Optimal Strategy for Living a Good Life

Xavier Hilbert

April 2023

1 Introduction

Why do you want to be a chef?
My father is a chef, and I want to make him proud.
Why do you want to make him proud?
Because it would make me happy.

Why do you volunteer at the homeless shelter?
Because it is the right thing to do.
Why do you want to do the right thing?
Because doing the right thing makes me happy or conversely doing the wrong thing would make me sad.

At 2 am, police were called in response to a gunshot. Upon arrival, they found Adam dead with a note that says *I no longer want to be unhappy.*

As exemplified, regardless of cost or virtue, all actions are performed because the agent believes the action will promote happiness or avoid unhappiness [Mil79].

It is reasonable to believe the chef and volunteer are living a good life; after all, they are doing what makes them happy. But for Adam, if you were able to ask him if he had lived a good life, he would likely say no despite the fact he did what he thought would make him happy. Therefore, there is a disconnect between performing actions that you believe will make you happy and a good life.

There are three reasons for this disconnect: misjudgement, misfortune, and bad strategy. After leaving a party, have you ever thought *that wasn't as fun as I thought it would be, I wish I stayed home*? This is misjudgement. We go to the party because we believe it will make us happier than the alternatives (staying home), but we misjudge, ending up with less utility than we thought we would. Misjudgement is not the fault of the agent. How were you supposed to know the party would make you unhappy until you have tried it? But if you go to a similar party again and end up worse off, the agent is at fault because this is a bad strategy. Why waste time doing something that made you unhappy the first time, why not try something new instead?

There is a luck component to life. You did not choose your genetics nor your upbringing. Even with perfect strategy and judgement, if one has an incurable, chemical imbalance causing depression, this could prevent one from having a good life. Or perhaps perfect strategy and judgement leads one to ski but one ends up injuring himself on the slope. This is also misfortune. Although one chooses to ski, unlike a chemical imbalance, we can not fault the agent for his unhappiness, because he performed the action that maximizes his happiness - he was just unlucky.

Lastly, how do you approach living a good life? This is strategy. Unlike misjudgement and misfortune, we can fault an agent for bad strategy, because we make our choices. Fortunately, there is an optimal strategy for living a good life even in a world of misjudgement and misfortune.

2 Finding the Optimal Strategy

2.1 Notation and Groundwork

2.1.1 Actions, Experiences, Utility

Every action costs time and/or money and in turn, an action provides an experience. For example, spending time with family is an action, and after the action is complete, you gain the experience of spending time with family. However, sometimes we act to have a particular experience but end up having another. For example, one might want to experience Europe but experiences a plane crash. This is the unlucky, or in other cases, lucky, component of life. In addition to an experience, one can gain (or lose) time and/or money from an action.

Notationally,

$$action(time,\ money) \rightarrow [resulting\ experience,\ net\ time,\ net\ money]$$

So, if eating an apple takes 2 minutes, costs \$3, and extends your life by an extra second because it is healthy,

$$eating An Apple(2\,minutes,\ \$3) -> [eating\ an\ apple, 1\,second - 2\,minutes, -\$3]$$

If teaching takes 3 hours, costs \$5 to commute, and pays \$40,

$$teaching(3\,hours,\ \$5) -> [teaching,\ -3\,hours,\ \$35]$$

If chemotherapy takes 96 hours in total, costs \$15,000, but gives you back 30 years of your life,

$$chemotherapy(96hours, \$15,000) -> [chemotherapy, 30years - 96hours, -\$15,000]$$

An experience itself does not grant utility. Otherwise everyone would have the same utility for the same experience. Instead, everyone has a unique utility function.

Notationally,

$$U_{you}(eatingAnApple(2\,minutes,\ \$3)[0]) = 5\,utils$$
$$U_{friend}(eatingAnApple(2\,minutes,\ \$3)[0]) = -3\,utils$$

where

$$action(time,\ money)[0] = experience$$

2.1.2 Imperfect Information

We do not know how happy an experience will make us until we have it. Following, when we are born, we have no experiences so we do not know what makes us happy. As a result, our utility function has no terms initially. But when we experience things for the first time like hunger, we begin to discover our utility function...

$$U_{you}(hunger) = -20(hunger)$$

Then we sleep for the first time and gain 10 utils, discovering more of our utility function...

$$U_{you}(hunger,\ sleep) = -20(hunger) + 10(sleep)$$

To maximize our utility, we find the gradient with respect to each experience, choose the experience that has the highest, absolute gradient, and if positive, choose the experience and if negative, avoid it. For example,

$$\frac{\partial U}{\partial hunger} = -20$$
$$\frac{\partial U}{\partial sleep} = 10$$

Since the magnitude of the gradient with respect to hunger is highest and negative, we should choose to eat.

As we continue to have new experiences, we discover what makes us happy. If we had perfect information, meaning we knew our entire utility function at any given moment, maximizing happiness would be easy as following the above. However, it is impossible to know what will make us the happiest at any given moment, because we can not try every possible experience. Moreover, there

are constraints one must take into account, because regardless of how much one enjoys vacationing, one may lack the time and/or money to always be on vacation.

Time is how much time you have left, and is defined as

$$Time = biological\ clock + \sum_{action \in All\ Acions\ Performed} (action[1])$$

where biological clock is the amount of time one would have if their choices had zero impact on one's lifespan and

$$action(time,\ money)[1] = net\ time$$

As an example, let's say a baby will live to be 80 years old (if his actions have no impact on his lifespan). And assume, since birth, the baby has performed only two actions: sleeping and eating applesauce (and apple sauce has given him a second extra of life).

$$eatingAppleSauce(3minutes, \$0) \rightarrow [eatingAppleSauce, -3minutes{+}1second, \$0]$$

$$sleep(8\ hours,\ \$0) \rightarrow [sleep,\ -8\ hours, \$0]$$

$$Time = 80\ years + (-3\ minutes + 1\ second) + (-8\ hours)$$

Money is how much money you currently have, and is defined as

$$Money = \sum_{action \in All\ Acions\ Performed} (action[2])$$

$$action(time,\ money)[2] = net\ money$$

For the following actions: accepting $50 from your parents, working 8 hours at $20 per an hour, and paying $100 worth of bills,

$$\begin{aligned} Money = {} & acceptingMoneyFromParents(10\ seconds,\ \$0)[2] \\ & + working(8\ hours\ ,\ \$0)[2] + payingBills(2\ minutes,\ \$100)[2] \end{aligned}$$

$$Money = 50 + 160 - 100 = \$110$$

2.2 The Armed Bandit Problem and Living a Good Life

2.2.1 Introduction

Imagine a casino. There are K slot machines (also called bandits) each with an "arm" to pull, and once pulled, you receive a random reward. Each machine has its own unknown, reward distribution, and your goal is to maximize the total amount of rewards you receive given a fixed number of pulls. Finding the optimal strategy for this problem will give insight into the optimal strategy for living a good life.

2.2.2 Proving the Analogy

At any moment, we have an infinite number of actions to choose from, but realistically, we consider a finite (K) number of actions. For example, when planning a vacation, we hardly consider more than 50 locations even though we have the world to choose from. Likewise, there are a finite (K) number of slot machines in our imaginary casino. Bandits and actions are interchangeable hereafter.

$$\text{K actions} \longleftrightarrow \text{K bandits (slot machines)}$$

Recall every action grants an experience. However, there is variability in the experience we receive, and consequently, there is variability in the utility we will receive. For example, we might play the slot machine but the utility we receive varies based on if we experience a win or a loss. Even actions we take for granted such as eating an apple pull from a random distribution where there might be a .000001% chance you choke on the apple.

$$\therefore \text{After acting, earn random amount of utility} \longleftrightarrow \text{After pulling a bandit, earn a random reward}$$

Again, every action costs time and/or money. Given a wage, w, where w is the amount of money you can generate per unit of time, we can treat an action as if it costs time xor money [Per20]. In the case of time,

$$action\,(time,\, money) = action\left(time,\, \frac{money}{w}\right) = action\left(time\, +\, \frac{money}{w}\right)$$

Considering this, we can fuse our money and time constraints; let Budget (B) equal how much time you have left,

$$Budget = biological\ clock + \sum_{action \in all\ Actions\ Performed} \left(action(time,\, money)[1] - \frac{money\ to\ perform\ action}{w} \right)$$

Let's say one's $Budget = 5\ years$. Then, after chemotherapy, one gains an additional 30 years of life. But it costs him \$15,000, and he makes \$30,000 per

year,

$$Budget\ New = 5\ years + 30\ years - 6\ months$$

This model seems to make sense. If one is uber wealthy, meaning one's w is very high compared to the financial costs of all one's actions, one's budget is

$$Budget = biological\ clock + \sum_{action \in all\ Choices\ Made} (action(time,\ money)[1])$$

In the case of an affluent individual who is treated with chemo,

$$Budget\ New = 5\ years + 30\ years$$

Whereas, most must consume their budget by an additional amount to perform actions that cost money. Moreover, the largeness of one's wage does not affect their budget directly since a higher wage doesn't give you a higher budget (more life), it only improves how the budget decays with respect to the financial cost of an action.

$$\text{Budget} \longleftrightarrow \text{Number of pulls}$$

This is because if one uses up all their budget, one can not generate any more happiness (since he or she is out of time), and if you use all your pulls at the casino, you can not generate any more rewards.

Regardless of w, to maximize happiness, we should choose an action that 1) costs little and 2) provides high utility and keep repeating the action until our budget equals 0. Therefore, we should choose the action with the highest

$$efficency = \frac{utility\ from\ action}{Time\ required\ to\ perform\ the\ action + \frac{financial\ costs\ to\ perform\ the\ action}{w}}$$

$$efficency = \frac{utility\ from\ action}{cost\ of\ action}$$

$\therefore$ Choose action with highest efficiency $\left(\frac{utility}{cost}\right) \longleftrightarrow$ Choose the bandit with the highest efficiency $\left(\frac{reward}{cost}\right)$

Therefore, the best strategy for the armed bandit problem is equivalent to the best strategy for living a good life.

2.2.3 Exploration vs Exploitation

Each time we pull a bandit, we are given a reward. To find the empirical efficiency for each bandit, we can use

$$Calculated\ efficiency = \frac{\left(\frac{Sum\ of\ all\ the\ rewards\ given\ by\ bandit\ b}{Number\ of\ pulls\ of\ bandit\ b}\right)}{Cost\ of\ bandit\ b}$$

which boils down to

$$Calculated\,efficiency = \frac{average\,reward\,of\,bandit\,b}{Cost\,of\,bandit\,b}$$

If we had perfect information, we would always choose the bandit with the highest efficiency. Instead, we are faced with an exploration vs exploitation dilemma. That is, to maximize the rewards we receive given a fixed number of pulls, we should choose the bandit that has the highest calculated efficiency, yet to actually maximize the rewards we receive, we need to explore all bandits in order to to approximate the true efficiency of each one and exclusively choose (exploit) the most efficient one.

$\therefore$ Exploit: Choose action with highest (calculated) efficiency $\longleftrightarrow$ Choose the bandit with the highest (calculated) efficiency

$\therefore$ Explore: perform an action that is different from your most historically efficient action $\longleftrightarrow$ pull a bandit that is different from your most historically efficient bandit

In other words, "the exploration versus exploitation dilemma can be described as the search for a balance between taking the empirically best action as often as possible and exploring other actions to cover the case of there being other actions that are better" [ACF02].

I will consider 3 strategies that try to maximize utility: naive approach, epsilon greedy, and upper confidence bound (UCB).

2.2.4 Naive

Always explore: If you pull each slot machine a sufficient number of times so that you essentially know the true efficiency of each machine, then you should exclusively choose the machine that has the highest calculated efficiency instead of continuing to try all the machines.

Take away: we shouldn't always explore; at some point, we should do what has historically made us happy.

Always exploit: Imagine a child, and he plays video games for the first time. He enjoys gaming more than any other activity. Consequently, he plays them for the rest of his life, since he is exploiting what is most efficient. Instead, if he explored more, he would have probably found experiences that are more efficient, exploited them, and would have been a lot happier.

Take away: we also shouldn't only pursue what has historically made us the most happy.

2.2.5 Epsilon Greedy:

Epsilon greedy is based on the idea that we should explore some $\epsilon\%$ of the time and exploit the rest of the time [ACF02]. The following algorithm describes the strategy [Hil23]:

```
def epsilon_greedy(bandits, epsilon):
 #generate a random number between 0 and 1
 p = np.random.random()
 if p < epsilon:
  #Choose bandit at random.
  return random.choice(bandits)
 else:
  #Choose bandit with highest calcualted efficency.
    return np.argmax([b.calculated_efficiency for b in bandits])
```

Equivalent Strategy for Living a Good Life: Write down a list of experiences you want to have. Then predict how efficient each experience will be for you and pursue the most efficient one. If another experience becomes superior, choose the better one. And while continuing this cycle, add some randomness to your life. Try a new restaurant. Try a new sport. This is because as you have new experiences, you may find one that is more efficient compared to any of the ones you have had.

2.2.6 UCB:

$$\underset{bandit}{\operatorname{argmax}}\left(bandit.calculated\_efficiency + \sqrt{\frac{2\ln(step)}{bandit.number\_of\_pulls}}\right)$$

UCB says you should choose the bandit that gives the highest value for the formula above [ACF02]. *Step* is the total number of pulls you done so far, and $bandit.number\_of\_pulls$ is how many times you have pulled the particular bandit.

UCB is divided into two components [Rob20]. If we consider the first term,

$$\underset{bandit}{\operatorname{argmax}}\left(bandit.calculated\_efficiency\right)$$

then we would only exploit since with every turn we are choosing the action with the highest calculated efficiency. The second term, the exploration term,

$$\underset{bandit}{\operatorname{argmax}}\left(\sqrt{\frac{2\ln(step)}{bandit.number\_of\_pulls}}\right)$$

makes bandits we haven't tried in a while more appealing (since, for these bandits, step will continue to increase and $number\_of\_pulls$ will stay constant)

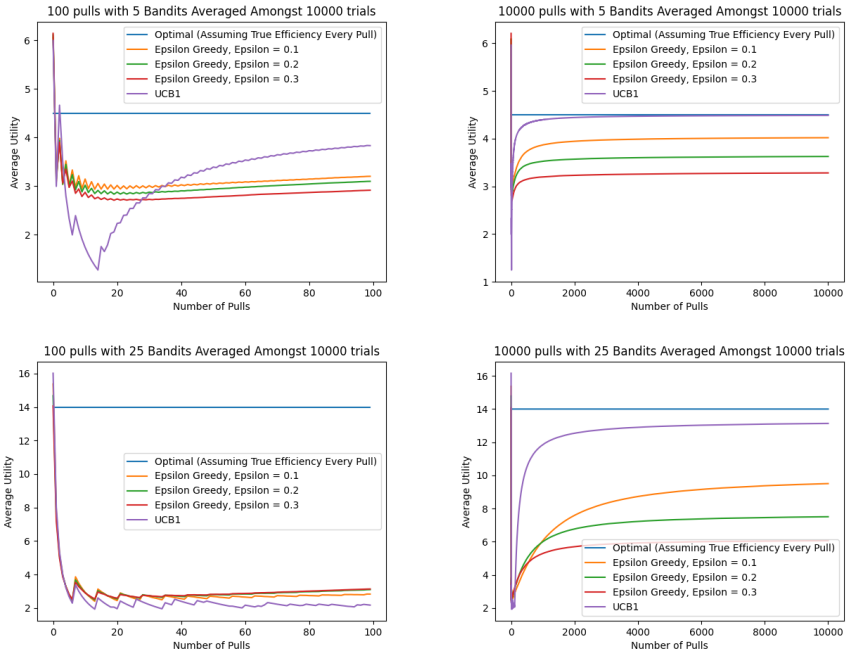
[Rob20]. In addition, if a bandit has never been pulled, UCB demands that we try it because

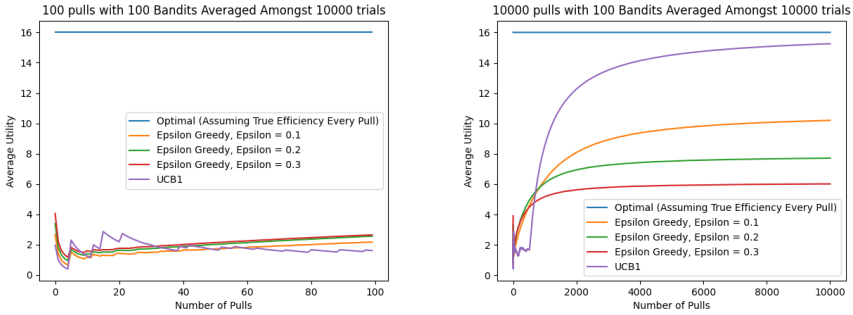
$$\sqrt{\frac{2\ln(step)}{0}} = \infty$$

Equivalent Strategy for Living a Good Life: UCB values high exploration initially and as time progresses, it values it less and less. This is because as step increases, the exploration term gradually decreases (since as 'n' approaches infinity $\frac{\ln(n)}{n}$ goes to zero), until eventually, actions are selected based solely on the exploitation term [Rob20]. Therefore, initially you should try as many experiences as possible. Then as time passes, explore less and exploit more. Then once you reach a certain age, you should only do what is most efficient.

2.2.7 UCB vs Epsilon Greedy

I simulated the armed bandit problem varying the number of pulls and number of bandits. The mean, standard deviation, and cost of each bandit was chosen at random. For each pull, a bandit's reward was pulled from $Normal(\mu, \sigma)$ and cost remained fixed [Hil23].





Interpretation The goal is to choose the strategy with the highest area under the curve since that area is total happiness. The "optimal" strategy represents when we have perfect information and always pull the bandit with the highest efficiency; thus, it will have the highest possible area under the curve. Lastly, for the ϵ-greedy algorithms, recall, a lower ϵ means that algorithm will perform more exploitative actions.

A commonality between all graphs is UCB dips at the the beginning. This is because UCB explores all bandits initially; while the ϵ-greedy algorithms start exploiting right away. Another commonality is as the number of bandits increase, each algorithm performs worse relative to the optimal strategy, which suggests that one should seek few experiences unless they have a large budget.

Comparing the right and left hand sides of the graphs, UCB always outperforms the ϵ-greedy algorithms when it has 10,000 pulls (budget is large). However, UCB underperforms (except for the case of 5 bandits) when it has 100 pulls (budget is small). This is because once UCB is done exploring, it has consumed it's budget before it can exploit what is has learned while ϵ-greedy exploits throughout.

Takeaway is UCB works great if

$$\frac{Number\ of\ pulls}{Number\ of\ bandits} = \frac{Budget}{Number\ of\ expereinces\ you\ want\ to\ have}$$

is large and

ϵ-greedy algorithms work better if the opposite. That is, if

$$\frac{Number\ of\ bandits}{Number\ of\ pulls} = \frac{Number\ of\ expereinces\ you\ want\ to\ have}{Budget}$$

is large.

As a consequence of these results, if your budget is low compared to all the experiences you want to have, you should exploit more at the beginning of life. While if your budget is large compared to the number of experiences you want,

you should explore exclusively. Then explore less and exploit more, and once you reach a certain age, exploit only.

My code is available [Hil23]. You can create your own bandits and see what strategy is optimal for you. However, it is hand wavy since the model assumes you know your budget in advance and you know the average utility of each of your experiences (albeit this can be compensated by a large standard deviation).

2.2.8 Model Assumptions

The model assumes you can pick k (finite) actions when in reality there are infinite options. So there is some responsibility for one to pick actions one believes to be highly efficient for oneself, and the model will figure out the optimal strategy given one's budget. As a result of this assumption, education is important. This is because when you look at data on what experiences generally make people happy, it is likely you can extrapolate to yourself. Consider Alice and Bob. If Alice researches different vacation destinations and chooses her top 5 based on what others suggests and Bob chooses 5 at random, when the model returns an optimal strategy for both, Alice will likely be happier. As the phenomena "garbage in, garbage out" suggests, if you choose k actions and they are all bad, the model will pick the best strategy, but strategy doesn't matter if all options are bad. Misjudgement can certainly happen, and it is not the fault of the person; if everyone recommends restaurant X but one ends up not liking it, one shouldn't be faulted for choosing eating at restaurant X as a bandit. But if no research was done in picking one's bandits, one can be faulted, because this is bad strategy.

The model also assumes the number of bandits does not decrease or increase overtime. I believe the results would be similar if bandits could be removed. This is because if you want to remove an experience from your list (i.e. don't want to have it anymore) and the reward for pulling it is negative enough to cause calculated efficiency to drop such that the bandit will never be pulled again, you have effectively removed it from your list. Adding experiences would affect the results. If you add an experience after each step, for example, UCB will never end up exploiting (since it has to explore all options), which also gives insight. That is, if you frequently add on experiences that you want to try, you are better off taking a ϵ-greedy approach over a UCB approach. Of course, this is speculation and a more complicated model would have to be created to confirm these predictions.

Also, the model assumes independent and identically distributed (iid) random rewards. However, in life sometimes an experience becomes more (or less) rewarding the more you have it. But if rewards do increase (or decrease), the calculated efficiency will increase (or decrease), and it is likely that the model will learn and choose (or avoid) the experience anyway. Again, speculation.

2.2.9 Further Analysis, Hyper Parameter Tuning, and Important Questions

Budget (how long you live for) can be small for a variety of reasons. Sometimes, we are unfortunate and are diagnosed with diseases like cancer. Other times, we make poor choices like not exercising or putting ourselves in dangerous circumstances. Since total happiness is the area under the curve, which is a function of budget size, it is important to maintain a large budget. Therefore, we should ideally choose actions that increase our life expectancy such as a healthy diet and exercise. The financial costs of experiences are also in the Budget equation, but they are factored into the efficiency. So if the model chooses a high cost experience, it means that the experience adequately compensates in utility. It then follows that extravagant experiences are fine as long as they are efficient.

On a humanity macro level, there are many, many experiences we can have, yet, in comparison, we live short lives. Our budget is small and the amount of bandits is large. This means our best strategy should be ϵ-greedy, where we exploit most of the time. That is, the best strategy to living a good life is to exploit and do what makes you happy as life is too short to figure out what makes you the happiest.

To turn this paper on its head, it might be that the amount of experiences is so large and our budgets are so small that it is akin to having one pull at the casino. Therefore, strategy does not matter. In fact, the idea of having a strategy is nonsense since there is no strategy in luck.

3 Justifying Maximizing Happiness

3.1 Collaboration

Maximizing your own utility has the potential to interfere with the happiness of others. For example, sometimes we can steal and not get caught. Benefiting yourself at the expense of others is morally greedy, but it seems being greedy is required to maximize you own happiness in certain situations. However, it is proven that taking does not maximize your happiness.

In a prisoner dilemma environment, a two player game, each person is trying to maximize their own utility. Each player has an option to cooperate or defect. If both cooperate, both players receive 3 utils. If one defects and one cooperates, the betrayer receives 5 utils and the victim receives 0 utils. And if both players defect, each player receives 1 util.

These reward distributions are akin to real life. This is because if two people are kind towards each other, both benefit. But there are situations where one can throw another under the bus for their own benefit. And if both have animosity towards each other, no one benefits.

In a study, 60 computer programs played a simulated prisoners dilemma as

described. Ultimately, the tit-for-tat strategy out preformed all others in maximizing utility [Axe80]. The strategy is to cooperate initially then choose what the other person did last. The author of the study noted that tit-for-tat demonstrates the value of being kind initially, of being somewhat forgiving, but also the importance of being provocable [Axe80].

The author, intentionally or not, made the games represent real life even more by introducing variability in the length of the game. That is, after the two players choose their action, there was a .35% the game would end [Axe80]. This represents real life, because there are some people you have a lot of interaction with and others not so much. In addition, you can never be certain when or if your interactions with someone will come to an end. Following, tit-for-tat is optimal in real life situations.

Tit-for-tat also feels morally right. If someone is nice to you, you should reciprocate. But if someone is mean towards you, there is no duty to still be nice, and if someone is initially not nice, but then chooses to be kind, there is room for forgiveness.

3.2 Society

Only seeking to maximize your own happiness seems selfish since you are not helping society (unless volunteering, contributing to the world, etc. are efficient for you). But this is not the case.

A reasonable societal goal is to maximize the aggregate utility [McD23]. Mathematically, we want,

$$Max(U(all\ people\ in\ society))$$

where U is the aggregate utility of people in society. Given that there are n people in society,

$$Max(U(all\ people\ in\ society)) = Max(U(person\ 1,\ person\ 2,\ \dots\ person\ n))$$

Since U is the additive (aggregate) utility of all people and each person has their own utility function,

$$\begin{aligned} &U(person\ 1,\ person\ 2,\ \dots\ person\ n) \\ &= U_{person\ 1}(person\ 1) + U_{person\ 2}(person\ 2) + \cdots + U_{person\ n}(person\ n) \end{aligned} \tag{1}$$

then,

$$\begin{aligned} &Max(U(person\ 1,\ person\ 2,\ \dots\ person\ n)) \\ &= Max(U_{person\ 1}(person\ 1) + U_{person\ 2}(person\ 2) + \cdots + U_{person\ n}(person\ n)) \end{aligned} \tag{2}$$

But it is not intuitively true that

$$Max(U_{person\ 1}(person\ 1) + U_{person\ 2}(person\ 2) + \cdots + U_{person\ n}(person\ n)) \\ = Max(U_{person\ 1}(person\ 1)) + Max(U_{person\ 2}(person\ 2)) + \cdots + Max(U_{person\ n}(person\ n))) \tag{3}$$

This is because if person 1 maximizes their utility, it is conceivable, using the aforementioned prisoner dilemma, that he or she should take from another (let's say person 2) and gain 5 utils. In consequence, the statement above would be false since the maximum utility of society is 6 not 5. However, remember, that it is proven tit-for-tat maximizes happiness. So if everyone is truly maximizing his or her own utility, which again requires tit-for-tat, then everyone would choose cooperation first. Following the strategy, everyone would keep reciprocating cooperation. And if everyone chooses cooperates, utility of society will be maximized. Therefore, the statement above holds.

Ergo, to maximize the aggregate utility in society, each person needs to maximize his or her own utility by performing actions that are optimal. As discussed, since we do not have perfect information on our own utility function, we have to strategize how to maximize utility either through UCB or ϵ-greedy. Either way, maximizing your happiness is necessary for achieving the societal goal of maximizing the aggregate utility of all people. Therefore, it is not a greedy to maximize your own happiness. In fact, it is on the contrary.

4 Implications for Myself

When writing this paper, I realized that we have multiple armed bandit problems happening in several domains in our lives. For example, when I was young, I tried several sports: soccer, baseball, basketball, tennis, and golf. Then I tried fencing and found that it gives me the highest efficiency. I gradually stopped exploring other sports and started exploiting my love for fencing, which I have been doing for the past 9 years and plan on continuing after I graduate.

However, in other domains, I have just begun my version of the armed bandit problem. For graduate school versus industry, I am choosing the later. Although I have a high efficiency for learning, I have been a student my whole life and want to (and should) explore working, because if I have a higher efficiency for work than school, then the time I'd spend in graduate school would not be optimally spent. If industry gives me a lower efficiency than graduate school, then I'll probably go back to school and pursue academics instead.

In terms of work, I have also begun an armed bandit problem. I have several entrepreneurial experiences, and entrepreneurship gives me high efficiency. In addition, last summer, I was a tech consultant, which I enjoyed. But after I graduate, I want to try a technical position as a software engineer and/or data scientist, because it doesn't make sense to exploit consulting or entrepreneurship

when I have not explored other bandits that I also think could be good options. I hope to continue experiencing new roles, then start settling into roles I enjoy, and then choose the role I enjoy the most and exploit it.

I plan on working in New York. This is because I have never lived in a city before, and I want to see how much I enjoy it; as I have lived in Kansas, Maryland, and North Carolina, eventually I want to find a location that I enjoy most and exploit it. Also, New York is one of the largest casinos in the world in terms of bandits since there are so many new experiences I can explore, which will hopefully accelerate my progression through my multiple armed bandit problems.

Marriage and relationships are unique. We are not married nor in committed relationships for the better part of our youth, so it is only natural to want to try the relationship bandit. But once you commit to marriage, it is assumed that you are giving up your right to pull other bandits. Of course, you can always divorce or cheat and this is a sign that you probably didn't do enough exploration, because if you find a bandit gives you the most efficiency, you should always choose that bandit. That being said, my plan is to take the UCB approach. That is I plan to marry (exploit) once I am sure she is the one.

References

[Mil79] John Stuart Mill. *Utilitarianism*. The Floating Press, 1879. ISBN: 9781775410614.

[Axe80] Robert Axelrod. "More Effective Choice in the Prisoner's Dilemma". In: *The Journal of Conflict Resolution* 24.3 (1980), pp. 379–403. ISSN: 00220027, 15528766. URL: http://www.jstor.org/stable/173638 (visited on 04/16/2023).

[ACF02] P. Auer, N. Cesa-Bianchi, and P. Fischer. "Finite-time Analysis of the Multiarmed Bandit Problem". In: *Machine Learning* 47 (2002), pp. 235–256. DOI: https://doi.org/10.1023/A:1013689704352.

[Per20] Bill Perkins. *Die with Zero: Getting All You Can from Your Money and Your Life*. Houghton Mifflin Harcourt, 2020. ISBN: 9780358099765.

[Rob20] Steve Roberts. "The Upper Confidence Bound (UCB) Bandit Algorithm Multi-Armed Bandits: Part 4". In: (Oct. 26, 2020). URL: https://towardsdatascience.com/the-upper-confidence-bound-ucb-bandit-algorithm-c05c2bf4c13f.

[Hil23] Xavier Hilbert. *Armed-Bandit-and-Strategy-to-Live-a-Good-Life*. https://github.com/XavierHilbert/Armed-Bandit-and-Strategy-to-Live-a-Good-Life. 2023.

[McD23] DeForest McDuff. "Society 3 - Human Progress". 2023.