs have higher salaries, we sort the transition matrix such that for each row, the transition probabilities decrease as the salaries increase.

Different levels of wishful thinking are introduced into this simulation through a bias parameter $\gamma$. The transition matrices assumed by biased agents are tied to the rewards associated with different states. The perceived likelihood of transitioning to state $s'$ after quitting $s$ is proportional to the true probability of reaching $s'$ from $s$ multiplied by the salary of $s'$ raised to the power of $\gamma$:

$$p_{\text{bias}=\gamma}(s'|s, a = \text{quit}) \propto p(s'|s, a = \text{quit}) \cdot R(s')^{\gamma} \qquad (2)$$

where $p(s'|s, a = \text{quit})$ is the true transition probability. When $\gamma = 0$, there is no bias: the agent's beliefs are the same as the true transition matrix. When $\gamma > 0$, the agent is optimistic: states with larger salaries will be deemed more probable outcomes than in reality, while states with smaller salaries will be deemed less probable. When $\gamma < 0$, the opposite occurs, resulting in a pessimistic agent. $\gamma$ with larger magnitudes result in greater degrees of optimism or pessimism. We considered $\gamma = -5, -1, 0, 1, 5$.

We simulated $10,000$ episodes for each horizon, $\alpha$, $\gamma$, and reward distribution. As in Simulation 2, we record the proportion of possible earnings earned in each episode to enable comparison across episodes with different rewards.

### Results

The results of the simulations are similar across the three reward distributions: as in Simulation 2, the absolute proportion of rewards earned does vary, but the strategies' performance relative to one another remains the same. We thus show only the results of the exponential distribution in Figure 3. As this figure shows, the benefits of a bias towards optimism are dependent on the characteristics of the transition matrix. When distributions are very sparse, as occurs with small $\alpha$, there is little potential to move between jobs, so all strategies perform relatively similarly. With slightly larger $\alpha = 0.1$, there is a clear disadvantage for extreme pessimism ($\gamma = -5$) and extreme optimism ($\gamma = 5$); mirroring the results of Simulation 2, the highly optimistic strategy is the worst performing strategy. However, this level of $\alpha$ also begins showing a limited benefit for a slightly optimistic strategy ($\gamma = 1$), dependent on reward distribution. There is no advantage for the exponential distribution, while the slightly optimistic strategy outperforms the realistic strategy by at least two standard errors for horizons of two and four for the uniform distribution and horizon two for the power law distribu-

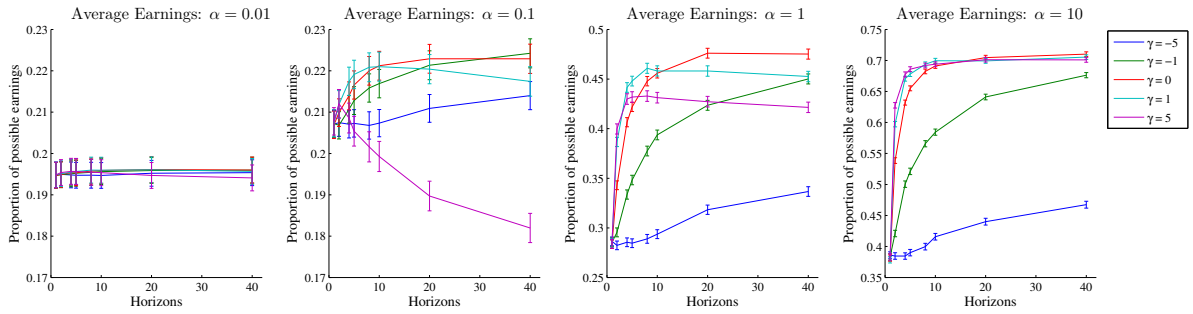


Figure 3: Average lifetime earnings as a proportion of possible earnings in Simulation 3, with rewards sampled from an exponential distribution. $\gamma$ dictates the degree and direction (optimistic or pessimistic) of the agent's bias. When $\alpha$ is large enough, optimistic agents ($\gamma > 0$) have an advantage over the realistic agent ($\gamma = 0$) at small horizons. Error bars reflect 1.96 standard errors.

tion. As $\alpha$ increases further, the benefit to an optimistic strategy also increases. The highly optimistic strategy tends to be best at the lowest horizons, eventually being outperformed by both the slightly optimistic and realistic strategies.

Overall, the results of this simulation suggest that an advantage for optimism in cases with very limited lookahead holds for many more transition and reward matrices than the example in Simulation 1. The effect is strongest where there is non-negligible probability on all possible states, as occurs with the larger $\alpha$ values, and does not fade even when the transition distributions are unlikely to be skewed ($\alpha = 10$). The advantage for optimism with short horizons across a range of environments suggests that there be many scenarios in which people must make decisions in which a similar advantage holds.

## Simulation 4: General Transition Matrices

In Simulation 3, we ensured that the lowest rewards were paired with the easiest-to-achieve states, mirroring the idea of high paying jobs being in limited supply. However, the results of the non-sparse transition matrices in Simulation 3 suggest that this condition may not be necessary for an optimistic advantage: the non-sparse transition matrices actually resulted in the largest advantage for optimism, even though high reward jobs were not much less likely than other jobs. Our final simulation explores this more general case: is there an advantage for optimism in cases where transition matrices and rewards are unrelated?

### Methods

The methods for this simulation were identical to Simulation 3, except that the transition matrices were not sorted. Thus, if one quits the *banker* job, one might be highly likely to transition to the *movie star* job, while if one quits the *police officer* job, the most likely transition might be to the *waiter* job. As in Simulation 3, we sampled rewards from exponential, power law, and uniform distributions with the same mean, and sampled the transition matrices from a symmetric Dirichlet distribution, considering $\alpha = 0.01, 0.1, 1, 10$. The same five agents were used, with biases set using Equation 2.

### Results

The results of Simulation 4 are very similar to those of Simulation 3, demonstrating that the sorting constraint does not have a large impact on relative earnings. Earnings are in general higher in this simulation, reflecting the fact that higher paying jobs are no longer the hardest to achieve. As in Simulation 3, the three reward distributions result in similar relative advantages for optimistic and pessimistic strategies: none of the reward distributions show any advantage for pessimism, but with larger $\alpha$, there is a bias for optimistic strategies at short horizons. As shown in Figure 4, smaller horizons with $\alpha \geq 0.1$ result in advantages for the highly optimistic and slightly optimistic strategies, with greater advantages for $\alpha \geq 1$ and a much more robust advantage for the slightly optimistic strategy. Only at $\alpha = 0.01$ are the results of Simulation 3 characteristically different than in Simulation 4. In Simulation 3, the sparse, sorted transition matrices meant that the lookahead horizon had very little impact on rewards, as quitting one's job rarely held any possibility of improved salary. In Simulation 4, rewards and transition probabilities are unassociated, so improvements in rewards from longer lookahead are possible. However, this $\alpha$ still has the smallest range of possible rewards, demonstrating the limited impact of strategies when transition probabilities are sparse.

## Conclusion

In this paper, we have explored whether an optimistic or wishful thinking bias can improve performance when agents have limited foresight into the consequences of their actions. Using Markov decision processes, we have shown that when the horizon an agent can consider when planning is relatively limited compared to the true time horizon of a task, some bias towards optimism results in higher total reward. As the horizon that the agent can consider increases, the gain for an optimistic policy decreases, and optimism eventually becomes detrimental to performance. Overall, these results demonstrate the possibility that wishful thinking could be a computational heuristic for improving performance rather than simply a mistake in people's reasoning.

Our model provides a proof of concept for the possibil-

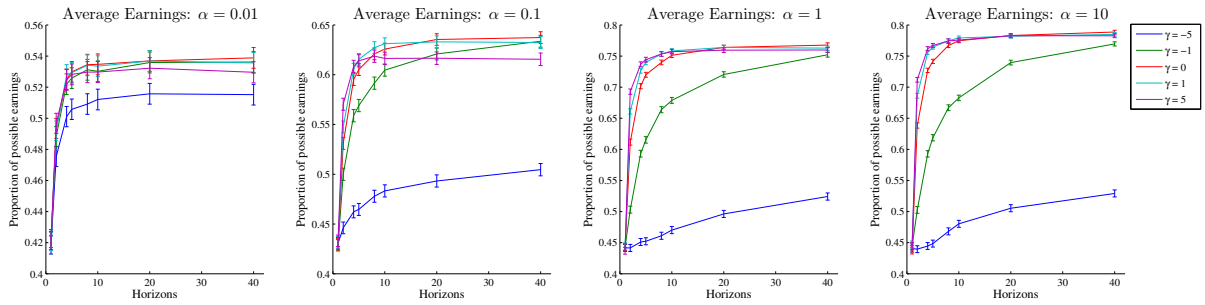


Figure 4: Average lifetime earnings as a proportion of possible earnings in Simulation 4, with rewards sampled from an exponential distribution. Just as in Simulation 3, optimistic agents ($\gamma > 0$) have an advantage over the realistic agent ($\gamma = 0$) at small horizons, demonstrating that the advantage for optimism does not require a link between relative salary and ease of acquiring a job. Error bars reflect 1.96 standard errors.

ity that wishful thinking is a strategy. The simple model is only an approximation for real world decision making, which is typically more complex. Additionally, our model examines only one possible computational approximation: limiting the time over which one considers the consequences of one's actions. Other approximations, such as feature-based reinforcement learning algorithms to model time without an explosion of the state space or forward search approximations (e.g., Ross, Pineau, Paquet, & Chaib-draa, 2008), are possible. Considering these other types of approximations would further develop our understanding of whether there are situations in which wishful thinking is advantageous across a broad set of rational process models.

The model we have presented provides a starting point for future work. First, simulations of wider set of decision problems are necessary to establish whether an optimistic bias holds in more complex situations. In the case of choosing a job, we assumed that people could retain the same job for indefinite periods of time and that each job had a constant salary; thus, if one job was better than some other job for one year, it would be much better than the other job if retained over many years. In this type of structure, it is intuitive that an optimistic bias might help to compensate for a bounded horizon. However, other situations may be more complex, such as a job that initially has a higher salary leading to less potential to switch to other high paying jobs than a job with a lower initial salary. Exploring such situations will allow us to establish a more general theory for what features of a situation result in an advantage for optimistic biases. Experiments are also a necessary next step for determining whether wishful thinking is used mainly in situations where it provides an advantage or is modulated by the computational complexity of the task. For example, one might ask participants to make decisions in situations with short or long time horizons. If wishful thinking is used to deal with computational complexity, one would expect it to be less common for simpler decisions with shorter time horizons. While there are a number of steps necessary to establish whether the model we have presented bears on how humans actually cope with predicting the long term results of their choices, our results suggest that it is not necessary to assume that all cases of a wishful thinking bias are detrimental or irrational: instead, this bias may represent a better approximation to an optimal solution when only limited computational resources are available.

**Acknowledgements.** This work was supported by ONR MURI grant number N00014-13-1-0341 to TLG.

## References

Akerlof, G. A., & Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *The American Economic Review*, *72*(3), 307–319.

Babad, E. (1997). Wishful thinking among voters: Motivational and cognitive influences. *International Journal of Public Opinion Research*, *9*(2), 105–125.

Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ, USA: Princeton University Press.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review*, *89*(1), 306–318.

Coquelin, P.-A., & Munos, R. (2007). Bandit algorithms for tree search. In *Proceedings of the Twenty-Third Annual Conference on Uncertainty in Artificial Intelligence* (pp. 67–74).

Kaelbling, L. P. (1993). *Learning in embedded systems*. MIT Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Knox, R. E., & Inkster, J. A. (1968). Postdecision dissonance at post time. *Journal of Personality and Social Psychology*, *8*(4, Pt. 1), 319.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*(1), 95–121.

Larwood, L., & Whittaker, W. (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology*, *62*(2), 194.

Lyles, M. A., & Thomas, H. (1988). Strategic problem formulation: Biases and assumptions embedded in alternative decision-making models. *Journal of Management Studies*, *25*(2), 131–145.

Mayraz, G. (2011). *Wishful thinking* (Tech. Rep. No. CEP Discussion Paper 1092). Centre for Economic Performance, London School of Economics.

Munos, S. G. W., & Teytaud, O. (2006). Modification of UCT with patterns in Monte-Carlo go. *Technical Report RR-6062*.

Ross, S., Pineau, J., Paquet, S., & Chaib-draa, B. (2008). Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*, *32*(1), 663–704.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. MIT Press.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*(2), 143–148.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806.