Issue Brief

April 2025
No : 430

Human Biases and Errors: Impacting Military Decision Making & Implications for AI Based Decision Support Systems

Col Nachiket Kodilkar

Human Biases and Errors: Impacting Military Decision Making and Implications for AI Based Decision Support Systems

Col Nachiket Kodilkar

Abstract

Artificial Intelligence is being increasingly incorporated in various Military Decision Support Systems. To ensure accuracy and reliability of such systems, it is important to understand the effect of Human Biases and Errors on AI based Decision Support Systems (AIDSS) and take measures to obviate their negative effects. This aspect is especially critical for military decision making, as minor errors can result in catastrophic consequences. This Issue Brief analyses Human Biases and Errors affecting Military Decisions and their likely impact on AI models employed in AIDSS. The insights in this Issue Brief will facilitate creation of more effective AI models to power Military Decision Support Systems.

Keywords: Artificial Intelligence, Military, Decision Making, Human Biases

Background

Decision making is a complex activity which involves consideration of multiple external (situational) and internal (cognitive) factors. Quality of decisions is influenced by multiple factors viz. information available, capabilities to process the information, ability to separate 'noise' from 'facts' and ability of putting into context collated information in order to derive optimal solution for the issue at hand.

Obtaining complete information is never possible, hence it is imperative for any decision maker to take an optimal decision under conditions of uncertainty or partially available information. The same is more important for a military decision maker, as he/she has to operate in an environment which is Volatile, Uncertain, Complex and Ambiguous (VUCA).

In such a scenario, it is important that 'Human biases' of the decision maker and errors in decision making do not distort the information processing, resulting in sub-optimal decisions. This issue is also relevant during Artificial Intelligence (AI) modelling, as cognitive biases of the programmers or the biases in data can affect the output of the AI System.

A lot of research has been done in ascertaining the impact of these biases in the civilian domain, however, research is limited in ascertaining impact of human errors and biases in military decision making. This Issue Brief will attempt to ascertain the impact of human errors and biases in military decision making and recommend measures for modelling AI systems being employed for the same.

There are numerous cognitive biases and errors. The ones which impact military decision making the most are discussed in this Issue Brief.

Anchoring Bias

This bias indicates the tendency to 'rely too heavily on one data point' while making decisions, that is we tend to anchor our decision to the chosen data point which can be :-

- The first piece of information acquired.
- Data believed to be obtained from a reliable source.
- Data that reinforces the decision makers' intuition / pre-conception.

Once a decision maker anchors his decision to the chosen anchor data point, it is difficult to change the decision, even when new data / evidence is provided. Also, the change in the decision from the anchor point is minimal and incremental, even if the data indicates otherwise.

A famous example from military history is that of Normandy landings in World War II. The allies launched 'Operation Bodyguard'— an elaborate deception plan to deceive the German High Command with respect to the time and location of allied invasion of North Western Europe. The operation focused on reinforcing the German belief that the invasion is likely through Pais de Calais rather than Normandy. It was a classic utilization of Anchor bias, wherein the German High Command opinion with respect to likely Allied invasion was 'anchored' to Pais de Calais and despite intelligence inputs to the contrary, the German High Command could not re-orient their forces towards Normandy (English Heritage).

Implications for AI Models

The first implication is with respect to Value Function used in an AI based Decision Support System (AIDSS). If weightage assigned to few factors is more than the rest, decision of the AI model will get 'anchored' to these dominant factors. In case the relative importance of 'these' decision factors change, then it will be difficult for the AIDSS to adapt to the new situation fast.

Training data initially supplied to the AI model is more likely to shape the output of the AI model, compared to data supplied later. For example, supposedly, AIDSS is fed with 100 data points on number of artillery rounds required to be fired to degrade a particular enemy locality. Hypothetically, with these data points, AIDSS figures that 300 rounds are required for requisite degradation. If, 100 more data points are fed (which suggests that 500 rounds are required for requisite degradation), it will be difficult for the AIDSS to revise its output from 300 to 400 (average of 300 and 500) rounds due to 'Anchoring Bias' to initial data.

This bias would be more prominent in AI models adopting gradient descent algorithms. Let us presume for a moment that programming builds impartiality in the AI model, wherein successive data also gets equal weightage compared to initial data. In such a scenario, the model will be sensitive to noise or errors implying that the model will again not be accurate. The point to highlight is that there is a trade-off in accuracy for mitigating anchor bias.

Apophenia

Apophenia is the 'tendency to perceive meaningful connections between unrelated things' (Ness Labs). The bias stems from 'co-relation without causality'.

As a generally accepted military principle, an attack which has failed is not reinforced. However, there are multiple cases where failed attacks have been reinforced and success has been achieved. The success could be attributable to other factors such as effective degradation of the enemy or better logistics backup etc. Multiple examples with co-relation between reinforcement of failed attacks and success may lead us to believe that it is a successful military tactic.

Implications for AI Models

When AI models are based on regression, causality needs to be kept in mind. At times, there may be very high co-relation between variables, but military decisions may go wrong if causality is not correctly assessed.

AI models are trained using Loss Functions. If all independent variables, affecting the dependent variable, are not taken into account, it may lead to incorrect interpretations with respect to causality thereby incorrectly affecting future decisions.

Few events are actually random or are non-deterministic. Any tendency to derive meaning or inferences from such events can lead to wrong training of AI models. A rudimentary example is the Battle of Midway during World War II, wherein an American scout plane 'accidently' discovered the Japanese Aircraft Carriers, when they were most vulnerable (their planes were refuelling and rearming). Without this discovery, it would not have been possible for US to destroy the Japanese Naval Fleet (Parshall and Tully, 2007). Hence, attributing success of US in Battle of Midway to any other factor may not derive correct lessons as it was the chance event which had maximum impact on the outcome in this particular battle.

Availability Heuristic

This heuristic is the tendency to base decisions on available information. Implicit in this fact is the 'non-consideration' of information that is not available but may have an impact on the quality of the decision taken.

Majority of military decisions are based on VUCA conditions and military commanders are trained to seek information that is critical for their decision making rather than passively relying on available information. The process of 'Intelligence Preparation of Battlefield', has been designed to seek critical information that is lacking for effective decision making. Though, decisions will still have to be taken in situations where information is not complete, when there is consideration of non-available information in decision making, risk assessment with respect to selected decision is better.

A subset of Availability Heuristic is the 'Survivorship Bias' which indicates availability of biased data due to 'survival of a particular class of sample population'. Data which would have been provided by people who did not survive an event / process is inadvertently overlooked. Let us say, a study aims to determine the co-relation between firing standards of soldiers and probability of their survival in conflict. Such a study will not provide effective results due to survivorship bias as soldiers who are alive but not good firers will also contribute to the sample data but good firers who laid down their lives will not form part of the sample population.

Implications for AI Models

Survivorship bias results in 'overfitting of data'. AI models are trained on data that is available, utilizing the 'Loss Function'. Programmers attempt to minimise the loss function when back tested with available data. This results in AI models being able to output efficient decisions based on training data but are less efficient while dealing with real world scenarios where non availability of requisite data due to survivorship bias can result in sub-optimal decisions.

AI models are not trained to ask— which data am I missing to make a holistic decision? To address the availability heuristic issue, AI models need to be proficient in General AI. AI models need to be able to identify the decision sought and all the factors contributing to that decision. This is a key requirement to address the availability heuristic. Any information that is missing, needs to be actively sought by the AI model in order to make an informed decision. It is felt that there is lot of scope for further research in this field.

Baader–Meinhof Phenomenon

This phenomena states that, once something has been noticed, then that thing/event is noticed more often, leading to the belief of 'higher than normal' frequency of occurrence of that event. For instance, if a friend tells us about a new car launched by a company, we will start noticing that car on the road more often. We may have come across that model of the car earlier, but now instances of noticing that particular car would be more often. In military decision making, this phenomenon affects analysis of intelligence inputs and strategic assessments.

As an example, some military strategist established a corelation between Chinese Game of 'Go' or Weichi' and China's actual foreign engagements thereby coining terms such as 'strategic encirclement', when, in reality, there may or may not be any such intent of China to learn from the game of 'Go' and execute 'encirclement'. This co-relation, having 'noticed' by later strategic thinkers, have been replicated in numerous Western Publications.

Implication for AI Models

A very interesting manifestation of Baader-Mienhof phenomena is the 'compounding effect' which is noticed in algorithms used by social media websites. Watching a particular genre of videos on Youtube results in similar videos being recommended by the algorithm resulting in a non-linear compounding effect, wherein majority of the content consumed by the user belongs to a narrow class of genres, unless the user actively searches for a more diverse content. Similarly, algorithms on e-commerce websites may result in disproportionate sales of a particular item due to such compounding effect, when, in reality, there may be very little difference between the 'most popular product' and its closest competitor.

For AI algorithms, which are based on supervised learning, this phenomenon comes into play. Consider an AI algorithm trained on analysing satellite images; the algorithm is likely to report higher occurrences of those enemy entities which it has learnt to identify /classify than those which it has not learnt to identify/classify. In order to mitigate the effect of this bias, AI models, especially those dealing with Intelligence, Surveillance and Reconnaissance need to incorporate both supervised and unsupervised learning and may have to be programmed to identify new unlabelled objects and entities as well.

Salience Bias

Salience Bias is the tendency to focus on items that are more prominent or emotionally striking. A prominent example from World War II is the 'Battle of Stalingrad'. The plan to capture Stalingrad was not strategically sound, however Hitler insisted on the same, as the city was named after Joseph Stalin, the then General Secretary of the Communist Party of Soviet Union – clearly a Salience Bias.

In military operations this bias is exploited in the form of Cliff Chop Assault, wherein a large quantum of force assaults a defensive position in mountains along relatively gradual slope. While the defender concentrates on defending against this 'prominent' attack, a small body of troops assaults from an unexpected direction.

Implications for AI Models

Salience Bias will be more prominent in computer vision systems. Aspect of 'attention' is a pre-requisite to enable computer vision systems to correctly interact with the environment. The system needs to identify which object to focus on. In the absence of attention mechanism, all objects in the vision field will get equal weightage and may overwhelm the computer vision system with too much data. The collateral cost of attention mechanism is

Salience Bias, wherein an object, that is more prominent, would grab the 'attention' of the computer vision systems. Similarly, while considering search algorithms, data with higher frequency / prominence would affect the analytics of the AI model. Data which is subtle may get ignored as the value function of the algorithm is likely to get biased by 'prominent data'. AI models thus need to be programmed to identify subtle but important data elements for effective military decision making.

Confirmation Bias

Confirmation bias is the tendency to seek, interpret or focus on information that confirms one's existing preconceptions. This bias can be placed under the category of biases that results from the egocentricity of humans (How can I be wrong?). Ego in itself is a reward mechanism, which motivates humans to make efforts towards achieving their goals. For neural networks, where there is no incentivisation for the AI model to do its work better—ego based biases may not come into play, however, as we progress towards Artificial General Intelligence, the aspect of 'ego' will have to be built into ML algorithms and that is when, 'ego based biases' will start affecting the AI system outputs.

Illusion of Validity

This bias is subtly different from Confirmation Bias. It is the tendency to overestimate the accuracy of one's judgements, especially when incoming information is consistent with the judgement. This bias is especially detrimental in military decision making, as military planners have to deal with 'Black Swan events' i.e. ones which have very low probability of occurrence, cannot be foreseen, but have very high impact.

Illusion of validity may lead military planners into a sense of false security thereby resulting in inadequate preparation when the black swan event actually occurs. Post independence, India did not anticipate any conflict with China, resulting in comparatively lesser military preparedness. Diplomatic parleys till late 1950s also supported this judgement. Thus, we were not adequately prepared for the 1962 conflict. Despite this fact, we were again inadequately prepared in April 1971. Historically, the accuracy of the adage 'diplomacy can avert war' has been overestimated on multiple occasions due to Illusion of Validity.

Implications for AI Models

While estimating the statistical confidence level of its assessment, the AI model needs to look at not only the probability of events but also its impact. The confidence level assigned by the AI model needs to be more conservative where the impact of the assessment is high, so that the military planners can undertake risk mitigation accordingly. Consider an AI model assessing Enemy Courses of Action (COA) as an example. If the AI model assesses a particular enemy COA as a low probability event with a very high confidence level, the model should moderate the confidence level to a lower number if this enemy COA is likely to have significant impact on success /failure of our mission. This will ensure that this enemy COA is not discounted by the military planners. This aspect is already built into the military appreciation process by consideration of 'most dangerous enemy COA' and needs to be suitably dovetailed while programming AI models.

This bias reinforces the requirement of 'Explainable AI' in military decision making. The AI model should be able to give out the rationale for making a particular decision. This will help military planners to decide how much they can rely on the decision/analysis of the AI model.

Bias Blind Spot

This is a 'bias which makes us blind to own biases'. We perceive that we are observing and assessing a given situation objectively, whereas in reality, our biases may be affecting our decision making.

Implications for AI Models

AI algorithms give outputs based on probability. For classification tasks, where binary choice is based on a probability threshold, bias blind spot may result in the algorithm predicting outcomes with higher degree of certainty than it actually is. Let's take an example of a satellite image analyser which predicts deployment of artillery guns. In case enemy has deployed dummy guns, the algorithm will still predict deployment of guns with very high confidence level. It will be very difficult to teach the algorithm to question its own assessment and for it to be more amenable to its confidence level, while dealing with issues of surprise and deception, which are a common phenomenon in military operations.

Similar issues are likely to be faced with AIDSS trained on past conflicts, wherein the AI System may base its decisions on assessment of past conflicts and may be oblivious to enemy tactics and strategies which are novel or which have not been orchestrated in past conflicts.

Sample Size Estimation Error

This class of biases implies errors in judgement resulting from incorrect consideration of the sample size of the observations. Biases and errors under this category and its resultant impact on AI models is as follows:

Base Rate Fallacy : Base rate fallacy or base rate neglect is the 'tendency to ignore the overall statistics and base one's judgement on a specific occurrence or instance'. This fallacy generally works in conjunction with recency bias resulting in more weightage being given to recent observations. This fallacy is likely to impact weightages that are assigned to various factors in a Military AIDSS. Employment of tanks at Zoji La in 1948 was a novel manoeuvre which contributed to victory. An AIDSS trained on this data would assign a high weightage for employment of tanks in restricted terrain astride a mountain pass. However, employing the same manoeuvre in future wars, without consideration of other factors, would result in Base Rate Fallacy (In how many wars have tanks been successfully employed in restricted terrain? Relying on assessment derived from a single or a small sample size may not be correct). Employment of tanks at a narrow pass, in today's era of pervasive surveillance and new age weaponry such as sensor fused weapons / Unmanned Combat Aerial Vehicles (UCAVs), would be suicidal.

Hyperbolic Discounting : Hyperbolic discounting is the 'tendency to have a stronger preference for more immediate payoffs (even if smaller) relative to later payoffs (even if larger)' (Renascence). A 'quick attack' i.e. an attack on the enemy's hastily prepared defences is likely to facilitate capture of enemy positions in an earlier time frame. However, the captured positions are liable to be vulnerable to enemy's counter attacks. Unless requisite logic is built into AIDSS, it may recommend a quick attack in majority of the situations as compared to a deliberate attack due to hyperbolic discounting. This bias is being countered in AI algorithms by doing a depth first tree search till all possibilities are considered, thus also taking into consideration payoffs in future. With this, AI algorithms for playing Chess or the game ‘Go’ have been able to identify maverick moves far ahead in the game. However, where AIDSS have to consider options in a non-deterministic, partially observable universe, the algorithm is more likely to be affected by hyperbolic discounting compared to AI algorithms behind chess and ‘GO’ engines where outcomes are deterministic and complete information of opponent moves is available.

***Creeping Normality** :* Creeping Normality refers to the ‘tendency to overlook small incremental changes until they become significant’. This cognitive effect has been effectively exploited by China in her international strategy which is termed as ‘Salami Slicing’. It implies incremental aggressive activities below the threshold, such that they go unnoticed, but result in considerable gain over a prolonged duration. AIDSS will have to be trained to look for incremental actions by the enemy aimed at surprise and deception. The incremental changes may be in preparation of defences, build-up of forces, capability development of new weapons / equipment, logistics buildup etc.

Logical Fallacies

Logical Fallacies occur where there is error in reasoning. It is an important class of biases which have considerable real world implications. These biases are subtle but have considerable impact on decision making. They are discussed in succeeding paragraphs.

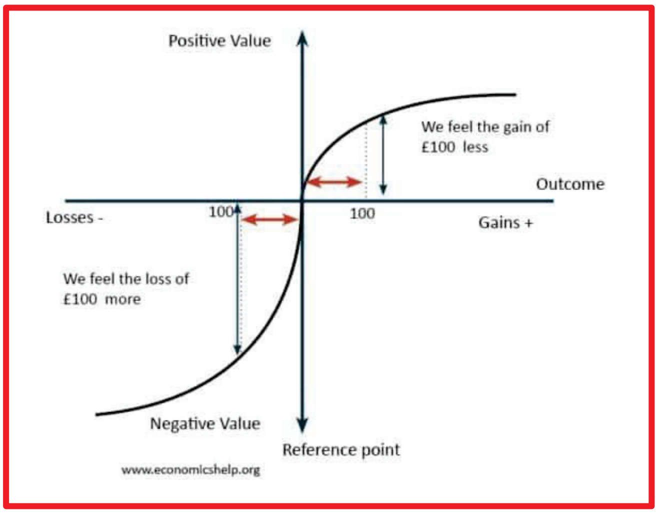
Sunk Cost Fallacy & Gambler’s Fallacy : Sunk Cost Fallacy is a situation where people justify/support a decision taken, based on total prior investment of efforts or resources, despite new evidence suggesting that the decision was probably wrong. In military parlance, this bias results in the cardinal mistake of reinforcing a failure. Let us say that a military commander chooses COA 1 compared to COA 2 or COA 3. Even if COA 1 results in reverses, the military commander thinks that they are temporary, and that, since adequate resources are already employed in executing COA 1, allotting few additional resources will result in success.

At what point of time does one realise that the chosen decision is wrong? Since the cost benefit analysis does not output a discrete number, but a plot on a continuum, it is difficult to nominate a threshold in human decision making, after which, wrong decision is negated and the correct one is taken. In AIDSS, there is a feasibility of using an activation function with pre-fed thresholds, akin to functions employed in binary classification tasks.

Similarly, when a value function output dips below an identified threshold, earlier incorrect decision can be discounted and a new COA can be executed. Another issue which comes up and is more subtle is with respect to the value function itself. Since more resources have been utilised in COA 1 as compared to COA 2, tendency is to believe that COA 1 is 'better placed' than COA 2 to succeed. This is the **Gamblers Fallacy** (Gambler's Fallacy is the tendency to think that future probabilities are altered by past events, when in reality, they are unchanged) and Sunk Cost Fallacy working in tandem. Human mind is not capable of grasping the fact that there may be situations where allotting additional resources for COA 2 can provide better success than COA 1 even when considerable resources have already been employed for COA 1.

Loss Aversion Bias *:* Closely related to the Sunk Cost Fallacy is the Loss Aversion Bias. This bias states that 'people prefer avoiding loss rather than maximising gain'. The perceived payoff of loss and gain is not linear as indicated in the figure below.

Figure 1: Loss Aversion Bias



Source: https://www.economicshelp.org/blog/glossary/loss-aversion/

- Consider the following two options after buying a lottery ticket of Rs 1 lakh:
  - **Option 1.** If you win, you get Rs 3 Lakhs. If you lose, you forfeit Rs 1 Lakh paid for lottery ticket.
  - **Option 2.** If you win, you get Rs 2 Lakhs. If you lose, you get Rs 50,000 back.
- People are more likely to choose Option 2 than Option 1. The extra loss of Rs 50,000 in Option 1 has more impact on decision making despite extra Rs 1 Lakh being gained in the same option.
- Military implication of this effect is that, given a finite set of resources and *ceteris paribus*, a military commander will be more inclined to use it for loss aversion rather than maximising gain.

Pseudo Certainty Effect *:* This is the 'tendency to disregard uncertainty in decision making, especially multi-stage decision making'. It results in people making risk-averse choices if the expected outcome is positive, but making risk-seeking choices to avoid negative outcomes. This may seem counter intuitive, but has been observed in various real world examples. In stock markets, when people are in profit, they book it early, even when they know that real value of the stock has not been realized (risk averse decision), however when stock suddenly dips 20-30%, people keep holding it or try to average it out despite understanding that the stock has fallen due to change in fundamentals (risk seeking decision).

In military terms, it implies that enemy commander may undertake low risk missions when he is considerably confident of victory, however, may be willing to take high risk missions when he is not confident of success or is confident of failure. This bias can be exploited to have an exponential detrimental effect on enemy's combat power. Psychological operations can be executed to make the enemy believe that he has lesser chances of success, thereby urging him to take higher risk missions. Higher risk missions on average should lead to weaker enemy combat power, which in turn will force him to take further risks. This cycle if properly exploited can subdue the enemy effectively.

Implications of Logical Fallacies on AI Models

Correct understanding of such biases can facilitate in correct assignment of weightages in an AIDSS for Wargaming enemy COAs.

These biases may inadvertently creep in the programming of AIDSS from their programmers as they affect the logic built into programming Programme code needs to be checked for biases during training of AI models.

AI models assisting in psychological and cognitive operations can exploit these biases to output effective psychological / cognitive warfare plans.

Zero-sum Bias

Zero-sum bias implies considering a situation incorrectly to be like a 'zero-sum game' (i.e., one person gains at the equal expense of another). While conflicts are usually considered zero sum, retaining this bias in decision making will result in foreclosing options which result in a win-win situation. Even in conflict, it is feasible to negotiate and work out a solution, to avoid uncontrolled escalation. This aspect is more relevant in grey zone or operations short of war.

Implications for AI Models

AIDSS will have to be trained to look at alternatives where the payoffs of the opposing parties are not considered inversely proportional to each other.

Tree search algorithms employed for working out options thus, also need to consider alternatives where anticipated probabilities may be less but outcomes may be better than 'Nash Equilibrium'. A classic example is the negotiation between Major General JFR Jacob and Lt Gen AAK Niazi during 1971 India- Pakistan War (Vivekananda International Foundation). The chances of Pakistan accepting surrender were considered remote initially. An AIDSS based on zero sum game would have rejected the possibility right away. However, it turned out to be the game changer in deciding the outcome of 1971 war.

Belief Bias

Belief Bias is an effect where evaluation of the logic in an argument is based on its 'believability'. This bias plays an important role in surprise and deception in military operations. German advance in World War II, is an apt example wherein allied forces did not expect advance of mechanised columns through forested area — it was not believable and hence not expected. However, during the Battle of France in 1940, German Panzer divisions advanced through the Ardennes Forest. The French forces had left the Ardennes Sector less defended, which resulted in their eventual defeat.

Implications for AI Models

Implication for AIDSS is that there would be certain options that might be considered as low probability by the AI engine, however, feasibility of the same should not be discounted. A potent representative example is the fourth game of 'Go' between Le Sedol and AphaGo AI engine wherein the 78th move made by Le Sedol was unexpected for AlphaGo as it had a very low probability score in AlphaGo's calculations. This move resulted in a deteriorating situation for AlphaGo with Le Sedol finally winning the match (Google Deepmind).

Thus, it must be realised that traditional approach of 'minimising the loss function', based on training data, would not work for a AIDSS involved in military operational decision making. There is a requirement of considering probability distributions of each COA and pitching it against all enemy COAs to generate an output probability distribution giving out likelihood of success.

Conclusion

Human Biases and Errors have been largely impacting military judgement. When an AIDSS is employed for military decision making, it not only should avoid biases and errors found in humans but should also compensate for them through diligent programming. Only then will an AIDSS be a force multiplier, capable of providing insights and advice to military commanders which would be qualitatively better than human advisors.

While there are large number of human biases, only the important ones have been discussed to derive key implications for AI models being employed in military decision making. Consideration of additional human biases and programming approaches to counter each bias can be a topic for further research.

Works Cited

D-Day Deception: Operation Fortitude South. English Heritage. https://www.english-heritage.org.uk/visit/places/dover-castle/history-and-stories/d-day-deception/.

Google DeepMind Challenge Match (March 8-15, 2016). https://deepmind-media.storage.googleapis.com/alphago/pdf-files/english/ls-vs-ag4/LS%20vs%20AG%20-%20G4%20-%20English.pdf.

Parshall, J. and Tully, A. (2007). *Shattered Sword: The Untold Story of the Battle of Midway*. Potomac Books Inc.

India's Greatest Military Victory : The Role played by Lt Gen JFR Jacob in the Liberation of Bangladesh (Jan 14, 2016). Vivekananda International Foundation https://www.vifindia.org/article/2016/january/14/indias-greatest-military-victory-the-role-played-by-lt-gen-jfr-jacob-in-the-liberation-of-bangladesh

The dangers of apophenia: not everything happens for a reason. Ness Labs. https://nesslabs.com/apophenia

Hyperbolic Discounting: Preference for Immediate Rewards Over Future Benefits (Aug 03, 2024). Renascence. https://www.renascence.io/journal/hyperbolic-discounting-preference-for-immediate-rewards-over-future-benefits#:~:text=Hyperbolic%20Discounting%20is%20the%20tendency,in%20psychology%20and%20behavioral%20economics.

About the Author

Col Nachiket Kodilkar was commissioned in Infantry in Jun 2002. He has a diverse service profile having served in Siachen Glacier, desserts, Counter Insurgency areas of J&K and Manipur and Andaman & Nicobar Islands. He completed his dissertation based on Artificial Intelligence during Higher Defence Management Course in 2022 and is currently pursuing his PhD from SRM Institute of Science and Technology, Chennai in the same field.