Quantifying Risk Aversion and Risk-Seeking Behavior Through Utility-Based Policy Learning in Blackjack

Praveen Bandla, Tanishq Sardana, Dipesh Tharu Mahato

Center for Data Science, New York University

praveen.bandla@nyu.edu, ts5430@nyu.edu, dm6259@nyu.edu

Abstract

Modeling human decision-making under risk and uncertainty remains a significant challenge, with implications in fields like economics and cognitive science. Many decision-making models fail to fully account for individual risk preferences and utility functions. This work investigates the relationship between payoff and utility across diverse risk rofiles, using Blackjack as a simulation to quantify rewards and outcomes. By modeling the game as a deterministic environment, we apply dynamic programming to compute action values for each state and subsequent subgame encountered during gameplay based on different utility functions, and select the highest-value action, weighted by probabilistic outcomes[1]. Drawing on cognitive theories such as Prospect Theory and Expected Utility Theory, our findings show that variations in utility functions significantly influence decision-making and financial outcomes, offering new insights into decision-making under uncertainty.

Keywords: Markov Decision Process; Blackjack; Expected Utility Theory; Risk Preferences; Dynamic Programming; Bellman Equation; Prospect Theory

Introduction

Modeling human behavior is a crucial aspect of numerous machine learning (ML) applications, particularly in areas that involve decision-making algorithms that reflect human behavior or tendencies such as recommendation engines, autonomous systems, and reinforcement learning. Understanding and replicating human decision-making processes plays an integral role in the success of model learning within these spaces. In the context of reinforcement learning, for instance, the degree to which reward functions accurately reflect human tendencies have significant bearing on the success of a given model. However, designing reward functions that accurately reflect human behavior is a challenging task due to the inherent complexity of human decision-making, especially under conditions of uncertainty and risk.

A wealth of research across disciplines such as behavioral economics, cognitive science, and decision theory has provided valuable insights into human behavior. For example, Kahneman and Tversky's Prospect Theory (1979) introduced the idea that people evaluate outcomes based on subjective value rather than objective payoff, emphasizing concepts like loss aversion and diminishing sensitivity. Similarly, behavioral models like Expected Utility Theory (Von Neumann Morgenstern, 1944) and Cumulative Prospect Theory (Tversky Kahneman, 1992) attempt to explain decision-making under risk. However, much of this research remains qualitative or theoretical, offering limited tools for directly modeling or quantifying human behavior in applied settings.

In this study, we propose a quantitative framework to bridge this gap by leveraging Blackjack as a controlled environment. Blackjack offers a simplified but representative model of decision-making under risk and uncertainty. By incorporating utility functions into reward determination for each state of the game, we aim to quantify the effects of different behavioral theories on financial outcomes - payoffs - through observing player actions. While the scope of our work is limited in terms of the range of utility theories explored and the variations of simulations ran, our goal is to establish a foundational framework that can be expanded upon in future research through exploring more complex utility functions and expanding the bounds of our simulation to collect more comprehensive data.

Literature Review

Optimal Policy in Blackjack

Blackjack is a well-studied problem in game theory and decision-making. Early research established the game's optimal policy under the assumption that players act as perfectly rational agents aiming to maximize their expected payoff (Thorp, 1962). This approach equates utility with monetary gain, ignoring the variability in individual preferences and subjective decision-making processes. Modern implementations of optimal blackjack strategies (e.g., counting cards or basic strategy charts) further reinforce this assumption of rationality, leaving little room for alternative models that incorporate behavioral variability.

Behavioral Economics and Prospect Theory

Theories like Prospect Theory (Kahneman Tversky, 1979) provide a more nuanced understanding of decision-making. Prospect Theory introduces the concept of a value function, which is concave for gains, convex for losses, and steeper for losses than for gains, capturing human tendencies like risk aversion and loss aversion. These insights are particularly relevant for modeling decision-making in blackjack, where players frequently face trade-offs between potential gains and

[1] Code used to create the game environment and run simulations can be found on our GitHub here

risks of loss. Similarly, Expected Utility Theory (Von Neumann Morgenstern, 1944) offers a framework for understanding rational choice, while deviations from it highlight the importance of subjective utility in real-world decisions.

Limitations of Existing Models

Despite their strengths, traditional theories often fail to account for the complexity and variability of human preferences. Most research in behavioral economics and decision theory remains theoretical, providing limited tools for empirical testing or application. For example, while Prospect Theory describes decision-making biases qualitatively, its practical implementation in dynamic environments like Blackjack is rare. This gap highlights the need for quantitative frameworks that can integrate these theories into actionable models.

Applications to Current Study

Our study builds on this body of research by using Blackjack as a testbed to explore and quantify utility functions derived from behavioral theories. Unlike traditional models that assume rationality and equate utility with monetary payoff, our approach incorporates variable utility functions to reflect different risk preferences and behavioral traits. This allows us to evaluate how utility theories influence decision-making outcomes in a controlled environment. By doing so, we aim to provide a practical foundation for future research that can expand this framework to include a broader range of decision-making models and applications.

Methodology

This project aims to model Blackjack as a simulation environment to investigate decision-making under risk and uncertainty, with a focus on how utility functions, influenced by individual risk preferences, impact strategic decisions. The methodology involves creating a Blackjack game modeled as a Markov Decision Process (MDP), running simulations using dynamic programming to generate optimal action at each state given varied utility function, and conducting a survey to validate the model's alignment with real-world human decision-making.

Game Model and Environment

For our simulation, *we define a Blackjack round as a series of Blackjack games played consecutively*, where the bankroll at the end of one game after adjusting for payoffs, becomes the initial bankroll for the subsequent round.

The Blackjack game is modeled as an MDP where the state space S is defined as the Cartesian product:

$$\mathcal{S} = \{P_c \times D_c \times k \times c \times m_b\}$$

Here the parameters represent the list of player cards (irrespective of suit and card order), the list of dealer cards (irrespective of suit and card order), the initial bankroll, the current bankroll, and the minimum bet. Thus, each state is uniquely represented as a combination of these parameters.

The action set consists of "Hit" and "Stand," and transitions between states depend on the player's actions and the probabilistic nature of the deck given the cards that have already been dealt. Terminal states occur when both the player and dealer bust or stand. The dealer follows a fixed policy, standing on 17 or higher and hitting on 16 or lower (McDermott, 2015). The game proceeds with the player making decisions until they reach a terminal state, followed by the dealer doing the same.

Key Assumptions

- **Independence of Utility**: We assume that the utility function depends solely on the initial and current bankroll, remaining unaffected by the game number or the earnings from previous games. This represents our most significant assumption, as research indicates that biases often arise from sequential dependencies.
- **Independence of Card Draws and Deck Reset**: We assume probability of drawing each card is uniform, and that the deck is reset after each game. This ensures there is no historic tracking or card counting across games
- **Action Set:** The available actions at each decision point are "hit" or "stand". If the player's hand value is less than 12, they always take a "hit".
- **Fixed Minimum Bet, Initial Bankroll:** The player's initial bankroll and bet amount is fixed at the start of the round. For consistent comparisons between our utility functions, we fix the initial bankroll and minimum bet to be the same for all simulations run.
- **Utility Functions:** The player's utility function is chosen based on their self-identified risk profile. Utility functions include risk-averse, risk-seeking, and risk-neutral profiles. The player utility function is denoted as: $U(k,x)$ where k denotes the initial bankroll, and x denotes the financial payoff, and the output $U(k,x)$ yields the utility or reward.

Utility functions are chosen based on the player's self-identified risk profile: risk-averse, risk-seeking, or risk-neutral.

Round Procedure

Each round begins by initializing the player with a utility function that reflects their risk profile. The dealer is fixed to follow the casino standard. The process for each round is as follows:

1. The player starts a round with an initial bankroll.
2. A game starts with the player receives two initial cards, and the dealer reveals one card (equivalent of one face up card)
3. The player turn begins as the player continues to make decisions, until they either stand or bust. Then, the game moves to the dealer states

4. The winner of the round is determined based on the outcome (win, loss, or draw), and the player's bankroll is updated accordingly. If the player runs out of money at this step, the round terminates.

5. The next game is initialized, and the process repeats until the round is complete

Simulations

We simulate 300 iterations of rounds, where each round involves running a Blackjack game using the Bellman equation to calculate the optimal action at each state visited during the simulation. For each card draw, we recursively evaluate the expected value $V(s)$ of each subgame by applying the Bellman equation:

$$V(S) = \max_a \left(R(S,a) + \gamma \sum_{S'} P(S'|S,a) V(S') \right)$$

where:

- S represents the state as is a function denoted by $S(k,x,P_c,D_c,m_b)$ where the parameters represent initial bankroll, current earnings, player cards, dealer cards, and minimum bet respectively.
- $R(S,a)$ is the reward function, representing the reward of taking action a in state S. This is computed as

$$R(S,a) = \begin{cases} U(k,x+m_b) - U(k,x) & \text{if player wins with } a \\ U(k,x-m_b) - U(k,x) & \text{if player loses with } a \\ 0 & \text{if action } a \text{ results in push} \end{cases}$$

  Note that here, the payoffs are only processed once the game reaches a terminal state and is 0 otherwise.
- $P(S'|S,a)$ is the transition probability of moving from state S to S' by taking action a.
- γ is the discount factor, which discounts future rewards. For the purposes of our paper, we set to 1.0 as the game reaches a natural conclusion within a few stages.
- $V(S')$ is the value function for the next state S'.

This process continues recursively to compute the optimal action for each state encountered during the game.

Survey

To assess how well our model aligns with real-world decision-making, we collect probability density functions (PDFs) for each of the five utility functions representing varying risk preferences (extremely risk-averse, moderately risk-averse, risk-neutral, moderately risk-seeking, and extremely risk-seeking). Participants are first asked to self-identify their risk profile and then choose the PDF that best represents their preferred playing style. This helps us match their self-identified risk preferences to the modeled risk assessments in the utility functions.

The survey results are used to validate the relevance of the utility functions in reflecting different risk preferences and to assess the overall fit of the model to real-world decision behaviors.

Results

The simulations were conducted using five different utility functions - Linear, Quadratic, Exponential, Logarithmic and Quartic. For each simulation, we ran 300 rounds where each round contained 15 games. Players started with a base bankroll of 500, with a bet size of 20, and played with exactly one deck in shuffle (relevant for computing probability of draw) [2].

Aggregated Statistics

We compute our aggregate metrics for each function from our simulation and tabulate our results below:

Utility Function	Mean Bankroll	Standard Deviation
Linear	489.27	60.32
Quadratic	497.33	64.65
Exponential	494.13	60.60
Logarithmic	497.33	62.77
Quartic	492.00	58.10

Table 1: Table of Utility Function Statistics

The mean bankroll values align closely with the expected long-term return for a player using optimal strategy. The house has an edge of approximately 0.5% in standard blackjack play, as found in the study (Thorpe, 1962). This yields an expected bankroll of 497.5 for the perfectly rational policy. Given that all but our linear function are optimizing utility not expected earnings, our mean bankrolls - that range from 489.27 to 497.33 - are intuitively explainable on average.

Interestingly, the standard deviation values are not as large as anticipated. Despite the quartic utility function typically modeling extreme risk preferences, it does not exhibit the highest standard deviation. Instead, **Quadratic** and **Logarithmic** utility functions show higher variability, suggesting that our sample size may not be large enough to capture the true risk-seeking behavior represented by the quartic function. This outcome hints at the potential need for more data to better reflect such risk-based utility in future studies.

It is surprising that the **Linear utility function** exhibits the lowest mean bankroll, despite modeling a "perfect" or balanced strategy. This could indicate that the linear utility, which models a risk-neutral player, is less suited to the blackjack environment than the other utility functions. A more risk-averse or risk-seeking utility (such as **Exponential** or **Quadratic**) might better capture the outcomes of real-world players who often exhibit such preferences in high-risk scenarios like blackjack.

[2] Code used to create the game environment and run simulations can be found on our GitHub here

Distributions

Here, we plot the distributions of the payoffs for each round in Figure 1. Through inspecting the curves we observe various trends.

Firstly, we see that the Quartic has the tallest peak with multiple local max points. This tells us the Quartic distribution is most volatitle, with the highest probability of doing marginally better than the baseline payoff, as well as the highest probability of doing marginally worse than the baseline payoff.

Focusing on the distribution tails after the 600 reward, we see that the linear curve has the lowest area, implying the lowest probability of ending the round with a payoff of 600 or greater. While it is expected to see the payoff lower to the Quadratic and Quartic functions, we see that it seemingly also is outperformed by the more risk averse curves.

As was expected, we see the Quadratic have the highest probabilities within about the $[550, 600]$ U $[400, 450]$ reward values. This is reflective of the Quadratic function's tendencies to take greater risks

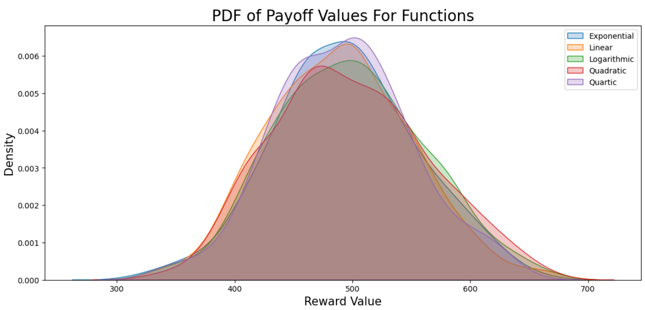

Figure 1: Distribution of payoffs

Cumulative Returns by Utility Function

Here, we plot the cumulative normalized returns for each function over the 300 rounds that we simulate. What we first see, as to be expected is that each function starts at about the same point. What we begin to notice after is that the curves continue to diverge towards points.

We see that the linear curve is the closest to a line, having the least fluctuations out of all the functions observed. What we do see is that round to round, the quartic has the most significant jumps - this is to be expected given the raised power of the function as compared to say the Quadratic function.

We also see that the Quadratic and Logarithmic functions appear to have the most variation overall, going through periods of increase followed by decrease. This is reflective of the sensitivity that these functions share in their variability from round to round. What would be further interesting to evaluate here, is the cumulative progression of returns by game as opposed to by the round.

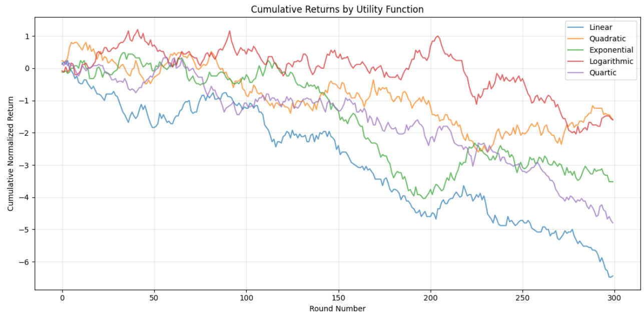

Figure 2: Cumulative Returns

Survey

We were able to gather self risk evaluation data for around 35 respondents. The aim of the survey was to analyze whether their self-identified risk preferences matched with the modeled risk assessment PDF's from the utility functions.

For the question **'How would you characterize your general risk tolerance when playing card games or gambling?'**

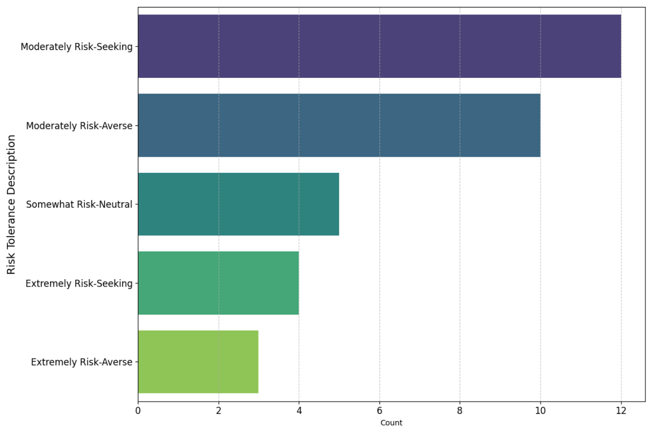

Figure 3: Respondents Self-Evaluation of Risk Tolerance

Most participants termed themselves closer to the central threshold of being moderate or neutral risk takers.

For the next question respondents were explained the game setting and then showcased the various risk utility functions and their PDF's following which they were asked **'Which risk profile would you characterize yourself as?'**

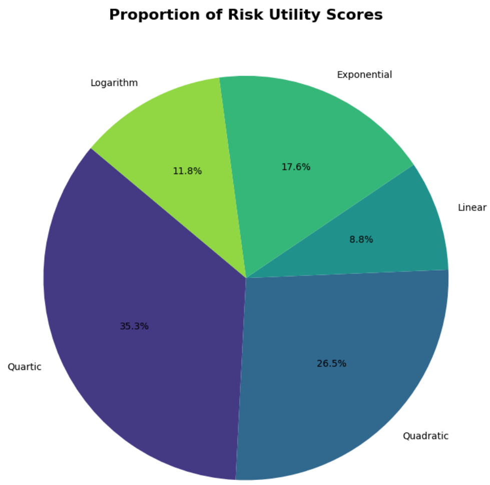

Figure 4: Respondents Results

The participants Risk Utility PDF selection is interesting that while people did not mostly categorize themselves as 'Extremely Risk Seeking', they have selected themselves highly in line with the Quartic utility function which is considered 'Highly Risk taking'. In further analysis, it was observed that the 'Moderately Risk Seeking' respondents were finding themselves relating to higher risk tolerance than they expected. 'Moderately Risk Averse' participants also classified themselves as a mix bag between Exponential utility function and Quadratic. The rest of the responses by the risk Profiles are in line with the other utility functions.

Discussion

Analysis of Survey in Connection with Other Findings

The survey results provide valuable insights into the alignment between individuals' self-reported risk preferences and the modeled risk profiles derived from the utility functions. Interestingly, while a majority of respondents identified themselves as moderate or neutral risk-takers, their selections of the risk profiles from the utility functions reveal a more nuanced understanding of their risk tolerance. The high alignment with the Quartic utility function, which models highly risk-seeking behavior, is particularly notable.

This suggests that many participants may perceive themselves as more risk-tolerant in practical gambling scenarios than they initially self-identified. We also speculate that participants/players do not operate with a simple categorical risk profile, but one that varies based on the scale of payoff, and current amassed payoff (current bankroll and initial bankroll in this case). That said, the discrepancy could also arise from the fact that our distributions showed greater volatility and 'risk' factor for the Quadratic simulations, as opposed to Quartic.

Strengths of the Paper

One of the strengths of this study lies in its detailed exploration of different utility functions and their implications on decision-making in a Blackjack environment. By modeling diverse risk preferences through utility functions such as Linear, Quadratic, Exponential, Logarithmic, and Quartic, we have quantified the varying degrees of risk tolerance that players exhibit in gambling contexts with uncertainty. This approach allows for a thorough analysis of how different risk attitudes impact overall performance, providing valuable insights for both theoretical and practical applications in decision-making research.

Additionally, the inclusion of survey data enriches the study by bridging the gap between theoretical models and real-world behavior. The combination of simulated gameplay with real participant data allows for a more comprehensive understanding of how different utility functions relate to actual human preferences, making the study more robust and grounded in human behavior.

Limitations

While the results of this study provide valuable insights, there are several limitations that must be acknowledged.

First, the sample size used in the survey was small, with only around 35 respondents. This limited sample size may not fully represent the broader population's risk preferences, which could affect the generalizability of the findings. Moreover, as explored in the Analysis of Surveys section, players likely deploy a variable risk profile that fluctuates based on the payoff levels, temporal factors etc., making it difficult to capture through a single metric. Furthermore, the self-reported nature of the survey data relies on participants' ability to accurately assess and articulate their own risk tolerance, which may introduce biases or inaccuracies in the results.

As for the simulations, we acknowledge that our data is limited, and that our analysis would have benefitted significantly from having more data. Moreover, our analysis would have strengthened from exploring various hyperparameter configurations for the functions, or performance at different initial bankroll values.

And finally, for the simulation setup itself. While the game was modeled as a deterministic environment with predefined utility functions, real-world Blackjack involves stochastic elements such as card shuffling and player decisions influenced by external factors (e.g., emotional states, fatigue). The absence of any consideration for bias, temporal factors, limits the effectiveness of the paper. That said, we acknowledge that we would not able to aptly consider all potential biases - the goal of our paper serves to understand what functions/conditions map closest to humans' true decision-making process comparitvely. It would be unrealistic to expect to model human gambling methods in its entirety.

Future Work

Several avenues for future work arise from this study. First, it would be valuable to explore additional utility functions that could better capture the nuances of human decision-making under risk. For example, incorporating functions that model diminishing marginal utility of wealth could offer deeper insights into how players balance gains and losses over time. Expanding the utility functions to include varying degrees of loss aversion or nonlinear risk preferences would allow for more realistic modeling of player behavior in Blackjack.

Additionally, expanding the scope of the game by introducing more complex betting strategies (doubling down, raising), multi-player dynamics, or incorporating elements such as splitting, Megabets present a more comprehensive view of decision-making under uncertainty. This would allow for further exploration of how utility functions affect strategic choices in less deterministic and more dynamic settings.

Finally, as opposed to collecting surveys that rely on self-assessment, the analysis would be strengthened significantly by providing participants the opportunity to play the game directly. That way, we would be able to more accurately and extensively capture their risk profile, along with having a mapping that our models could attempt to fit to, to quantifiably evaluate models that are closest to representing human behavior.

Conclusion

This study investigates the impact of different utility functions on decision-making in Blackjack, simulating various risk preferences and analyzing their effects on bankroll progression. The results highlight the importance of modeling risk attitudes, with risk-seeking behaviors being most pronounced in the Quartic utility function and risk-averse behaviors in the Exponential and Logarithmic functions. Survey data further reveals discrepancies between self-reported and actual risk tolerance, emphasizing the complexity of human decision-making under uncertainty. While the study provides valuable insights, future work could expand the utility functions, include more dynamic game settings, and collect additional data to refine the models and better capture real-world player behavior.

Acknowledgments

The foundations of the concepts for the work on this project have been taught by Professor. Todd M. Gureckis (New York University) who has been a great support throughout the implementation of this project.

References

Carlin, B. I., & Robinson, D. T. (2009, August). Judgment and decision making. *Judgment and Decision Making*, *4*(5), 385–396.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

K. Preuschoff, P. B., & Quartz, S. R. (2006). Neural correlates of expected risk and risk prediction error in economic decision-making. *Proceedings of the National Academy of Sciences*, *103*(42), 15841–15846.

McDermott, D. (2015). *The rules of blackjack*. Cardoza Publishing.

Neumann, J. V., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Sundali, J., & Croson, R. (2006). Fear and loathing in las vegas: Evidence from blackjack tables. *Journal of Behavioral Decision Making*, *19*, 123–134.

Thorp, E. O. (1962a). *Beat the dealer: A winning strategy for the game of twenty-one*. New York: Random House.

Thorp, E. O. (1962b). The optimum strategy in blackjack. *Transactions of the American Mathematical Society*.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Appendix

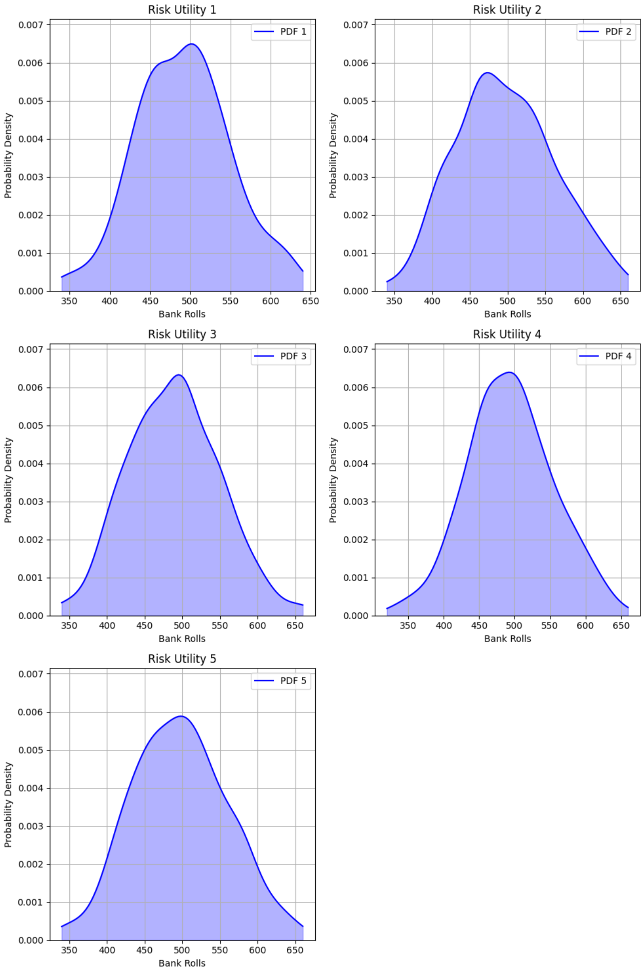

Figure 5: Risk Utility Functions PDF's showcased

Risk Utility Functions PDF's showcased to Survey Respondents for Q2. Utility function relation to the PDF was hidden from the responders to only get their blind review on which Risk Utility function they preffered and were most aligned with. The Risk Utility function in order are - Quartic, Quadratic, Linear, Exponential and Logarithm