Faculty of Sciences Departement of Computer Science

PREPARATORY WORK FOR THE MASTER THESIS

Fairness in decision-making under uncertainty using contextual multi-armed bandits

Author : Jean-Nicolas Grégoire

Promoter : Professor Tom Lenaerts (Université libre de Bruxelles)

Academic year 2024–2025

Contents

Chapter 1

Introduction

The collective decision of experts in a group can outperform the best expert in the group. But can it surpass another expert? In other words, is it optimal in absolute terms. This work explores decision-making under uncertainty, focusing on trade-offs between optimization and fairness.

We consider rational decision-making, where choices are based on logic and available information. In contrast, decision-making in some areas, such as politics, is often influenced by personal or hidden interests. Another approach is heuristic-based decision-making, which relies on mental shortcuts but is not the focus of this work. However, even rational decision-making is subject to biases, which must be accounted for and ideally minimized.

One of the challenges we face in our work is to be mindful of the prevalence of hypocrisy in today's world. That is to say, we must consider the increasing dissemination of false information on the internet.

Chapter 2

Introduction to Collective Decision-Making

2.1 Human decision-making

Each of us makes numerous decisions every day. Decision-making involves choosing among several possible actions in response to a problem, with the objective of resolving it by translating our choice into a series of actions. Cognitive psychology has demonstrated that purely rational approaches to decision-making do not adequately explain how we actually decide. The classical approach suggests that we make decisions through a rational analysis, calculating probabilities to identify the choice whose consequences best align with our interests. However, empirical research has revealed significant difficulties in processing probabilities accurately. We often commit errors in judgment due largely to limitations inherent in our cognitive systems. Kahneman and Tversky's work highlights that human decision-making is frequently non-rational, especially under uncertainty [Tversky and Kahneman, 1974]. Rather than relying strictly on logical reasoning, we tend to be guided by intuition and exhibit a pronounced aversion to loss. In uncertain situations, we often rely on heuristic rules. While heuristics are generally helpful in simplifying complex decisions, they can introduce biases into our reasoning. There are at least two methods for the human brain to make non-rational decisions: intuition and instinct. Intuitive decisions are based on experience [Klein, 2017, Gladwell, 2010]. Instinctive decisions are made "with our guts", but the exact cerebral mechanism involved is still unknown [Damasio, 2006]. Given these cognitive limitations, it logically follows that team-based decisions are likely superior to individual decisions. Indeed, empirical findings suggest that decisions made by small groups tend to yield better outcomes, on average, than those made by individuals alone[Navajas et al., 2018].

2.2 Algorithmic decision-making

Algorithmic decision-making follows strict rules; for instance, decisions can be made by formulating a problem as an optimization task, aiming to maximize or minimize a specific objective function. Algorithms can also incorporate uncertainty, making decisions based on probabilities and expected outcomes, as exemplified by Bayesian decision theory, Markov decision processes, and multi-armed bandits. However, algorithms lack intuition or instinctive decision-making capabilities. In this sense, unlike humans, algorithms are consistently rational. Nevertheless, problems leading to poor algorithmic decisions often arise from biases present in data or the assumptions underlying the algorithm itself. Algorithmic biases and human biases differ in their origins and manifestations. Human biases stem from individual experiences, cultural backgrounds, and personal prejudices, often operating unconsciously. In contrast, algorithmic biases arise from the data used to train machine learning models and the design choices made during algorithm development. For instance, if an algorithm is trained on historical data that

reflects societal inequalities, it may perpetuate those biases in its outcomes. However, unlike human biases, algorithmic biases can be systematically identified and corrected by analyzing and adjusting the underlying data and algorithms.

2.3 Collective algorithmic decision-making

Just as with human biases, researchers have investigated whether collaboration among different algorithms can enhance decision-making. This exploration led to the development of techniques known as ensemble methods. These methods involve combining multiple machine learning models to achieve better predictive performance than could be obtained from any individual model alone. Common examples include bagging, boosting, and stacking.The main objective of ensemble learning is to find an optimal performance through the combination of multiple classifiers. The ensemble learning a single or multiple algorithms is deployed to generate different base classifiers. These base classifiers are strategically combined together through a combination method of decision making in classifying new data instances. Since an ensemble is a combination of multiple methods, assessing the prediction of an ensemble requires a lot of computation compared to that of a single model.

2.3.1 Max Voting

In classification problems, each base model votes on the output, and the class with the majority of votes is selected. Each base model in the ensemble votes on the predicted class, and the class with the majority votes becomes the final prediction. This method works well when all base models have similar performance and contribute equally to the decision-making process.

2.3.2 Averaging

For regression tasks, the outputs of all models are averaged to produce the final prediction. This smooths out individual errors and produces a more accurate overall prediction. For instance, when predicting house prices, averaging the outputs of three different models can yield a more reliable result than relying on any one of them.

2.3.3 Bagging

Bagging creates multiple subsets of the training data by sampling with replacement. Each subset trains an independent base model, and their predictions are aggregated, typically through averaging, for regression problems, or voting, for classification tasks. Models are trained independently on different subsets of data and their predictions are averaged (regression) or voted on (classification). Random Forest is the most famous bagging technique. It creates multiple decision trees using random subsets of data and features, then combines their predictions. This approach reduces overfitting and increases model stability.

2.3.4 Boosting

Boosting tackles errors iteratively by training base models sequentially. Each new model focuses on correcting the mistakes of the previous ones. Techniques like AdaBoost assign higher weights to misclassified samples, ensuring that they receive more attention in subsequent iterations. The models are trained sequentially, each focusing on errors made by the previous ones. This reduces bias and improves predictions.

2.3.5 Stacking

Stacking combines predictions from multiple base models by using a meta-model, often a simpler algorithm like linear regression, to learn how to best combine them. Unlike bagging and boosting, stacking allows for heterogeneous base models, such as decision trees, support vector machines and neural networks, all working together. Predictions from multiple models are combined using another model (meta-model) to improve performance. This approach allows for using diverse base models, such as neural networks and decision trees, together.

Chapter 3

Modeling Collective Decision-Making in Contextual Multi-Armed Bandits

Chapter 4

The Importance and Measurement of Fairness in Decision-Making

Fairness in computer science has become an important topic, especially as numerous reports have highlighted concerns about algorithmic bias. Some experts argue that algorithms are so deeply affected by biases that their decisions cannot be considered fair or reliable. Consequently, they recommend rigorously evaluating AI decision-making systems with standards similar to those applied in pharmaceutical drug testing before approval for widespread use.(https://www.theguardian.com/technology/2019/dec/12/ai-end-uk-use-racially-biased-algorithms-noel-sharkey). Algorithmic bias occurs when systematic errors in machine learning algorithms produce unfair or discriminatory outcomes, often reflecting or reinforcing existing socioeconomic, ethnic, and gender biases. In this case, algorithmic bias is not caused by the algorithm itself, but by how the data science team collects and codes the training data. For example, United States courts use the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool to assess defendants' recidivism risk; but in this risk assessment, black defendants were twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism (https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm). Another example is the lack of fairness in the electronic health records of patients due to undergo coronary artery bypass surgery [**?**].

It is very difficult to estimate the degree of fairness because we want to achieve equal opportunities for all. However, the prevalence of the parameters we want to test is already variable at the outset. This prevalence will therefore have to be taken into account both before and after comparisons between the different groups. For example, let's take the case of atrial fibrillation, a disease that is more common in men than in women and more common in Caucasians than in blacks and Asians. Fair treatment will have to take into account the differences in prevalence before and after the decision is made by the algorithms. Note that our algorithm will also have to take positive discrimination into account if it is to be truly fair (https://www.michaelmauro.co.uk/thought-leadership/positive-action-vs-positive-discrimination?).

How do you measure if you model is fair? How do you decide which measure of fairness is appropriate? Subtlety 1 : Different groups can have different ground truth positive rates Subtlety 2 : your data is a biased representation of ground truth Subtlety 3 : it matters wether the model decision's consequences are positive or negative, en vertu de l'aversion au risque.

Chapter 5

Evaluating Fairness in Existing Contextual Multi-Armed Bandit Methods

We measure the performance using regret.

Chapter 6

Research Objectives

Chapter 7

Proposed Methods, Metrics, and Datasets for Fairness Evaluation

Meta-CMAB

Chapter 8

Definition to remember

8.1 Fairness definition and concepts

8.2 Approaches for measuring fairness

There is no universal means to measure fairness and no clear guidelines. Evaluating fairness in decision-making systems, including multi-armed bandits (MABs), requires a combination of statistical, algorithmic, and theoretical techniques.

8.2.1 Statistical Fairness Metrics

These metrics assess whether different groups receive equitable treatment.

- *Demographic Parity* (Statistical Parity): Ensures that the probability of receiving a positive outcome is the same across all groups.
- *Equalized Odds*: Ensures that both true positive and false positive rates are equal across groups.
- *Equal Opportunity*: A relaxed version of equalized odds, requiring only equal true positive rates.
- *Calibration*: Ensures that predicted probabilities are consistent across groups.

8.2.2 Individual Fairness

These methods check whether similar individuals receive similar treatment.

- *Lipschitz Fairness*: Ensures that similar inputs receive similar outputs.
- *Counterfactual Fairness*: A decision is fair if it remains unchanged when an individual's protected attributes (e.g., gender, race) are counterfactually changed

8.2.3 Fairness in Multi-Armed Bandits (MABs)

When fairness is considered in MABs, we evaluate:

- *Regret-Based Fairness*: Measures if certain groups suffer from higher regret.
- *Fair Exploration*: Ensures that all arms (e.g., demographic groups) are sufficiently explored before making a final decision.
- *Envy-Free Allocation*: Ensures that no group envies an other's received reward distribution.
- *Fair Reward Disparity*: Compares rewards received by different groups over time.

8.2.4 Group Fairness in MABs

- *Fairness-Constrained MABs*: Incorporates constraints to ensure groups receive equitable rewards.
- *Meritocratic Fairness*: Ensures rewards are proportional to the actual merit of the arms.
- *Fairness-aware Thompson Sampling or UCB*: Modifications to traditional bandit algorithms to ensure fair treatment across groups.

8.2.5 Long-Term Fairness

- *Temporal Fairness Constraints*: Ensures fairness over time rather than just in a static snapshot.
- *Cumulative Fairness Bounds*: Looks at fairness over a long-term horizon rather than immediate decisions

8.2.6 Metrics for binary classification

8.3 Databases

These datasets should help assess whether a model exhibits bias against certain groups based on characteristics like race, gender, age, or socioeconomic status.

8.3.1 General datasets

8.3.2 Fainess databases for MABs

8.4 Collective intelligence

The idea of collective intelligence refers to the enhanced decision-making ability that emerges from the collaboration, collective learning, or aggregation of multiple agents (humans, algorithms, or both). Instead of relying on a single optimal strategy determined by an individual, collective intelligence leverages the wisdom of the crowd to improve decision outcomes.

1. **Concepts in Collective Intelligence for Decision-Making:**
    - *Decentralized Knowledge*: Information is distributed among multiple agents instead of being controlled by a central entity.
    - *Diversity of Opinions*: Different agents may have unique perspectives or experiences that contribute to better decisions.
    - *Aggregation Mechanisms*: Methods such as voting, consensus algorithms, or statistical models combine multiple viewpoints into a single decision.
    - *Adaptive Learning*: Groups learn and evolve over time, refining their decision-making strategies based on past outcomes.
2. **Collective Intelligence vs Individual Decision-Making:**
3. **Collective Intelligence in Multi-Armed Bandits (MAB):** In the context of multi-armed bandits, collective intelligence can be applied by allowing multiple agents to participate in decision-making, rather than having a single decision-maker optimize rewards. This allows us to use multiple approaches.
    - *Majority Voting Bandits*: Multiple agents explore different arms and vote on the best action.
    - *Multi-Agent Bandits*: Agents share knowledge dynamically (e.g., federated learning-based bandits).

Feature	Individual Decision-Making	Collective Intelligence
Decision Basis	Optimized for a single agent	Aggregates multiple inputs for group benefit
Bias	Prone to individual biases	Reduces bias through diversity
Performance	Optimal for short-term goals	More robust for long-term decisions
Exploration-Exploitation	Exploits known best actions	Encourages exploration through diversity

Table 8.1: Comparison between Individual Decision-Making and Collective Intelligence

- *Fairness-Enforced MAB*: Decision rules prioritize group-level fairness over individual exploitation. For example, in an online recommendation system, rather than allowing a single AI to optimize for an individual user's engagement (which may reinforce biases), a collective intelligence approach could balance fairness by considering multiple users preferences before making a recommendation.

4. **Benefits of Collective Intelligence:** Increased Robustness: Reduces over-reliance on a single viewpoint. Fairer Decision-Making: Ensures diverse groups are considered in outcomes. Better Generalization: Leverages shared experiences to improve decisions. Resilience to Noise and Bias: Individual biases are reduced through aggregation.
5. **Challenges in Collective Intelligence:** Coordination Complexity: Requires effective mechanisms to aggregate decisions. Trade-off Between Fairness and Reward: Ensuring fairness may lead to suboptimal short-term rewards.

8.5 Experts

8.5.1 Reward-Based Experts

8.5.2 Fairness-Aware Experts

8.5.3 Reinforcement Learning (RL) Experts

8.5.4 Meta-CMAB expert

8.5.5 EXP4-IX expert

8.6 Bandits

A bandit problem is a sequential decision-making problem where a learner interacts with an environment over a series of rounds. In each round t, the learner selects an action $A_t$from a given set of possible actions A, and the environment responds by providing a reward X_t. The learner's objective is to maximize the cumulative reward over a fixed number of rounds n.

Mathematically, the problem arises because the environment is unknown, meaning the learner must balance **exploration** (gathering information about different actions) and **exploitation** (choosing the best-known action to maximize rewards). The **regret** is a key performance measure, defined as the difference between the total reward the learner could have obtained by always choosing the best possible action and the reward actually obtained by following a given policy.

Bandit problems are widely used in computer science and machine learning, particularly in reinforcement learning and online optimization. There are different types of bandits:

- **Multi-Armed Bandits (MABs)**: The simplest form, where each action (or "arm") provides rewards drawn from a fixed but unknown probability distribution.
- **Stochastic Bandits**: A case where rewards are generated from a fixed distribution, such as a Bernoulli or Gaussian distribution.
- **Contextual Bandits**: A more advanced setting where the learner observes contextual information before choosing an action, making it a stepping stone to full reinforcement learning.
- **Adversarial Bandits**: A setting where rewards are determined by an adversary rather than a stochastic process, making the problem significantly harder.

The competitor class plays an important role in defining optimal performance. It represents the set of policies against which the learner is compared. In some settings, the regret is measured relative to the best fixed policy, while in others, the best adaptive policy may be considered.

In summary, bandit problems provide a fundamental framework for learning under uncertainty, balancing the trade-off between exploration and exploitation. They have applications in recommendation systems, A/B testing, adaptive routing, and finally decision-making under uncertainty which will be the main topic of this work.

8.6.1 Contextual multi-armed bandits

Multi-Armed Bandits (MAB) and Contextual Multi-Armed Bandits (CMAB) are frameworks for sequential decision-making under uncertainty. While MAB focuses on learning the best actions solely from observed rewards, CMAB extends this by incorporating contextual information, allowing for more informed decision-making.

Multi-Armed Bandits (MAB)

In a standard MAB problem, a decision-maker, or learner, repeatedly selects from a set of arms, each associated with an unknown reward distribution. The objective is to maximize cumulative rewards over time while balancing exploration (gathering new information) and exploitation (leveraging known information to optimize rewards).

In MABs, decisions are made without any external context, the learner relies solely on past rewards to refine its strategy. Common algorithms used to solve MAB problems include:

- ϵ-*greedy*: Randomly explores with probability ϵ, otherwise exploits the best-known arm.
- *Upper Confidence Bound (UCB)*: Selects arms based on an optimism-in-the-face-of-uncertainty principle.
- *Thompson Sampling*: A Bayesian approach that samples from a posterior distribution to balance exploration and exploitation.

The most common example of an MAB problem is a gambler at a casino choosing between multiple slot machines (arms). Each machine has an unknown probability of payout, and the gambler must learn over time which machine offers the highest expected reward while minimizing losses due to poor choices.

Contextual Multi-Armed Bandits (CMAB)

CMAB extends the traditional MAB framework by introducing contextual information that influences reward distributions. At each decision point t, the learner observes a context vector x_t (e.g., user preferences, environmental conditions) and selects an arm based on both this context and past experiences.

The reward function is no longer fixed per arm but instead depends on the combination of the selected arm and the given context.

The goal in CMAB is to learn a policy:

$$\pi_t : \mathbb{R}^d \to \Delta([K]) \tag{8.1}$$

This policy will map contexts to probability distributions over the arms, enabling adaptive decision-making based on available contextual data. This is particularly relevant in real-world applications where decision-making does not occur in isolation but depends on additional information.

A practical example of CMAB is a news recommendation system that selects articles for users based on their demographic information, past reading behavior, and current interests. Unlike traditional MAB, where the same actions are evaluated independently, CMAB takes into account user-specific features to personalize recommendations dynamically.

Challenges and Considerations

One of the main challenges in CMAB is balancing exploration and exploitation while accounting for the high dimensionality of contextual information. Estimating the reward function $f(k, x_t)$ can require a significant amount of data, making early exploration costly in sensitive applications like medical diagnostics. For instance, choosing treatments based on insufficient patient data can lead to suboptimal or even harmful decisions.

To mitigate this, CMAB algorithms can incorporate expert advice or prior knowledge to guide learning, reducing the reliance on trial-and-error. However, expert knowledge itself can be biased due to cognitive biases such as outcome bias or external influences, making it crucial to develop mechanisms that adaptively aggregate expert opinions while minimizing their potential distortions.

An advanced extension of CMAB, known as bandits with expert advice, explicitly models this setting by allowing the learner to integrate multiple sources of information and refine its strategy over time, shifting the challenge from purely estimating f to optimally aggregating expert knowledge.

8.6.2 Algorithms for Contextual Multi-Armed Bandits

Linear Upper Confidence Bound (LinUCB)

The Linear Upper Confidence Bound (LinUCB) algorithm is a widely used approach for contextual multi-armed bandit (CMAB) problems. It extends the classical Upper Confidence Bound (UCB) algorithm by incorporating contextual information to make more informed decisions.

In this kind of multi-armed bandit setting, UCB selects arms based on an upper confidence estimate of their expected rewards, balancing exploration and exploitation. LinUCB builds upon this idea by assuming that the expected reward of an arm follows a linear model:

$$\mathbb{E}[r_t|x_t, a] = x_t^T \theta_a \tag{8.2}$$

where x_t is the d-dimensional context vector, θ_a is an unknown parameter vector specific to arm a, and r_t is the observed reward. Since θ_a is initially unknown, LinUCB estimates it from past observations using ridge regression. Finally, To select an arm, LinUCB computes an upper confidence bound for each action:

$$\hat{r}_a = x_t^T \hat{\theta}_a + \alpha \sqrt{x_t^T A_a^{-1} x_t} \tag{8.3}$$

where $\hat{\theta}_a$ is the estimated parameter vector, A_a is a covariance matrix tracking feature information, and α controls the exploration-exploitation trade-off. The algorithm selects the arm with the highest upper confidence bound.

Contextual Thompson Sampling (CTS)

Contextual Thompson Sampling (CTS) is a Bayesian approach to solving the CMAB problem. It extends the standard Thompson Sampling algorithm by incorporating contextual information when making decisions. Unlike deterministic strategies, CTS maintains a *posterior distribution* over the unknown parameter θ, which represents the relationship between context features and rewards.

At each time step t, the algorithm samples a parameter estimate from the posterior distribution:

$$\theta_t \sim \text{Posterior}(\theta|\text{past rewards}) \tag{8.4}$$

Given this sampled parameter, CTS selects the arm a that maximizes the expected reward under the sampled model:

$$a_t = \arg\max_{a \in [K]} x_t^T \theta_t \tag{8.5}$$

Once the reward is observed, the posterior is updated using Bayesian inference, refining the model's understanding of the reward distribution.

Neural Contextual Bandits (Neural UCB / Neural TS)

Neural contextual bandits extend traditional contextual bandit models by replacing the typical linear reward model with a deep neural network. Instead of assuming that rewards are linearly dependent on features, these models learn a more flexible function:

$$r = f(x, \theta) + \epsilon, \tag{8.6}$$

where $f(x, \theta)$ is a neural network parameterized by θ, and ϵ represents noise. This approach allows for modeling complex, non-linear relationships between context x and rewards, which is particularly useful in settings where linear assumptions fail to capture the true reward structure.

The two main algorithms in this category, Neural Upper Confidence Bound (Neural UCB) and Neural Thompson Sampling (Neural TS), adapt standard exploration strategies to neural network models.

- **Neural UCB** estimates uncertainty using a confidence bound based on the neural network's predictions. The algorithm constructs an upper confidence bound on the reward for each arm and selects the arm with the highest upper bound. Since neural networks do not provide inherent uncertainty estimates like Gaussian processes, techniques such as bootstrapped ensembles, last-layer Laplace approximations, or Bayesian linear regression on the final layer are often used to quantify uncertainty.
- **Neural TS** follows the Thompson Sampling framework, where uncertainty is modeled probabilistically. The algorithm samples parameters from a posterior distribution over the neural network's weights and selects actions based on these sampled parameters. Similar to Neural UCB, the challenge lies in estimating uncertainty, which is often handled using variational inference, dropout-based Bayesian approximations, or deep ensembles.

8.6.3 Random Forest Bandits

Random Forest Bandits use an ensemble of decision trees to estimate rewards and guide exploration. Instead of modeling the reward function with a single function (linear or neural network-based), this approach utilizes an ensemble of decision trees, forming a random forest. Each tree in the forest provides an independent estimate of the reward, and the bandit algorithm selects arms based on an exploration

strategy such as UCB or TS again. The first one estimates an upper confidence bound by computing the mean and variance of the reward estimates across trees. The model selects the arm with the highest upper bound, allowing for exploration in areas with high uncertainty. While the other strategy provides a posterior-like distribution over rewards, from which samples are drawn to determine which arm to select probabilistically.

8.6.3.1 Comparison of CMAB Algorithms

Algorithm	Assumption	Exploration	Pros	Cons
LinUCB	Linear rewards	UCB	Fast, scalable	Fails for non-linearity
CTS	Bayesian prior	TS	Adaptive exploration	Computational cost
Neural UCB/TS	Deep learning	UCB or TS	Handles non-linearity	Data-hungry
RF Bandits	Decision trees	UCB or TS	Captures structure	Expensive

Table 8.2: Comparison of CMAB Algorithms

8.6.4 Bandits with experts advice

8.6.4.1 Popular Approaches

8.6.4.2 Follow the Best Expert

8.6.4.3 EXP4: Exponentially Weighted Average Forecaster

8.6.4.4 Hedge Algorithm (Weighted Majority)

8.6.5 Advantages of Bandits with Expert Advice

8.7 Stacking

We are going to use Stacking, which is a machine learning method that combines several machine learning models to improve predictive performance [Wolpert, 1992]. Unlike bagging (e.g., Random Forest) and boosting (e.g., XGBoost), which focus on reducing variance and bias, respectively, stacking aims to learn an optimal way to aggregate model predictions. The principles of stacking are as follows, it uses two layers to perform its analysis. The First layer (base learners) includes several base models, such as regressions, decision trees, SVM. which are trained independently on the same data. The Second layer (meta-learner) is a higher-level model (meta-model) trained on the predictions of the First layer models to produce the final prediction. Finally, stacking enhances predictive performance by leveraging model diversity, reducing overfitting, and improving generalization. However, its direct application to bandits with expert advice is challenging since the expert set is predefined, and decision-making must balance exploration and exploitation rather than just optimizing prediction accuracy.

8.8 Biases

The concept of bias in decision-making was significantly shaped by Daniel Kahneman and Amos Tversky in *"Judgment under Uncertainty: Heuristics and Biases"* [Tversky and Kahneman, 1974]. Their work demonstrated that people rely on mental shortcuts, or heuristics, which often lead to systematic errors in judgment. People use heuristics (mental shortcuts) to simplify decision-making under uncertainty. These heuristics, while useful, introduce predictable biases. Kahneman and Tversky refined their model to explain decision-making under risk. Losses are more psychologically impactful than equivalent gains (loss aversion).

Researchers like Richard Thaler integrated biases into [Leonard, 2008]. Biases explain anomalies

in markets and consumer behavior. Bias research expanded beyond decision-making into social psychology. Studies on implicit bias reveal subconscious prejudices in hiring, law enforcement, and AI systems.

Collective decision-making based on expert groups aims to reduce individual biases, but it is not immune to them. Cognitive biases can still shape group decisions, influencing judgment in ways that may lead to systematic errors. Cognitive biases influence our decisions by distorting our perception of reality, often unconsciously. These are systematic thinking errors that occur when our brain processes information in a simplified way to cope with the complexity of the world. Here are some common cognitive biases and their impact on decision-making [Pohl, 2004] :

8.8.1 Confirmation bias

This bias pushes us to seek out, interpret and memorize information that confirms our pre-existing beliefs, while ignoring or downplaying information that contradicts them. This can lead to biased decisions because we rely on partial data and reject alternative perspectives. For example, we could imagine that an investor, convinced that a company's shares will increase, will only pay attention to positive news and ignores signs of decline, potentially resulting in poor financial decisions.

8.8.2 Availability bias

This bias occurs when we base our decisions on information that is easily accessible in our memory, often recent or striking, rather than on an objective analysis of all available information. This can lead to distorted risk assesment. After seeing numerous reports on plane crashes, for instance, a person might overestimate the risk of flying, even though statistically, air travel is one of the safest modes of transportation.

8.8.3 Anchoring bias

Anchoring occurs when we rely too much on the first piece of information received (the "anchor") to make a decision. Even if this information is arbitrary or incorrect, it strongly influences our judgments. In a salary negotiation, for example, if the initial offer is very high or very low, it sets a reference point that skews further discussions, even if the fair value should be different.

8.8.4 Status quo bias

This bias makes people prefer maintaining the current state of affairs rather than considering alternative options, often due to fear of change or cognitive effort. This can lead to passive decisions that neglect more advantageous options. A person, for instance, may stay in an unsatisfactory job simply because change seems risky to them, even if they have the possibility of a better paid and more fulfilling position elsewhere.

8.8.5 Framing bias

Decisions can be influenced by the way a situation or an option is presented, even if the underlying facts remains the same. Choices may vary depending on whether the information is presented in a positive or negative light. A medical treatment may be considered more attractive if it is said to have a "90% chance of success" rather than a "10% chance of failure", even though both statements are statistically identical.

8.8.6 Optimism bias

This bias leads us to overestimate the likelihood of positive results while underestimating risks. This can lead to unwise decision-making, particularly in contexts where there is a high degree of uncertainty.

An entrepreneur might, for instance, overestimate the chances of success of his project and neglect the financial risks, believing that the failures of others will not apply to him.

8.8.7 Loss aversion bias

People tend to be more motivated to avoid losses than to achieve equivalent gains, which can lead to overly conservative or irrational decisions. For example, an investor might refuse to sell a declining stock, hoping it will recover, even when an objective analysis suggests cutting losses would be the better decision.

8.8.8 Halo effect

The halo effect consists of judging a person or a situation in a generally positive or negative way based on a single characteristic. This can influence decisions in a biased way. In hiring decisions, a charismatic and well-dressed candidate may be perceived as more competent, even if another candidate has objectively better qualifications.

8.8.9 Authority bias

People tend to place excessive trust in authority figures, often without critically evaluating their arguments or evidence. A patient might accept a treatment proposed by a doctor without question, even if there are more suitable or less risky options.

8.8.10 Group effect (or groupthink)

The desire for group harmony can lead individuals to conform to majority opinions, sometimes to the detriment of their own critical judgment, in order to avoid conflict or to conform. For example, in a team meeting, a person might approve a questionable strategy simply because the majority supports it, even if they have doubts in private.

In conclusion, cognitive biases profoundly shape our decisions, influencing how we perceive and process information. In expert-driven decision-making under uncertainty, these biases can propagate from individuals to the collective, potentially leading to suboptimal outcomes. Recognizing these biases is the first step in mitigating their impact. Structured decision-making processes, bias-awareness training, and algorithmic interventions can help reduce the influence of biases in expert-driven collective decisions.

8.8.11 Algorithmic and AI Bias

Machine learning models inherit biases from human data. Fairness in AI is now a major research area, using methods like multi-armed bandits to reduce bias in decision-making. Critics argue biases might not be errors but adaptive strategies for real-world decision-making. Modern research explores ecological rationality, where heuristics are seen as optimized for specific environments.

Chapter 9

Objectives

Our work focuses on fairness in decision-making using bandit algorithms, a common framework in machine learning for sequential decision-making. Specifically, we investigate bandits with expert advice (BwEA) and bandits with regression oracles (BwRO)—approaches that rely on expert recommendations to optimize decisions over time.

Appendix A

Annexes

Bibliography

[Damasio, 2006] Damasio, A. R. (2006). *Descartes' error*. Random House.

[Gladwell, 2010] Gladwell, M. (2010). *Blink: The power of thinking without thinking*. Hachette Audio.

[Klein, 2017] Klein, G. A. (2017). *Sources of power: How people make decisions*. MIT press.

[Leonard, 2008] Leonard, T. C. (2008). Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness: Yale university press, new haven, ct, 2008, 293 pp, 26.00.

[Navajas et al., 2018] Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., and Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, 2(2):126–132.

[Pohl, 2004] Pohl, R. (2004). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Psychology press.

[Tversky and Kahneman, 1974] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

[Wolpert, 1992] Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.