Article

Not peer-reviewed version

Probability Paradoxes for Increasing Critical Thinking Skills

Priyantha Wijayatunga *

Posted Date: 14 October 2024

doi: 10.20944/preprints202410.0935.v1

Keywords: Probability paradoxes and puzzles; critical thinking; solving; exercises

Article

Probability Paradoxes for Increasing Critical Thinking Skills

Priyantha Wijayatunga

Department of Statistics, Umeå University, Umeå, Sweden; priyantha.wijayatunga@umu.se

Abstract: Critical thinking skills are important assets for scientific development. Therefore, there are many ways to enhance the critical thinking skills of students, researchers, etc., both in the field of education and scientific enterprise. In this article, we propose the practice of solving probability paradoxes and puzzles for the purpose since these problems are often complex, subtle and tricky; therefore, their solutions require many aspects of critical thinking. In the following, the Introduction outlines some of the current research on the topic without going into more details. We provide background descriptions of the paradoxes and probability paradoxes in Section 2. Section 3 provides a short description of critical thinking. We present several definitions since there is no clear consensus on one single definition, both in the literature and in the general literature. And in Section 4 current research on using puzzles for enhancing critical thinking is given. We then discuss several probability paradoxes, namely, the Monte Hall paradox, the two-envelope (exchange) paradox and the St. Petersberg paradox, in Sections 5, 6 and 7, respectively. In Section 8, a brief discussion of the statistical hypothesis testing is given emphasizing its controversial nature. We analyze these problems and concepts in such a way that critical thinking is evoked while keeping the technical details to a minimum.

Keywords: probability paradoxes and puzzles; critical thinking; solving; exercises

---

1. Introduction

"Students should be made to think, to doubt, to communicate, to question, to learn from their mistakes, and most importantly have fun in their learning."

– Professor Richard Feynman.

"Human mental and cognitive abilities are not biologically determined but instead created and shaped by the use of language and tools in the process of interacting with and constructing the cultural and social environment."

– Lev Vygotsky[1]

Critical thinking is often described as a metacognitive process, consisting of a number of subskills such as analysis, evaluation, inference, etc., that, when used appropriately, increases the chances of producing a logical conclusion to a given argument, or a solution to a given problem (Dwyer et al. 2014). In the famous Delphi Report (Facione, 1990), critical thinking is described as a purposeful, self-regulatory judgment that results in interpretation, analysis, evaluation and inference as well as an explanation of the evidential conceptual, methodological, criteriological or contextual considerations upon which that judgment was based. Furthermore, it is an essential tool of inquiry and a pervasive and self-rectifying human phenomenon. Therefore, it is one of the most important

---

1 https://en.m.wikipedia.org/wiki/Lev_Vygotsky?fbclid=IwAR0QnoGxF5XOW2j_fSAxTUiTaXbv3GGwD9H7bz1JHbLYNlVPyR-XAgkP0RI

 doi:10.20944/preprints202410.0935.v1

aspects needed for the progress of scientific enterprises. One of the central goals of science education is to teach students to think critically about data, information, theories and, applicable scientific models and methods (Holms, et al. 2015). That is, critical thinking is one of the most valuable skills that should be passed on by the school to its graduates throughout their education (Thompson, 2011).

The famous statistician George Box once said, "*all models are wrong, but some are useful*", meaning that the modeler must be aware that a model for a given phenomenon is ultimately for some kind of utilization. Therefore, it should function well in the context concerned by answering practical questions realistically, meaningfully, objectively, effectively and efficiently. Often, critical thinking makes our solutions to a given problem logical, efficient, feasible and optimal. Additionally, it can make the solutions convincing and communicable. These aspects are essential for the progress of science. Any model is an approximation of the reality. Therefore, thinking critically about the phenomenon under investigation and about the models that are built for it is highly important for using the inferences from them.

There has been strong emphasis on developing the critical thinking abilities of students of sciences over the past several decades, e.g., see Halpern and Dunn (2023), Raj et al. (2022), Holms, et al. (2015) and so on. Holms et al. (2015) presented a simple learning framework for improving the quantitative critical thinking skills of science students. It uses cycles of decisions and quantitative comparisons among different datasets or models. The framework was shown to be appropriate for assisting certain types of learning tasks in science. That is, the authors have experimentally shown that the framework significantly improves students' critical thinking behaviors and therefore improves their ability to learn scientific tasks. Their main argument is that the key element of developing and enhancing quantitative critical thinking behaviors is repetition of practice of making decisions, with relevant feedback. This is complying with the abovementioned second quote in the beginning—one's abilities can be shaped by using appropriate tools.

Raj et al. (2022) noted that critical thinking involves thinking clearly and reasonably about what to do or believe. The authors argued that when using critical thinking, students investigate, appraise, interpret, or synthesize information and use creative thoughts to construct arguments, solve problems, or reach conclusions. They have given many arguments for engaging in critical thinking practices for better learning activities. However, they have not given any empirical evidence on their claim. In Halpern and Dunn (2023), an introduction is given to critical thinking, which is grounded in psychological science, especially cognitive sciences. Their ideas are presented through humorous and engaging language and numerous practical and real-world examples. The authors focus on development and improvement of critical thinking skills that are characterized by clear, precise and purposeful thinking. Additionally, they provide some empirical evidence to support their claims. Quoting John Dewey, the pioneering American educator, the authors emphasize the act of "learning to think" as one of the primary purposes of education, which is also emphasized by Albert Einstein by saying "education is not the learning of facts, but the training of the mind to think."

This article contributes to this broad field in terms of how to improve the critical thinking abilities of the students of science, technology, etc. Our thesis is that we can use probability theory and its problems applicable in the real world, especially probability paradoxes and puzzles, for the task. Generally, solving mathematical puzzles are very good tools for stimulating critical thinking. Dispezio (1997) discussed an array of "best" puzzles for this purpose. They are mostly in real-world contexts; therefore, they are easy to comprehend, even if they are difficult to solve. Understanding the problem clearly is the main ingredient in the problem-solving process. It is generally accepted that solving and tackling probability puzzles, problems and related calculations can increase critical thinking skills. This is because they are often hard, complex, tricky and subtle. They are mostly real-world phenomena, requiring workable, efficient, practical, optimal and logical solutions. The interesting aspect of these exercises is that they can be tested through simulations, experiments, etc. However, arriving at a solution for any given probability puzzle or paradox requires not only analytical and logical reasoning but also clear, creative and critical thinking. Therefore, training in solving probability problems can be an efficient way to enhance students' critical thinking skills. Since applied probability problems and paradoxes deal with real-world phenomena, they can be used for students in various fields of sciences, not only mathematics students.

Here, we discuss a few probability paradoxes, such as the Monte Hall paradox, two-envelope (exchange) paradox, St. Petersberg paradox and even null hypothesis testing, in the context of statistics. We attempt to discuss how the practice of solving such problems can be used to increase the critical thinking abilities of students in science, technology, engineering and mathematics (STEM), etc. To the best of our knowledge, this topic has not been extensively discussed in the research literature in the fields of education and student learning. However, few attempts have been made recently. The reader is referred to the article collection in the journal Frontiers in Education[2] (**Wijayatunga et al. 2024**). In the following, some of the earlier research literature is discussed briefly, but a discussion of a comprehensive list of previous works is beyond the scope of this paper.

Before we proceed, we should look at what it is meant by a probability paradox and what it is meant by critical thinking. We can go through few definitions, both in general usage and in the research literature. These definitions clearly show that the practice of solving probability paradoxes and puzzle can be beneficial for enhancing and sharpening one's critical thinking skills. Note that most of the probability paradoxes are discussed well in the literature. Though we present some new approaches to solve them, we do not discuss their solutions found the research literature in great details. Our proposal here is to use the practice of solving paradoxes to enhance students' and researchers' critical thinking skills.

2. Probability Paradoxes

The probability theory is considered to be a hard part of the discipline of mathematics, at least as far as applications are concerned. It may not be as hard as other parts such as topology, etc., but people often go wrong in applying it correctly in the real world. However, the field of statistics may not be so, even though its discussions are built around the probability theory. Ths is because it deals mostly with data, observations and estimation of parameters of selected models. Many students do not prefer to study these important subjects because they are afraid of not understanding their concepts. Notably, the mathematician and the philosopher Bertrand Russel once said that "*the probability is the most important concept in modern science, especially as nobody has the slightest notion of it*." This saying highlights how important it is for scientists to grasp the concept in the probability theory and those in statistics for their scientific progress in their respective disciplines. We believe that familiarity with the probability theory and statistics should begin when one is a student of science. The exercises in the probability theory and its applications can improve not only one's scientific skills but also one's critical thinking abilities, as we argue in this paper. Probability paradoxes are a special class of probability problems that can be more effective at enhancing the critical thinking skills of scientists. This is mainly because the solutions to these puzzles can be obtained if one looks at the problems clearly, broadly, deeply and rationally, etc., and therefore critically.

Let us first consider what a paradox means in general. The online Cambridge Dictionary defines a paradox as "a situation or statement that seems impossible or is difficult to understand because it contains two opposite facts or characteristics." Similarly, the online Oxford English Dictionary describes it as "a statement or tenet contrary to received opinion or belief, especially, one that is difficult to believe." Therefore, the paradoxes are thought-provoking statements or situations that seem to defy general logic, mostly with some contradictions and inconsistencies. However, they contain some truth in them. Therefore, solving and clarifying paradoxes mostly involves exploring and rectifying them. To do so, one requires clear thinking, reflection, imagination, logical, analytical and rational reasoning, etc. Furthermore, it is stated in Sainsbury (2009) on its back-cover by Pascal Engel that "a paradox can be defined as an unacceptable conclusion derived by apparently acceptable reasoning and apparently acceptable premises." Therefore, the paradoxes are understood and resolved through a high level of intelligence—not only through logical and analytical reasoning but

---

[2] https://www.frontiersin.org/research-topics/39405/probability-and-its-paradoxes-for-critical-thinking

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 October 2024 doi:10.20944/preprints202410.0935.v1

also through critical thinking. To complement the above general definitions of paradoxes, let us look at what the Encyclopedia Britannia says; paradox is an apparently self-contradictory statement, the underlying meaning of which is revealed only by careful scrutiny. The purpose of a paradox is to arrest attention and provoke fresh thought. The statement "Less is more" is an example. Therefore, they are ideal for stimulating and enhancing critical thinking skills. More often, the paradoxes are easily solved through correct interpretation of the statements in it, removal of ambiguities in them, or falsification of logical premises or implications. In Section 7 of the paper, Bandyopadhyay et al. (2015) it is discussed what a paradox is and some of the views of the philosopher on it.

Probability paradoxes are special cases of paradoxes. They are in situations where randomness and stochasticity are involved. A popular way to quantify randomness and stochasticity for describing events, outcomes, etc., is to use probability theory. Inherently, probability theory and its calculations are somewhat difficult to grasp, at least in the first instance. Therefore, it can be mind-twisting to think about or to try to understand them. Often, it requires, among others, critical thinking, logical reasoning, numerical computation and nonnumerical analysis, etc. For the analysis of probability paradoxes, one can never depart from applying clear logical reasoning, precision and deep thinking. Additionally, to understand them, one needs to be very critical about potential solutions or steps toward them. Therefore, learning and analyzing probability paradoxes can improve critical thinking skills, so they can be used for educational activities in science, engineering, etc., to develop the skills of students in their respective fields. Importantly, some questions cannot be answered since their solutions may not exist. They are not paradoxical situations. For example, if I ask what color you get if you mix all the colors. To get it, logically you need to mix all the possible colors, including the resulting color itself, which should give another color. Therefore, simply the color should not exist! To do this in a feasible way, one should be interested in the color that we obtain if we mix all the basic colors or similar colors.

3. Critical Thinking

There are many definitions of the concept of critical thinking in the literature. As stated in Schmaltz et al. (2017), for unfortunate reasons, the definition of critical thinking has become so wide and broad that it can encompass nearly anything and everything (Hatcher, 2000, Johnson and Hamby, 2015). This situation is not helpful, and it places an extra burden, both on teachers who teach their students how to think critically and on students who seek to practice it during their learning process. The authors propose that educators need to clearly define critical thinking in more precise terms and that, in addition to teaching critical thinking skills, a strong focus should be placed on teaching students how to think like scientists and perhaps technologists. They argue that scientific thinking is the task of generating, testing, and evaluating claims, empirical data, and theories. Thinking critically and scientifically is necessary for the development of scientific enterprises. We do not want to adopt one single definition of critical thinking. We think that it is appropriate to consider several definitions of it; after all, there has not been a general consensus. Note that almost all the definitions have some kind of overlap.

Now, let us look at some of the definitions of critical thinking in the research literature on education and general usage. One of the definitions of critical thinking is *thinking clearly by distinguishing the attributes/aspects of the phenomenon under investigation and using laws of reasoning and possibilities/probabilities* (Raj et al. 2022). Therefore, critical thinking occurs when the thinker is well adapted to the situation and is highly concerned with real-world aspects such as possibilities. This approach results in novel, feasible, efficient and effective solutions to the problems that are key for scientific progress through overcoming difficulties, pitfalls, and errors in previous or ad hoc solutions. That is, progress in science can be regarded mainly as a result of good critical thinking practices.

Educational expert, Prof. Dewey, defines critical thinking as inclusive of reflective thinking and argued that the thinking process should also be taken as one of the objectives of education (Dewey, 1910). This finding strongly emphasizes the need to develop critical thinking skills in students. The practice of reflection is important in any learning task, as it encourages the learner to explore many other possibilities for the problem and to remember the facts. The reflection can be performed even after the learning task. Either way, the students can enhance their knowledge to a greater extent by doing so. Another definition of critical thinking is that it *is reasonable reflective thinking focused on deciding what to believe or do*, as defined in Ennis (1987) and (1989). In addition to Dewey's emphasis on reflective thinking, the key point here concerns what to believe or do.

Our third definition of the concept of critical thinking is from the Nation Council for Excellence in Critical Thinking. They adopt the definition given by Michael Scriven and Richard Paul at the 8th Annual International Conference on Critical Thinking and Education Reform in the summer of 1987: "*Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action. In its exemplary form, it is based on universal intellectual values that transcend subject matter divisions: clarity, accuracy, precision, consistency, relevance, sound evidence, good reasons, depth, breadth, and fairness.*"[3]

Finally, and most recently, in Halpern and Dunn (2023), the following definition is given. "Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome (of understanding well through obtaining broad and deep knowledge). It is used to describe thinking that is purposeful, reasoned, and goal directed—the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions—when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task. Critical thinking is more than merely thinking about one's own thinking or making judgments and solving problems—it is effortful and consciously controlled. Critical thinking uses evidence and reasons and strives to overcome biases."

In addition to the definitions in the research literature, it is worthwhile to understand what it means by the concept of critical thinking in layman's general language. These definitions can be less technical yet intuitive; therefore, they are important, especially for early students of science, engineering, mathematics, etc. The online Oxford English Dictionary[4] describes critical thinking as "*the objective, systematic, and rational analysis and evaluation of factual evidence in order to form a judgment on a subject, issue, etc*." The online Cambridge Dictionary[5] describes it as "*the process of thinking carefully about a subject or idea, without allowing feelings or opinions to affect you*." Therefore, from these definitions, it is clear how important the critical thinking skills for the students for their learning process and advancement in their discipline of studies are. Furthermore, comprehensive details are given in the Encyclopedia Britannia[6]–critical thinking is "*a mode of cognition using deliberative reasoning and impartial scrutiny of information to arrive at a possible solution to a problem*."

From the perspective of educators, critical thinking encompasses a set of logical skills that can be taught. It is a disposition toward reflective open inquiry that can be cultivated. Therefore, logical analysis is a main component of the critical thinking process. Therefore, it should be taught to the students either as a special subject on its own or along with the student's major subject through augmentation. Either way, ideally, the teaching of critical thinking skills should start with involving the students in problem-solving exercises, especially subtle and tricky problems, which we propose in this article. Abrami et al. (2015) showed through a meta-analysis that the exposure of students to authentic or situated problems and working examples seems to play an important role in promoting and enhancing critical thinking skills, particularly when problem-solving and role-playing methods are applied.

Clearly, these definitions show how critical thinking is evoked when solving probability puzzles, as they simply address the evaluation of information and evidence, inference methods, rational analysis and reasoning, and so on. Furthermore, when solving probability puzzles, we consider every possibility with logical and analytical reasoning, especially when the puzzle concerns

---

[3] https://www.criticalthinking.org/pages/defining-critical-thinking/766

[4] https://www.oed.com

[5] https://dictionary.cambridge.org

[6] https://www.britannica.com

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 October 2024 doi:10.20944/preprints202410.0935.v1

any real-world scenarios. We need to define events precisely, evaluate their possibilities accurately and perform analyses, comparisons and optimizations—all of which are critical. Therefore, when we solve such puzzles, we train our minds to think about the aspects that are related to critical thinking skills deeply, thus improving them. One can successfully solve probability puzzles if and only if logically consistent and analytically correct possibilities are taken into account, especially in timely manner. That is, we need to eliminate impossibilities, either through logical processes or through numerical and analytical processes. We need to determine which options should be accepted and applied at each stage of the solution process. Therefore, it is rather obvious that critical thinking skills can be improved immensely through the practice of solving these puzzles.

Compared with their findings, specifically, we believe that teaching students how to solve and analyze probability paradoxes and puzzles is an ideal way to enhance their critical thinking skills. This is because, often, the probability paradoxes and their calculations are in real-world scenarios, and they do not require much technical knowledge of any scientific discipline for understanding and solving them, thus making them a common ground for the students of the sciences. In particular, we are interested in the probability paradoxes that can be understood by the students of the sciences, technology, engineering and mathematics. However, our arguments can also be related to other disciplines, such as economics, especially when they are connected to the disciplines of interest. In addition to the paradoxes discussed below, Brittan and Taper (2024) provide an analysis of well-known paradoxes, such as the lottery, the old evidence, and Humphreys' paradoxes, in terms of enhancing critical thinking. Those are easy-to-understand puzzles that require no deep technical knowledge in probability theory.

4. Increasing Critical Thinking Skills with Practice of Solving Probability Paradoxes

Recently, in Brittan and Taper (2024), the authors noted the role of critical thinking in statistical inference—how data provide evidence for making conclusions. They tried to show it through the analysis of three well-known statistical paradoxes—the Lottery, the Old Evidence and Humphreys.' Additionally, they noted that there are many definitions of critical thinking; therefore, they avoided using any particular definition in their discussion, but they wanted to be critical about critical thinking, which they called meta-critical thinking. In Aizikovitsh-Udi and Amit (2011), the authors explored if the inclusion of a specially designed learning unit called "Probability in Daily Life" would enhance students' critical and creative thinking. The unit was taught to a group of tenth-grade pupils in a mathematics course with the purpose of encouraging critical thinking dispositions such as open-mindedness, truth-seeking, self-confidence and maturity. Pupils are required to analyze problems, raise questions and think critically about the data and the information. The authors have shown empirically, using a small group of pupils, that the unit enhanced the pupils' critical thinking. This work extends the previous work of Aizikovitsh-Udi and Amit (2009). The authors showed that including the special learning unit "statistics in everyday life situations" in a mathematics course improved pupils' critical thinking skills, such as raising questions, seeking alternatives and doubting.

In Haq and Sawitri (2021), a qualitative study is described to determine if the critical thinking skills of students could be enhanced by solving problems in probability calculations. Empirically, the authors could show that the practice of solving probability problems made the students try to know information well, providing reasons for problem-solving strategies and looking for alternative solutions. Furthermore, in Salido and Dasari (2019), the authors tried to differentiate students' learning styles, which are determined by their thinking skills and processing of information, through solving probability problems. The findings suggested that the types of students' errors in solving probability problems were due to a lack of skills in comprehension, encoding, transformation and processing of information. They can all be attributed to insufficient critical thinking used to solve problems. In Rauzah (2019), the author considered how solving probability problems enhances the critical thinking abilities of students when problem-based learning is used in teaching. Using a quasi-experimental research method of a nonequivalent pretest-posttest control group design implemented with a small group of 28 students, the author concluded that the application of the problem-based learning method for solving probability problems had a positive effect on improving students' critical thinking skills.

In Mora-Ramirez (2023), the author considers paradoxes as didactic resources that can help developing students' critical thinking skills during their education. Therein, a list of paradoxes is analyzed with the objective that they can be used in the classroom. Some of them are, the paradoxes of Achilles and Tortoise, Galileo's paradox, Hilbert's hotel paradox, Tristam Shandy's paradox, Protagoras' paradox, etc. Although it is not about using paradoxes, Tang et al. (2020) reported that taking part in playful design games in small groups improves participants' creativity and critical thinking while contributing to communication and collaborative teamwork skills. The participants reported that they felt motivated by taking part in the games; therefore, their creativity, critical thinking, communication, and collaborative engagement improved. In Matthee and Turpin (2019), a course focusing on the explicit development of critical thinking and problem-solving skills among first-year information science students at the University of Pretoria was reported. The critical thinking part of the course focuses on analysis, evaluation, and response to arguments. Class discussions and assessments are based on local, authentic arguments. In the problem-solving skills component of the course, students are required to understand the nature of a problem and to classify it to one of three categories: puzzles, problems, and messes. The study empirically shows that solving puzzles and complicated problems improves the critical thinking abilities of students.

Additionally, in an online discussion, the topic of probability puzzles for critical thinking is discussed[7]. Another similar discussion is "*What are some of the best/toughest probability puzzles*"[8]. There are numerous online discussions and blog posts related to the topic. It is tedious to discuss all of them here, and it is beyond the scope of this paper. Most of them do not discuss the topic directly, but there are many indirect relationships with the topic. In a blog post titled "*Bayes Theorem: A Framework for Critical Thinking*"[9], an introduction to the Bayes theorem and how to put it into practice are given. The author argues that the Bayes theorem is a framework for critical thinking. And in another forum called "*Physics Forum*" under the title "*Tough Probability Problems*[10]", it was argued that solving probability problems needs critical thinking skills in addition to some understanding of probability theory. Therefore, they argued that the practice of solving probability problems can enhance the critical thinking skills of students. Note that the concept of probability is commonly used in many sciences, and therefore it is ideal that students in science possess some technical knowledge in probability theory. In addition to the above works and perhaps some others, formal discussion on the use of probability puzzles and paradoxes to enhance the critical thinking skills of students is not very old. We believe that this article is triggering further discussion on this topic. Note that this article is not about why critical thinking skills are important for students.

The paradoxes considered in this paper are the so-called Monte Hall paradox, the two-envelope (exchange) paradox and the St. Petersberg paradox. They are easy to understand, yet their solutions are difficult to obtain; therefore, they are good example exercises for enhancing the critical thinking skills of students. Note that some parts of our discussions on paradoxes are somewhat closer to the work of Kelter et al. 2024. Furthermore, we discuss statistical null hypothesis testing and the difference between practical significance and statistical significance. Note that null hypothesis tests are developed for practical usages; however, they sometimes fail to yield practically valid conclusions. This is rather a paradoxical situation as far as their purpose is concerned. In particular, the science students find it difficult to use them in their academic activities (see Krishan and Idris 2015, Zaini et al. 2021 and Castro Sotos 2009 for details), especially when the result of such a statistical hypothesis says that there is a difference between two given quantities, but it is not practical. Therefore, it is worthwhile to have some insights into this topic.

---

[7] https://www.quora.com/Does-learning-probability-and-statistics-improve-critical-thinking

[8] https://www.quora.com/What-are-some-of-the-best-toughest-probability-puzzles

[9] https://neilkakkar.com/Bayes-Theorem-Framework-for-Critical-Thinking.html

[10] https://www.physicsforums.com/threads/tough-probability-puzzle.96903/#post-808793

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 October 2024 doi:10.20944/preprints202410.0935.v1

5. The Monte Hall Paradox

The so-called Monty Hall problem or puzzle (Freedman 1998, Wijayatunga 2022) is said to fool even mathematicians. This paradox can be demystified through correct interpretations of events, terms, etc., or avoidance of equivocation. Suppose you are on a game show. And you are given choices of three doors to select one, where behind one of the doors there is a car and behind the other two doors there are two goats, one for each. The goal is to select the door that has the car behind it. Therefore, you select one of the three doors (say, Door-1). The game host then reveals one non-selected door (e.g., Door-3) that does not have the car behind it. At this point, you can choose whether to stick with your original choice (i.e., Door-1) or switch to the remaining unopened door (i.e., Door-2). What are the probabilities that you will win the car if you stick, versus if you switch?

Most people believe, upon first hearing of this problem, that the car is equally likely to be behind either of the two unopened doors, so the probability of winning is $1/2$ regardless of whether you stick or switch. However, the probabilities of winning are $1/3$ if you stick and $2/3$ if you switch. Therefore, the confusion is between $1/2$ and $2/3$. This is the critical aspect of the problem. Therefore, your thinking should be directed toward why you come across two probability values instead of just one.

Now, let us see how these two numbers come. Let C be the selection of the door that has the car, G_1 and G_2 be the selections of the doors that have the goat-1 and the goat-2, respectively, and S denotes the event of switching (S' denotes sticking). Let the ordered sequence CG_1 denote the event of selecting the door with the car first and then the door with the goat-1 in the event of switching and so on. When switching is performed, the conditional probability of selecting the door with the goat-1 first and then selecting the door with the car is

$$P(G_1C|S) = P(G_1|S)P(C|G_1,S) = P(G_1)P(C|G_1,S) = (1/3)(1) = 1/3$$

Similarly, we have the other probability:

$$P(G_2C|S) = P(G_2|S)P(C|G_2,S) = P(G_2)P(C|G_2,S) = (1/3)(1) = 1/3$$

Therefore, the probability of winning the car, if it switches, is

$$P(C|S) = P(G_1C|S) + P(G_2C|S) = 2/3$$

And simply, if you stick, selecting the car is $P(C|S') = 1/3$.

Now, if switching is performed randomly, which is implicitly assumed here, i.e., $P(S) = P(S')$=1/2,, then the probability of winning the car on average is

$$P(C) = P(S)P(C|S) + P(S')P(C|S') = (1/2)(2/3) + (1/2)(1/3) = 1/2,$$

i.e., the probability of finding the car is $1/2$, on average when switching is done randomly. However, as we have seen, the probability is $2/3$ if you switch and $1/3$ if you do not switch. Note that the two probabilities $1/2$ and $2/3$ refer to two different events. The confusion arises when these two well-defined probabilities are interpreted as if they mean the same. The value $1/2$ corresponds to an unconditional event of winning the car (i.e., C) on average (switching or sticking), and the value $2/3$ corresponds to a conditional event of winning the car if it switches (i.e., $C|S$). Similarly, $1/3$ corresponds to a conditional event of winning the car if it sticks (i.e., $C|S'$). Therefore, the monthly Hall problem involves identifying the two probabilities of $1/2$ and $2/3$.

Saenen et al. (2018) presented a systematic review of the literature on the problem published between January 2000 and February 2018, addressing why humans systematically fail to react optimally to the problem or fail to understand it. Recently, Pearl and Mackenzie (2018) presented an attractive and intuitive solution to the puzzle. However, it is based on causal reasoning rather than probabilistic reasoning, which we are interested in here. Pearl and Mackenzie say that "*The Monte Hall paradox, an enduring and for many people infuriating puzzle, highlights how our brains can be fooled by probabilistic reasoning when causal reasoning should apply.*" We want to avoid causal reasoning here, as our interest is in understanding probability and resolving probability puzzles. We believe and are convinced that probabilistic reasoning can solve the puzzle easily. Although these authors claim that

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 October 2024 doi:10.20944/preprints202410.0935.v1

probabilistic reasoning can fool the human mind, especially in a causal context, we believe that it cannot occur when critical thinking is applied and when events are sequenced correctly. In fact, the latter is related directly to causal reasoning. Note that even in causal contexts, events can be described with probabilities with appropriate constraints. Therefore, in our solution, we use only marginal and conditional probabilities to describe events occurring over time and unity of probability for the totality of the events concerned. Although long expressions of probabilities seem to be involved, only very elementary probability operations are used.

The main concern that we reconstitute is that the people do not put enough critical thinking into practice in solving and explaining the puzzles such as this. There are many explanations for the paradox, but to the best of our knowledge, none of the articles claim that the problem the human mind has is differentiating between the conditional event and the marginal (unconditional) event, which is shown above. For this purpose, clear thinking, albeit not much mathematical ability, is needed in critical ways. Note that the first step of the mathematical way to solve and explain the problem that the human mind faces is to find interpretations of two values, $1/2$ and $2/3$ (alternatively, $1/2$ and $1/3$). Additionally, it is important to note that the value $1/2$ is obtained as the average probability of winning the car because switching is performed randomly, and furthermore, the two distinct probabilities of winning the car if it switches ($2/3$) and if it sticks ($1/3$) sum to 1. If they do not, then we may not obtain $1/2$ as the answer for this case. When you think clearly and critically about these numbers, then you understand the problem underlying the paradox. Therefore, the students of science, engineering, etc., can try this paradox to train their critical thinking abilities, thus enhancing them to a greater degree. Most importantly, we need very little knowledge about probability theory to solve this problem; however, some critical thinking is needed.

Recall that at the final stage of the game, the player grasps only that there are two doors shut and that the car must be behind one of them. Therefore, the player generally concludes that the winning probability is $1/2$. As shown above, this is the unconditional (marginal) probability as long as the switching is performed randomly. If the player tends to switch more often than not, his/her winning probability exceeds $1/2$ and becomes closer to $2/3$, even if he/she believes that it is $1/2$. In the extreme case of always switching $P(C) = P(C|S) = 2/3$, even if the player believes that switching makes no difference in the chances of winning the car!

6. Two-Envelope (Exchange) Paradox

Two envelopes enclosing money are presented to you and to your friend, where one envelope contains twice the amount of money of the other. You two are not informed which envelope contains twice the amount of money of the other. Each of you are then asked to select one envelope. Upon selection, you open your envelope and find that it contains x Dollars. Let random variable Y denote the number of dollars in the other envelope (since you do not see it). Let us drop the monetary unit for simplicity. Since your envelope is selected randomly, it can be that either $Y = x/2$ or $Y = 2x$, with equal probability of $1/2$. Thus, your expected winning amount by trading with your friend is $\frac{1}{2}\left(\frac{x}{2} + 2x\right) = \frac{5x}{4}$, which is larger than x, the amount that you currently have. Therefore, you would like to trade. Since the same reasoning can be used by a friend, trading can be performed immediately. In fact, you do not need to see the amount in your envelope for this reasoning as long as "any amount" is equally likely to be enclosed in your envelope.

According to the above reasoning, trading is reasonable; however, it is unreasonable to trade since, e.g., you may not stop trading if you both do not open the envelopes after a trading. That is, as soon as you get an envelope, you tend to trade if both of you do not know each of your amounts. Therefore, it is a paradox. Many studies have attempted to resolve this paradox in the literature, such as Albers et al. (2005), Christensen and Utts (1992), Nickerson and Falk (2006) and Wijayatunga (2019). In the second reference above, the authors identify the amount that is put in the first envelope, e.g., M, as the parameter of interest when the problem is solved statistically. In addition, their solution was based on the use of a "noninformative" prior distribution, and the conclusions are intuitive. However, they do not provide reasons why the above reasoning is wrong. The paradox is discussed in Kelter et al. (2023) in relation to science, technology, engineering and mathematics

(STEM) education. They try to show that if conditional expectations are calculated as done in the naïve solution to the problem, then a uniform distribution over an infinite set of points must be defined. They argue that such a model cannot exist mathematically. Unfortunately, their argument is not easy for STEM students to grasp, as they have included undefined symbols, namely, n.

If you want to simulate this game to see if we get conditional expectation of $E\{Y|X=x\}=5x/4$ at least approximately for any given amount x then you need to include countably infinite number of pairs of amounts that satisfy the condition that one amount is twice the other, e.g., if you want to see a certain pair of amounts, say, $(M,2M)$ in the two envelops at a certain instance then your simulation should include all the pairs of amounts $\ldots\ldots,(4M,8M),\ (2M,4M),\ (M,2M),\ (M/2,M),\ (M/4,M/2)\ldots\ldots$ which is a countably infinite sequence. This gives you, through sampling infinitely or at least a large number of times, that $E\{Y|X=x\}\approx 5x/4$ for any $x=\cdots,4M,2M,M,M/2,M/4,\ldots$. Clearly, this requires a definition of a uniform discrete probability distribution over an infinite set of points. In Wijayatunga (2019) it was talked about this earlier and in Kelter et al. (2023) it was argued such a probability model does not exist. It is clear that if you want to include other pairs of amounts too then you need to augment the above set of infinite number of pairs with similar and corresponding infinite set of amount pairs! If you think critically, since we cannot have a set of countably infinite number of amount-pairs practically we should avoid any conditional expectation calculations on such sample spaces, thus avoiding the paradox! This is because, the assessment of the conditional expectation is dependent on having a set of infinite number of amount pairs, that is an impossibility. It seems that this might be the argument presented in Kelter et al. (2023).

However, when the game host decides a single pair of amounts, say $(M,2M)$ then it can be played by two players! Then both players may only calculate their marginal expectations $E\{X\}=3M/2=E\{Y\}$ and therefore they avoid exchanging envelopes. Note that if one simulates the game with a single pair of $(M,2M)$ that is used in a given game, then it is impossible to get $E\{Y|X\}\approx 5X/4$ practically. Then we get that $E\{Y|X=M\}=2M$ and $E\{Y|X=2M\}=M$ but nothing else. Furthermore, since envelopes are selected randomly, $E\{Y\}=\frac{1}{2}2M+\frac{1}{2}M=\frac{3M}{2}=E\{X\}$. However, if we use the formula $E\{Y|X\}=5X/4$, then we do not obtain the above result, i.e., $E\{Y\}=E\{E\{Y|X\}\}=\frac{1}{2}\frac{5(M)}{4}+\frac{1}{2}\frac{5(2M)}{4}=\frac{15M}{8}$, which is incorrect. Therefore, the formula $E\{Y|X\}=5X/4$ cannot be acceptable!

Here, we present another simple but critical solution, as our point here is only to use critical thinking. Let us say that the two amounts that are put inside the envelopes are 100 and 50 in that order, written as (100, 50). Then, the person with the 100-envelope may think that the two possibilities in the other envelope are 200 and 50, each with a probability of 0.5. The other person with the 50-envelope may believe that the two possibilities in the other envelope are 100 and 25, each with a probability of 0.5. Note that both persons use these possibilities to calculate their own two conditional expectations separately but simultaneously. Now, let us write these (above) assumed amounts of the two persons as ordered pairs. They are (100, 200), (100, 50), (25, 200) and (25, 50). It is easy to note that all these pairs satisfy the initial condition that one amount is twice the other but the third combination (25, 200), which is the instance where the first person assumes that the second is having 200 while the second person assumes that the first person has 25. That is, when both persons calculate their conditional expectations, they include a possibility that is prohibitive. This simply says that the use of conditional expectation calculations together is not allowed because they have included a prohibitive combination in their calculations. Therefore, both persons should avoid the conditional expectation calculation altogether, thus nullifying the paradoxical situation. One can argue that the conditional expectation does not exist as shown above! What they should do is select an envelope and expect that it contains an amount of *3M/2* on average when two amounts are *M* and *2 M*. This is the marginal expectation of both of you. There is no prohibitive possibility for use in this calculation. Since both have the same marginal expected value, no trading is desired.

Note that our solution mostly uses critical thinking in the sense of applying some basic statistical definitions, in this case, identifying events and conditional probabilities desirably, correctly, clearly and convincingly, and providing clarity to the solutions through simple logical arguments. Note that in the literature, there are many solutions that are complicated and more technical. Unfortunately, they lack clear arguments (not those that are technically logical) and aspects of critical thinking.

However, some technical knowledge is needed for puzzles such as this, as they need clear definitions, interpretations, etc.

7. The St. Petersberg Paradox

This paradox was introduced by Nicolas Bernoulli in 1713. Suppose you offer the following game for another person and then you want to find an entry fee for the game to be charged from the other since you are going to pay the winning money. You have a fair coin, and you are going to toss it. If it turns the Heads in the first round of tossing, you pay 2 dollars to the other, and the game is over. If it falls the Tails, then you toss the coin once again. In this second round, if it falls the Heads, then you pay the other 4 dollars, and then the game is over. Otherwise, you toss the coin yet once again. Then, in this third round of tossing, if it falls the Heads, you pay the other 8 dollars and finish the game. Otherwise, you go for another round of tossing. That is, the tossing continues until it falls the Heads. The winning money is 2^n dollars if the Heads falls in the n^{th} round of tossing where n is 1, 2, 3, etc.

With some knowledge of advanced theoretical probability theory, one can prove that the game should finish with probability one; therefore, it should end in a finite number of tosses. However, we cannot know when it finishes for any given occasion! Therefore, we need to consider all possible tosses. This leads us to consider an infinite number of tosses for the game (running it forever!)

Now, assume that you have to find the entry fee for the game so that no one loses money, on average. Therefore, the fee should be the expected pay-out of the game. Since the probability that the Heads falls in the first round is $\frac{1}{2}$, the probability that the Heads falls in the second round (the Tails and the Heads in the order) is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, and so on, the expected amount of money that should be paid (dollars) is infinite;

$$E\{S\} = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 4 \times \frac{1}{8} + \cdots = \sum_{n=1}^{\infty} 2^n \times \left(\frac{1}{2}\right)^n = \sum_{n=1}^{\infty} 1 = \infty$$

That is, you should charge an infinite amount of money from the other for him/her to play the game with you. If so, no sensible person will ready to play the game! In fact, for the game, the expected payoff does not exist mathematically.

Any critical thinker finds that there does not exist an entry fee if we consider that the tossing can continue forever because, then it cannot be a game at all, since we are not considering there a finish even if it exists (in an unknown finite time). There should be an entry fee only when the game has an end in all possibilities, so there is a winner/loser. If we consider that the game tends to go forever, we should define an end to it by our intervention to define it as a game.

There are solutions for this problem in the literature, but there are many arguments against or for it. Most of the solutions are rather technical and do not deal critical thinking, e.g., in Yukalov (2021), a rigorous mathematical resolution is given. Therefore, here we are not going to review those solutions.

Now, if you try to look at the problem critically, we should ideally think of the game in other ways but not forget that the expected pay-out is infinite or rather does not exist. Note that here, our emphasis is on how we can look at the problem critically. One way to think is to understand that both of you may not finish the game in principle! Therefore, playing is not meaningful, as is defining the entry fee. This argument says that you should bound or limit the number of tosses to a maximum.

Now, let the number of tosses that the game plays until the heads fall be N, where $N = 1$ with probability $\frac{1}{2}$, $N = 2$ with probability $\frac{1}{4}$, $N = 3$ with probability $\frac{1}{8}$, and so on. Therefore, the expected number of tosses that the game is played is the mean of a geometric distribution with a success probability of $\frac{1}{2}$, which is

$$E\{N\} = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + \cdots = \sum_{n=1}^{\infty} n \times \left(\frac{1}{2}\right)^n \to 2$$

That is, on average, you need to play two tosses to finish the game! However, this does not mean that the expected pay-out is only 4 dollars. It is beyond the scope of this paper to talk about how to use this statistical information to calculate the entry fee. However, it shows that game is expected to end,

 doi:10.20944/preprints202410.0935.v1

therefore one can critically think that the two players should agree to stop the game if it tends to go forever or at least for a large number of tosses. Recall that, mathematically, the number of tosses needed is finite with probability one; therefore, the game finishes with a finite number of tosses, but as long as the entry fee is concerned, the expected pay-off does not exist since we consider all possible tosses. The other person may agree to pay a fee that may be "somewhat" lower than the maximum possible winning money in the case of the game being finished.

Now, we can find the probability that the game is finished within, e.g., 8 tosses is $\sum_{n=1}^{8}\left(\frac{1}{2}\right)^n = 0.9960938$. That is, the probability that the game goes beyond 8 tosses is just 0.00390625 (rather small!). When the game needs more than 8 tosses, the game can be ended by paying the other person the amount of money as if the Heads falls on the 9th tosses, i.e., the winning money of $2^9 = 512$ dollars. Therefore, we can set an entry fee of

$$E\{S\} = \sum_{n=1}^{8} 2^n \times \left(\frac{1}{2}\right)^n + 2^9 \times \left(1 - \sum_{n=1}^{8}\left(\frac{1}{2}\right)^n\right) = 10$$

dollars. Note that the selection of the number of tosses 8 can be arbitrary or rather based on agreement between the two players. It could well be 6 or 10, etc., but not very small such as 1 or 2.

The critical aspect of solving the paradox is determining that it should be a practically well-defined game to have an entry fee for it. No one should be willing to pay an entry fee for a game that is not guaranteed to end at least in a reasonable amount of time. This identification (that it should be a well-defined game practically) is important for solving the paradox. One arrives at this identification through critical thinking about the game—appropriate questioning about it and identification of concepts of a game properly. Highly technical mathematical skills are of less use if one cannot come to this identification. Therefore, this paradox is good for students to test their critical thinking skills. However, some knowledge in probability theory is needed to obtain a solution to the problem, e.g., the game ends with a probability of one (with the application of a strong law of large numbers).

8. Statistical Null Hypothesis Testing

In science, engineering, etc., so-called statistical null hypothesis testing, aka statistical significance testing, is used extensively for their technical advancements and applications since most of the new results in these fields are obtained with the support of statistical inferences from empirical data, both observational and experimental. Therefore, the students and the researchers of these disciplines should know about statistical hypothesis testing sufficiently, in order to apply it correctly for their contexts. However, it has been argued and shown in research studies that these tests are among the most misused statistical tools. For example, when hypothesis tests are applied, researchers pay *almost* no attention to the uncertainty of the p-values of hypothesis tests, even though they clearly have some random variations (Yanagawa, 2020). Any critical thinker should know that the study should be repeated several times before arriving at a firm conclusion, especially due to this random variation. This is what Fisher, the pioneer of null hypothesis testing, advised (see below). As stated in Yu and Kumbier (2020), *"resource constraints can limit how data are collected, resulting in samples that do not reflect the population of interest, distorting the probabilistic interpretations of traditional statistical inference."* Therefore, there are uncertainties in the observed data. The question is that if we have taken them into account when performing hypothesis tests for desired statistical conclusions. This may be one of the main causes for the current replication problem/crisis in science (Wijayatunga 2022). The students of science should be aware of the nature and the uncertainty in p-values, and also that hypothesis tests are misused in the practice.

Critical thinking is needed to understand and to apply the theory of hypothesis testing correctly. Here, the most appropriate description of critical thinking is that it is about *thinking clearly by distinguishing the attributes/aspects of the phenomenon under investigation and using laws of reasoning and possibilities/probabilities* (Raj et al. 2022). Arguably, the practice of applying hypothesis tests can train students in using critical thinking skills, thus causing their development to a greater extent. However, to the best of our knowledge, no empirical studies have been conducted on this topic.

In Ioannidis (2005), it is argued that most of the published research findings that are based on statistical significance are false. It is stated that "*the high rate of non-replication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by the p-values, but, unfortunately, there is a widespread notion that medical research articles should be interpreted on the basis of only the p-values.*" Therefore, the main problem seems to be the over-reliance of statistical significance tests. The students find themselves in a paradoxical situation since they are taught that the statistical hypothesis testing is used for obtaining significant differences between quantities, effects of treatments, etc. However, it does not take much time for them to understand the reality if they investigate the theory. The catch is that the result is often based on one single study. In hypothesis testing, there is no real paradox involved, but such situations are rather due to failure to follow Fisher's advice of doing carefully conducted multiple studies when making statistical inferences (see below).

There is a large body of literature about correctly handling null hypothesis tests, handling their strengths and weaknesses through the use of effect sizes, confidence intervals, etc. and alternatively using Bayesian hypothesis testing. The reader is referred to works such as Kruschke (2011) and Lakens (2021) for more details. We are not going to review these works since it is not the objective of our study. However, they are very important for researchers and students who are working with statistical inferences from the real-world data. Note that our scope here is to discuss how critical thinking is used in statistical inference. Therefore, we want to highlight our way of analysis rather than all other valid analyses available in the literature. If this were a paper on statistical inference, then we would need to go through all these approaches and summarize them before presenting ours.

According to Fisher, one of the pioneers of statistical hypothesis testing, a single study is not enough to conclude existence of an effect, a difference, etc. in given phenomenon (*Hubbard and Bayarri, 2003*). We need to conduct carefully planned experiments several times under the same conditions. The original thesis of Fisher is that the *p*-value of a test should be considered a weak guide for assessing the strength of evidence for/against the null hypothesis. That is, the meaning of expression such as "*p-value is less than, say, 0.05*" is that the respective experiment should be repeated a few more times until the experimenter is convinced that the null hypothesis cannot be true most likely! In the case of observing such small *p*-values repeatedly, one can conclude that the observed effect, difference, etc., may not be solely due to chance. In this way, the user can minimize the occurrence of any exceptionally rare event in the data. Unfortunately, such a practice that is advised by Fisher is rarely performed today. Furthermore, it is important to keep in mind that the methods based on *p*-values was originally developed for experimental research by Fisher to analyze the effects of randomized assignment of treatments for small groups of subjects. That is, they were defined for causal analyses. However, the *p*-values are now being used in other contexts, such as statistical model building and predictions, and in machine learning, where the data are nonexperimental (just observational) and are often massive.

One of the simplest questions that can be asked about null hypothesis testing is why it does not always imply practical significance when the conclusion of the test is statistically significant. In short, the statistical significance tests fail to capture practical aspects of the application domain, yet they are developed for practical contexts. The students of sciences and some researchers seem not to understand this often. For them, this contradictory situation is somewhat paradoxical; therefore, they try to stick to the statistical conclusion rather than understanding what is going on with the test. This is one of the reasons for the replication crisis in science, which is a burning issue in the practice of science. Therefore, although it is not a completely paradoxical situation as far as probability theory is concerned, it is worthwhile to discuss it here.

Let us consider the following problem for simplicity. A total of 307 boys and 85 girls are selected randomly from a certain sports camp for small children. During free time, these children are offered access to candies. One of the organizers of the camp noted how many candies each child had taken. Then, he performs a two-sample *t test* to determine if there is a difference in the mean number of candies taken by boys and girls. Note that it is expected that a child may generally take approximately 15 candies. The test *results in a p*-value of 2.58x10^{-5}, which is much lower than the level of significance, e.g., 0.05. Therefore, the test revealed that there was a significant difference (a statistical significance) in candy consumption between boys and girls. However, he noted that boys consumed an average of 15.25 candies, whereas girls consumed an average of 16.61 candies. The empirical difference between the two means is just 1.36 candies. Practically, this is not a difference that should be considered a real difference, i.e., the difference may not have any practical significance unless for very special reasons. Note that the overall average is 15.54 candies, which is $(15.25 * 307 + 16.61 * 85)/(307 + 85)$. The difference is approximately 9% of the overall average.

However, if you have a rather small sample of boys and girls where the mean difference in their ability to eat candies is the same as above (1.36 candies), then the two-sample *t test* would not show any statistically significant difference because the samples contain little information about the boys' and girls' ways of eating candies; therefore, such a small difference in the means is attributed due to some randomness rather than due to a systematic difference. Two persons doing the test, one with a large sample and the other with a small sample, will have two different and opposing conclusions. The statistical laws dictate when to account for a difference or effect as significant. In this case, the main ingredient in deciding if the difference is considerable or meaningful is the central limit theorem, which states that as the sample size increases, the sample mean becomes increasingly normally distributed around the true mean of the population, with its variance becoming inversely proportional to the sample size. Thus, it allows for sample mean values that are only closer to the true mean as the sample size increases. In our case, the possibility of two sample means having overlapping ranges diminishes as sample sizes grow; therefore, a significant difference between population means exists.

Sometimes the students of science fail to understand that statistical conclusions are highly sensitive to the amount of data that is used for them. Statistical conclusions are often drawn on observed numerical values as if they are different from the desired hypothesized values or numerical differences, e.g., if there is a difference between males and females in terms of a given quantity of interest. A tiny difference can become considerable if the amount of data used to assess it is very large. This is because, even for such a tiny difference, the empirical support is greater for it being not equal to 0 numerically, so it is a real difference. This is true as far as statistical laws are concerned, that say that the random nature of the context is not sufficient to cause the variation of the respective estimate to exceed the observed difference between the estimate and its hypothesized value. This property is called the consistency of an estimate of a parameter of a probability distribution. In case of large sample, an instance where a small difference is inferred as a real difference is when the hypothesized value is not the true value of the parameter but very close to it.

9. Conclusion

We argue that the practice of solving probability puzzles, paradoxes and similar problems can be used to enhance the critical thinking skills of students and researchers in sciences, technology, mathematics and other fields. This is simply because they are often complex, subtle and tricky; therefore, to understand and solve them, one needs not only logical and analytical reasoning but also clear, sharp, deep and rational thinking, reflection, etc. which are some of the main ingredients of critical thinking. The idea of using probability puzzles is not new. People have been using it occasionally but often. Here, we want to provide formal treatment of this idea; thus, educators and scientists may apply the concept in various ways more often in the future. Of course, there is much work to be done in this topic, for which we have given a formal head start here. One of the main aims of this paper is to discuss this topic among the researchers and practitioners of education.

We have taken several paradoxes, namely, the Monte Hall paradox, the two-envelope (exchange) paradox, and the St. Petersberg paradox, and what may be thought of as a paradoxical situation in statistical null hypothesis testing to show how they can be used for the task. Note that, to solve these paradoxes, deep knowledge of probability theory and only some knowledge of the applications of its concepts in the real world are needed. It is important to use such simple and day-

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 October 2024 doi:10.20944/preprints202410.0935.v1

today applicable puzzles since, then, we can target the general students of science, technology, etc., who might not have deep knowledge in probability theory.

We have shown that the Monte Hall problem can be solved or rather explained through correct interpretation of probability values involved in the context. The two-envelope paradox is resolved by showing inconsistent elements in the sample space; therefore, it is prohibitive to define conditional expectations simultaneously. Furthermore, in order to have conditional expectation we need a discrete probability distribution over a countably infinite set of points. Regarding the so-called St. Petersberg paradox, we have shown that, to define an entry fee, the game should be defined so that it never tends to run forever. Finally, we discuss why we observe paradoxical conclusions in the statistical null hypothesis test. We have shown how statistical laws (e.g., the central limit theorem implied by the law of large numbers) define a difference between two numerical values according to the information in the context. However, such a decision does not consider the desired practical importance of the difference. All these conclusions use critical thinking skills more than the technical knowledge of probability theory does.

Other paradoxes that may be used include Simpson's paradox, Lord's paradox, and Sleeping Beauty's problem. We believe that the solutions to these paradoxes deal mostly with critical thinking, and they require less technical probability knowledge, thus making them attractive to the task. In fact, we plan to conduct further research on the aspects of critical thinking with such paradoxes in the future.

Finally, we would like to return to the two quotes, one by a professor of physics and the other by a psychologist, on student learning at the beginning of this paper. The first quote is directly connected to critical thinking through the practices of questioning, doubting, reflection, thinking and communicating (through presentations, demonstrations and so on). These actions, which are some of the aspects of critical thinking, help the student understand and be convinced well of the concepts that they have studied by being able to teach them to others, applying them in the real world and going beyond the acquired knowledge whenever needed. In addition, the second quote says that only practice of anything, in this case the critical thinking, make us better in it. Therefore, students need to practice critical thinking through various means if they want to improve their skills. Our proposal here is to practice it through solving probability problems, puzzles and paradoxes.

References

1. Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A. and Persson, T. (2015). Strategies for Teaching Students to Think Critically: A Meta-Analysis. *Review of Educational Research*, 85(2), 275–314. https://journals.sagepub.com/doi/10.3102/0034654314551063
2. Aizikovitsh-Udi, E and Amit, M. (2009). An innovative model for developing critical thinking skills through mathematical education. In L. Paditz & A. Rogerson (Eds.), *Proceedings of the International Conference of the Mathematics Education into the 21st Century Project: Models in developing mathematics education, Dresden, Germany: University of Applied Sciences* (pp.19–22).
3. Aizikovitsh-Udi, E and Amit, M. (2011). Developing the skills of critical and creative thinking by probability teaching. *Procedia Social and Behavioral Sciences (WCES-2011)*, 15, 1087–1091. https://doi.org/10.1016/j.sbspro.2011.03.243
4. Albers CJ, Kooi BP and Schaafsma, W (2005). Trying to resolve the two-envelope problem. *Synthese*, 145, 89–109. https://doi.org/10.1007/s11229-004-7665-5
5. Bandyopadhay, P. S., Raghavan, R. V., Dcruz, D. W. and Brittan, G. (2015). Truths about Simpson's Paradox: Saving the Paradox from Falsity. In: Banerjee, M., Krishna, S.N. (eds) *Logic and Its Applications. ICLA 2015. Lecture Notes in Computer Science*, 8923, 58–73. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45824-2_4
6. Brittan, G. and Taper M. L. (2024). Meta-critical thinking, paradox, and probabilities. *Frontiers in Education*, 8:1217790. https://doi.org/10.3389/feduc.2023.1217790
7. Christensen, R. and Utts, J. (1992). Bayesian Resolution of the "Exchange Paradox". *The American Statistician*, 46(4), 274–276. https://doi.org/10.1080/00031305.1992.10475902
8. Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2009). How confident are students in their misconceptions about hypothesis tests? *Journal of Statistics Education*,17(2). https://doi.org/10.1080/10691898.2009.11889514
9. Dewey, J. (1910). *How we think.* D. C. Heath & Co Publishers, New York. Retrieved July 7, 2024, from https://bef632.files.wordpress.com/2015/09/dewey-how-we-think.pdf

10. Dispezio, M. A. (1997). *Great critical thinking puzzles*. Sterling Publishing Company, Inc. USA
11. Dwyer, C. P., Hogan, M. J., Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, 12, 43-52, https://doi.org/10.1016/j.tsc.2013.12.004
12. Ennis, R.H. (1987). A Taxonomy of critical thinking. Dispositions and abilities. In J. B. Baron, and R. J. Sternberg (Eds.), *Teaching for Thinking.* New York: Freeman, (pp. 9–26).
13. Ennis, R.H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*,18(3), 4-10.
14. Facione, P. A. (1990). The Delphi Report: Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Millbrae, CA: California Academic Press.
15. Freedman, D. (1998) Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly. *The American Economic Review*, 88, 933–946. https://www.jstor.org/stable/117012
16. Halpern, Diane F. and Dunn, Dana S. (2023). *Thought and Knowledge: An Introduction to Critical Thinking*, the 6th Edition. Routledge, Taylor and Francis Group, New York and London.
17. Haq, I. M and Sawitri, F. W. (2021). Students' Critical Thinking Skills in Solving Probability. *Indo-MathEdu Intellectuals Journal* 2(2), 108–117. https://doi.org/10.54373/imeij.v2i2.23
18. Hatcher D. L. (2000). Arguments for another definition of critical thinking. *Inquiry* 20(1), 3–8. https://doi.org/10.5840/inquiryctnews20002016
19. Holmes, N. G., Wieman, Carl E. and Bonn, D. A. (2015). Teaching Critical Thninking. *Proceedings of National Academy of Science*, 112(36), 11199-11204. https://doi.org/10.1073/pnas.1505329112
20. Hubbard, R., and Bayarri, M. J. (2003). Confusion Over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing. *The American Statistician*, 57, 171–177. https://doi.org/10.1198/0003130031856
21. Ioannidis J. P. A. (2005). Why most published research findings are false. PLoS Medicine 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124
22. Johnson R. H., Hamby B. (2015). A meta-level approach to the problem of defining 'Critical Thinking'. *Argumentation*, 29, 417–430. https://doi.org/10.1007/s10503-015-9356-4
23. Kelter R, Schnurr A and Spies S (2023) A toolbox to demystify probabilistic and statistical paradoxes. Frontiers in Education, 8:1212419. https://doi.org/10.3389/feduc.2023.1212419
24. Krishnan, S. and Idris N. (2015). An Overview of Students' Learning Problems in Hypothesis Testing. *Jurnal Pendidikan Malayasia*, 40(2), 193-196. Retrieved July 20, 2024, from http://journalarticle.ukm.my/9445/1/Chap12new.pdf
25. Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. Perspectives on Psychological Science, 6(3), 299–312. doi: 10.1177/1745691611406925
26. Lakens, D. (2021). The Practical Alternative to the p Value Is the Correctly Used p Value. *Perspectives on Psychological Science*, 16(3), 639–648. https://doi.org/10.1177/1745691611406925
27. Matthee, M. and Turpin, M. (2019). Teaching Critical Thinking, Problem Solving, and Design Thinking: Preparing IS Students for the Future. *Journal of Information Systems Education*, 30(4) 242–252. Retrieved July 25, 2024, from http://jise.org/Volume30/n4/JISEv30n4p242.html
28. Mora-Ramirez, R. F. (2023). Use of paradoxes as didactic resources that develop critical thinking in students. *Sophia, coleccion de Filosofia de la Education*, 35, 235–263. ISSN 1390-8626. https://doi.org/10.17163/soph.n35.2023.08
29. Nickerson, R. S. and Falk, R. (2006). The exchange paradox: Probabilistic and cognitive analysis of a psychological conundrum. *Thinking & Reasoning*, 12(2), 181–213. https://doi.org/10.1080/13576500500200049
30. Pearl, J. and Mackenzie, D. (2018). *The Book of Why*: *The New Science of Cause and Effect*. Basic Books.
31. Raj, T., Chauhan, P., Mehrotra, R. and Sharma, M. (2022). Importance of Critical Thinking in the Education. *World Journal of English Language*, 12(3), 126–133. https://doi.org/10.5430/wjel.v12n3p126
32. Rauzah, Kusnandi (2019). Improving students' critical thinking abilities in probability problems through problem-based learning, In Retnowati, E., Suprapto, Jerusalem, M., Sugiyarto, K., & Wagiran, (Eds.). *Innovative Teaching and Learning Methods in Educational Systems, the 1st Edition* (eBook ISBN 9780429289897): *Proceedings of the International Conference on Teacher Education and Professional Development (INCOTEPD 2018)*, October 28, 2018, Yogyakarta, Indonesia. Routledge. https://doi.org/10.1201/9780429289897
33. Saenen, L. Heyvaert, M. Dooren, W. V., Schaeken, W., and Onghena, P. (2018). Why Humans Fail in Solving the Monty Hall Dilemma: A Systematic Review. *Psychologica Belgica*, 58(1), 128–158. https://doi.org/10.5334/pb.274
34. Sainsbury, R. M. (1995). *Paradoxes*. Cambridge University Press
35. Salido A and Dasari D (2019). Students' Errors in solving probability problems viewed by learning style. Journal of Physics: Conference Series, 1211, 012067. https://dx.doi.org/10.1088/1742-6596/1211/1/012067
36. Schmaltz R. M., Jansen E., Wenckowski N. (2017). Redefining Critical Thinking: Teaching Students to Think like Scientists. *Frontiers in Psychology*, 8:459. https://doi.org/10.3389/fpsyg.2017.00459
37. Tang, T, Vezzani, V and Eriksson, V (2020) Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity*, 37:100696. https://doi.org/10.1016/j.tsc.2020.100696

38. Thompson, C. (2011). Critical Thinking Across the Curriculum: Process over Output. *International Journal of Humanities and Social Science,* 1(9), 1-7. Retrieved July 26, 2024, from http://ijhssnet.com/journals/Vol._1_No._9_Special_Issue_July_2011/1.pdf
39. Yukalov, V. I. (2021). A Resolution of St. Petersberg Paradox. *Journal of Mathematical Economics*, 97:102537. https://doi.org/10.1016/j.jmateco.2021.102537
40. Yanagawa, T. (2020). Reproducibility of Statistical Test Results Based on p value. *Japanese Journal of Biometrics*, 40(2), 69–79. Retrieved June 26, 2024, from https://www.jstage.jst.go.jp/article/jjb/40/2/40_69/_pdf#:~:text=If%20the%20second%20test%20leads,and%20second%20statistical%20test%2C%20respectively.
41. Yu, B. and Kumbier, K. (2020). Veridical data science. *Proceedings of the National Academy of Sciences of the United States of America*, 117(8), 3920–3929. https://doi.org/10.1073/pnas.1901326117
42. Wijayatunga, P. (2019). Resolution to four probability paradoxes: Two-envelope, Wallet-game, Sleeping Beauty and Newcomb's', In Luís Meira-Machado and Gustavo Soutinho (ed.), Proceedings of the 34th International Workshop on Statistical Modeling, Guimarães, Portugal**,** II: 252–257.
43. Wijayatunga, P. (2023). Some Cases of Prediction and Inference with Uncertainty. In: Abraham, A., Hanne, T., Gandhi, N., Manghirmalani Mishra, P., Bajaj, A., Siarry, P. (eds) *Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022). SoCPaR 2022*. *Lecture Notes in Networks and Systems*, 648. Springer, Cham. https://doi.org/10.1007/978-3-031-27524-1_25
44. **Wijayatunga, P.,** Bandyopadhyay, P. S. and Woodcock, S. (2024). **Editorial: Probability and its Paradoxes for Critical Thinking. *Frontiers in Education*,** 9:1474013**.** https://doi.org/10.3389/feduc.2024.1474013
45. Zaini, Aida, Haziq Jamil, and Elvynna Leong. (2021). I Think I Understand: Investigating Misconceptions Regarding Hypothesis Test Concepts Among University Students. *EdArXiv*. September 11. https://doi.org/10.35542/osf.io/dj6x9