Open University of Cyprus

Faculty of Pure and Applied Sciences

Postgraduate (Master's) Programme of Cognitive systems

***School* of Pure and Applied Sciences**

Postgraduate (Master's) Dissertation

The effect of argumentation on cognitive functions at human - machine interaction.

AIKATERINI TSIRAKI

Supervisor
Dr.Loizos Michael

May 2022

Open University of Cyprus

Faculty of Pure and Applied Sciences

Postgraduate (Master’s) Programme of Cognitive systems
School of Pure and Applied Sciences

Postgraduate (Master’s) Dissertation

The effect of argumentation on cognitive functions at human - machine interaction.

Aikaterini Tsiraki

Supervisor
Dr.Loizos Michael

The present Postgraduate (Master’s) Dissertation was submitted in partial
fulfilment of the requirements for the postgraduate degree
in Cognitive Systems
Faculty of Pure and Applied Sciences
of the Open University of Cyprus.

May 2022

BLANK PAGE

Abstract

Attention has been focused on building argumentation based XAI which reinforces mutual explainability. Nevertheless, there are not many researches that explore the cognitive compatibility in HMI through the user perspective. Therefore, the aim of the study was to investigate some effects of argumentation on cognitive aspects such as decision making, problem solving and cognitive load in a XAI context, namely, Machine Coaching. By using a 9^3 factorial design (N=258), we saw that example based explanations mostly affect the confidence level of participants while prototype based, influence the behavior of participants. A pattern of people with high level of expertise preference of Near miss and people with lower level of expertise preference of Far miss, was visible as well and needs further investigation. Argumentation promotes not only consistency of cognitive behavior but also, higher performance. In general, our findings suggest: 1) integration of psychometrics such as trust, comprehension and persuasiveness combined with self-report tools, 2) personalized XAI which distinguishes users based on cognitive skills and background knowledge, 3) examination of Cognitive load while using EEG, and 4) further empirical studies on this matter.

Key words: argumentation, cognitive compatibility, XAI, Machine coaching, mutual explainability

Περίληψη

Η προσοχή έχει επικεντρωθεί στην οικοδόμηση εξηγήσιμης τεχνητής νοημοσύνης, βασισμένης σε επιχειρήματα, η οποία ενισχύει την αλληλεπίδραση και την αμοιβαία κατανόηση μεταξύ ανθρώπων και μηχανών. Παρ'όλα αυτά, δεν υπάρχουν πολλές εμπειρικές μελέτες που να διερευνούν τη γνωστική συμβατότητα στο τομέα διεπαφής ανθρώπου-μηχανής, υπό το πρίσμα του χρήστη. Επομένως, ο στόχος της συγκεκριμένης έρευνας ήταν να εξερευνήσει τις επιδράσεις ενός επιχειρηματολογικού πλαισίου, όπως εκείνου του Machine Coaching, σε γνωστικές λειτουργίες, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων και το γνωστικό φορτίο. Χρησιμοποιώντας έναν 9^3 παραγοντικό σχεδιασμό (N=258), παρατηρήσαμε, ότι οι επεξηγήσεις που βασίζονται σε παραδείγματα επηρεάζουν κυρίως το επίπεδο εμπιστοσύνης των συμμετεχόντων ενώ επιχειρήματα βασισμένα σε ένα πρωτότυπο παράδειγμα, επηρεάζουν τη συμπεριφορά των συμμετεχόντων. Είδαμε ακόμη, ένα μοτίβο προτίμησης ατόμων με υψηλό επίπεδο εξειδίκευσης, σε Near miss αντιπαραδείγματα και ένα μοτίβο προτίμησης ατόμων με χαμηλότερο επίπεδο εμπειρογνωμοσύνης, σε Far miss αντιπαραδείγματα, το οποίο χρήζει περαιτέρω διερεύνησης. Η επιχειρηματολογία συμπερασματικά, προάγει όχι μόνο τη γνωστική συνέπεια αλλά και την υψηλότερη επίδοση. Σε γενικές γραμμές, τα ευρήματά μας προτείνουν: 1) ενσωμάτωση και συγχώνευση ψυχομετρικών εργαλείων όπως η εμπιστοσύνη, η κατανόηση και η πειστικότητα σε συνδυασμό με μετρήσεις αυτοαναφοράς, 2) εξατομικευμένη χρήση εξηγήσιμης τεχνητής νοημοσύνης που προσαρμόζεται στις γνωστικές δεξιότητες και στο θεωρητικό υπόβαθρο κάθε χρήστη, 3) μέτρηση του γνωστικού φορτίου κατά τη χρήση ηλεκτροεγκεφαλογραφήματος (ΗΕΓ) και, 4) περαιτέρω εμπειρικές έρευνες πάνω στο θέμα αυτό.

Λέξεις κλειδιά: επιχειρηματολογία, γνωστική συμβατότητα, εξηγήσιμη τεχνητή νοημοσύνη, Machine Coaching

Acknowledgments

I would like to thank the following people for helping me throughout the process of building this dissertation:

Dr. Loizos Michael, for making this work possible with both his guidance and patience. Mrs. Xenia Pavli, Marianna Prodromou and Finiki Stoupa for their support in matters of time management and counseling. All the professors in the master programme of Cognitive Systems for their lectures, the Open University of Cyprus and University of Cyprus for offering this field of studies.

My parents, family and friends for both supporting me and helping me collect participants. Open science tools for providing the tools of constructing online surveys and their support throughout the uploading of the experiment and the extraction of the data.

My friend Xanthippi Kontogianni and her friend Dionisis Kappatos for offering a deeper insight of mathematics and probabilities. My cousin Gerasimos Pnevmatikakis for helping me understand better the mathematical equations behind my statistical analyses.

Dylan James for supporting me building a healthy mindset and complete this work by believing in myself and being productive.

Table of Contents

Contents

Chapter 1
Introduction

Artificial intelligence (AI) includes a widespread spectrum of application such as science, healthcare, finance industry, travel and transportation, social media, e- commerce and marketing. The advances in Machine learning (ML) algorithms such as Deep neural networks (DNNs [Deep learning]) or Cognitive Architectures in combination with model based Reinforcement Learning (RL) or free model based systems, have reached scores equivalent to or higher than humans in many domains and tasks (Lake, Ullman, Tenenbaum & Gershman, 2016; Vernon, Metta & Sandini, 2007). However, such AI systems not only need an extended training period of time to achieve these scores but, they are also domain/task specific, that is, they are not able to apply their skills in different areas or tasks other than those trained (Lake, Ullman, Tenenbaum & Gershman, 2016).

More specifically, they have been developed ML models that can be assigned to the following categories: Simple models vs complex models, model specific vs model agnostic and models with local vs global interpretability (Adadi & Berrada, 2018). Simple models such as decisions rules or linear regression are based on post hoc methods that promote local explainability/interpretability, that is, they can explain their decisions (Zhong & Negre, 2021; Adadi & Berrada, 2018). On the other side, complex models such as DNNs or latent factor models, are designed to stimulate accuracy (Adadi & Berrada, 2018). The level of interpretability of a model varies inversely with the accuracy. Model agnostic approaches can be used with any ML model while model specific are selective (Adadi & Berrada, 2018). Local interpretability indicates transparency on specific behaviors of the machine, contrast to global interpretability that can reveal how a ML algorithm work to make predictions (Adadi & Berrada, 2018; Zhong & Negre, 2021).

Furthermore, AI simulating the perceptual system of humans, created systems, namely, Cognitive Architectures that can be divided to cognitivist, emergent and hybrid. Cognitivist architectures are based on symbolic knowledge representation and manipulation. This means that these systems are able to interpret the external world through manipulation of rules (high level human readable reasoning) but some basic knowledge is incorporated to the system a priori and thus, it is affected by the point of view of the designer (Vernon, Metta & Sandini, 2007; Langley, Laird & Rogers, 2009). Emergent systems form representations of the world through their interaction with it. For this reason, it is unclear how they manage their knowledge acquisition, as this mechanism is treated as a "black box", by learning labels instead of statements and thus, it cannot explain the reason behind its decisions-actions (Vernon, Metta & Sandini, 2007; Langley, Laird & Rogers, 2009). Last but not least, Hybrid architectures integrate the essential advantages of the two previous architectures, that is, the transparent symbolic representations and the interactive flexibility (Vernon, Metta & Sandini, 2007; Langley, Laird & Rogers, 2009).

Overall, ML algorithms try to counterbalance statistical security with explainability with post hoc methods, while Cognitive Architectures try to align explainability with embodied adaptation. However, according to Lake, Ullman, Tenenbaum & Gershman (2016) ML strategies need to incorporate and merge important aspects such as intuitive physics, intuitive psychology, compositionality, causality and model based RL. In the same manner, as stated by Vernon, Metta & Sandini, 2007 and Langley, Laird & Rogers (2009), the evaluation criteria of a Cognitive Architecture should demand principles such as embodiment, perception, action, anticipation, adaptation, autonomy and motivation, in a balanced ratio. In other words, the main purpose of the AI is to achieve human-like capabilities in order to support humans in a cooperative manner.

Consequently, transparency, or to rephrase it, the explainable AI (XAI), is the spotlight of the research on the AI field (Adadi & Berrada, 2018). Even though there are domains that do not need transparency (Burkart & Huber, 2021; Doshi-Velez & Kim, 2018), the necessity of XAI is highlighted by the scientific community, in order to establish trust and efficiency in human-machine interactions (HMI), especially in sensitive zones such as science, healthcare, law etcetera… As mentioned by Lake, Ullman, Tenenbaum & Gershman (2016), XAI should "build causal models of the world that support explanation and understanding, rather than merely solving pattern recognition problems".

Concretely, explainability is important to maintain safety (Doshi-Velez & Kim, 2018), increase task-performance (Doshi-Velez & Kim, 2018), avoid technical depth (Doshi-Velez & Kim, 2018), build trust (Zhong & Negre, 2021; Burkart & Huber, 2021), reveal causality (Burkart & Huber, 2021; Zhong & Negre, 2021), facilitate transferability (Zhong & Negre, 2021; Burkart & Huber, 2021), advance informativeness (Burkart & Huber, 2021; Zhong & Negre, 2021), develop fair and ethical decision making or, in other words, nondiscrimination (Doshi-Velez & Kim, 2018; Zhong & Negre, 2021; Burkart & Huber, 2021), enrich confidence (Zhong & Negre, 2021), offer accessibility, interactivity and privacy awareness (Zhong & Negre, 2021).

1.1 Machine Coaching

It has to be pointed out that interpretability does not presuppose fairness and reliability (Doshi-Velez & Kim, 2018). Generally speaking, XAI tools are susceptible to biases (Alikhademi, Richardson, Drobina & Gilbert, 2021; John-Mathews, 2021; Guidotti, 2020). Therefore, AI explanations should not only be immune to algorithmic biases but also be able to help humans eliminate their own (Guidotti, 2020). It's possible that a dialectical HMI could positively contribute to this mutual disadvantage. For this reason, some researchers have been focused on mutual explainability (Michael, 2019, 2020a, n.d.; Bruckert, Finzel & Schmid, 2020; Schmid & Finzel, 2020).

To be more specific, Bruckert, Finzel & Schmid (2020) indicate the necessity of interactivity and interpretability in ML algorithms by providing a framework which can be applied to many domains. Their approach is dialogue based consisting of Mutual Explanations (ME) that allows users to be included in the machine's cognitive knowledge acquisition. Likewise, Michael (2019, 2020a, n.d.) introduces Machine Coaching that can be distinguished from Supervised Learning (SL) and Machine Programming which are on the other end of the spectrum. In Machine Coaching, the machine is able to detect and revise either the rules of a hypothesis theory or the hypothesis itself through a dialectical-online communication.

The main differences between Machine Coaching and other XAI approaches, is that Machine Coaching is based on eXplanation In, eXplanation Out (XIXO) principle which at the same time retains ML features such as induction, generalization, statistical security, that is to say, Probably Approximately Correct framework (PAC) and thus, guarantees the quality of the explanations that machines can learn to produce. Moreover, this approach is built on the interactive learning through formal argumentation which is represented with symbolic-rube based knowledge that takes the form of statements instead of labels. For instance, a combination of variables can construct a rule and by extension a set of rules can build a statement or hypothesis.

Hence, we can say that argumentation can be considered as the link between ML and XAI (Vassiliades, Bassiliades & Patkos, 2021). It is undoubtedly obvious that we might be entering to a new AI era, where HMI will be unfolded through cooperation, function allocation and joint decision making. This is feasible only by establishing the mutual explainability (Tung & Chan, 2009). It seems that bilateral explainability can create a shared or balanced cognitive burden among the two agents, as well. Particularly, as explained by Michael (2019, 2020a, n.d.) in Machine Coaching, XIXO facilitates cognitive compatibility, that is, reduced cognitive load for humans and reduced algorithmic burden for the machines because of this dialectical cooperation.

Nevertheless, as it has already been mentioned, systems that utilize symbolic knowledge representations, are affected by the viewpoint of the designers. Similarly, in Machine Coaching, machines' knowledge base is molded by their coach, that is, their user. According to Michael (2019, 2020a, 2020b, n.d.), this factuality arises some ethical issues. Briefly, if the users are biased for whatever reason, they will condition the machine to conform to their beliefs. For this reason, he queries whether challenging the beliefs of a user or imposing explanations for these beliefs could help users rearrange their thinking instead of reinforce their limiting beliefs.

In Machine Ethics through Machine Coaching, Michael (2020b), explains the context under which the Machine Coaching could taking place. In sum, the machine's cognitive functionality can be linked to that of a human-personal assistant or that of a child. In both situations, we care about the intentions of a behavior (reasons behind an action), the training period is happening in real-life situations, we correct by giving explanations and we make sense of the world trying to be objective through exchanging of opinions. Although, this mechanism is cooperative, the final responsibility either for the decision or the "character" of the machine is being assigned to the "coach".

Despite the fact that Machine Coaching aims to promote the bilateral understanding in HMI, their investigations analyze more the machine's perspective of comprehension (Michael, 2019). Identically, Miller, Howe & Sonenberg (2017) and Ehsan et al. (2021) indicate that AI programmers and researchers design XAI models which are mostly focused to technology rather than the actual cognitive needs of the users. There is a plethora of investigation of XAI through qualitatively analyses and especially from the expert-designer-engineer perspective (Bodria etal., 2021; Galhotra, Pradhan & Salimi, 2021; Rehse, Mehdiyev & Fettke, 2019; Chazette, Brunotte & Speith, 2021; Liao, Gruen & Miller, 2020, Lampridis, Guidotti & Ruggieri, 2020; Adhikari, Tax, Satta & Faeth, 2019; Guidotti, et al., 2019; Ribeiro, Singh & Guestrin, 2016).

1.2 Related work

All things considered, the contribution of social sciences such as philosophy, psychology and cognitive science will be inevitable for designing adaptive and interactive cognitive systems (Miller, Howe & Sonenberg, 2017; Abdul, Vermeulen, Wang, Lim & Kankanhalli, 2018; Vultureanu-Albişi & Bădică, 2022), since we need more empirical studies based on user's measurement of interpretability (Vassiliades, Bassiliades & Patkos, 2021b; Michael, 2019, 2020a). Along those lines, Teso & Kersting (2019) proposed an explanatory interactive learning framework by using a model-agnostic approach called CAIPI and observed that the level of trust on machines was higher if subjects were provided with explanations. They also saw, that machine's performance increased as a result of receiving counter-arguments as a feedback. Similarly, Herlocker, Konstan & Riedl (2000) and Dzindolet, Peterson, Pomranky, Pierce & Beck (2003) as referred to Ribeiro, Singh & Guestrin (2016), showed that providing explanations can boost acceptance.

But what kind of explanations have been developed along XAI lines? We will be referred to the types of ML reasoning (MR) that have been tested on users, as our aim is to examine human's perspective. Nevertheless, we will first present a basic structure of XAI explanations that seem to cover the most research approaches. As seen by Cyras et al. (2020), a Human-Centric explanation proposal in AI could be as follows: Categorization of MR as attributive (minimal set of rules and facts), contrastive (minimal modification of the given information) and actionable (minimal change of goals or resources). The categorization is based on the questions that the explanations are aimed to answer, that is, what, why, what not, what if, when and how, respectively. Nonetheless, according to their views, attributive explanations can work as functional parts of a contrastive explanation while contrastive explanations can lead to actionable decisions.

Attributive or Feature based explanations could be summarized by Example based explanations and this could be the reason of the findings of Adhikari, Tax, Satta & Faeth (2019), where Feature based explanations (creation summaries of features that describe an instance belonging to a concept) had no effect while Example-based explanations (selection of instances belonging to a concept) performed better, supporting subjects on decision making processes. The combined effect of both, has no significance. Concerning contrastive explanations, Shulner-Tal, Kuflik & Kliger (2022), tested how different styles of explanations are being perceived from non-experts regarding fairness and understanding level. Sensitivity-based explanations, that is, how various changes in the values of input characteristics will modify the result, increase the understanding level of participants. On the other side, certification-based, that is, if the system is verified as a fair system, increase the trust level of participants. Here, the researchers use the term Sensitivity-based explanations to show how some variables could change in order to give the desired results and thus, this could be equated with a contrastive explanation.

A lot of researchers such as, Ramon, Martens, Provost & Evgeniou (2020) and Schmid & Finzel (2020) have been focused on contrastive explanations. Contrastive explanations and counter-arguments are being used interchangeably and for this reason, Karimi, Schölkopf & Valera (2020, 2021) suggest to use the term contrastive instead of counter-arguments. They also say that the research should be directed on explanations which are based on recourse actions (minimal consequential recommendations) instead of counterfactual examples (nearest contrastive explanations) that result from those actions. In contrast to this view, Wachter, Mittelstadt & Russell (2017) claim that counterfactuals should be preferred as an algorithmic form of providing explanations to users. This is because the abovementioned types provide reasons of why an outcome has occurred by offering a satisfied insight of how you can reach the desired result and without revealing their internal algorithmic mechanism.

Regarding this technique of reasoning presentation, Winston (1970) as referred to Rabold, Siebers & Schmid (2021) showed that presenting Near Miss examples (negative example that differ minimally from a given positive example) in relational domains, leads to faster ML. This efficacy can be applied on explaining a learned model as well (Rabold, Siebers & Schmid, 2021). With an empirical study, Kobayashi (2019) found that the simultaneous presentation of contradictive scientific and social consensus information on the safety of genetically modified foods affected the subjects' beliefs compared to the presentation of each of the information, separately. This might be explained due to the fact that individuals had the opportunity to compare these two contrastive versions and not due to the nature of the information provider as the paper is being claimed.

Rabold, Siebers & Schmid (2021) also used contrastive explanations. Specifically, they conducted an empirical study using Rule based (if-then statements, global explanation of the concept that the specific instance belongs to), Example based, Near and Far Miss explanations (negative example with high or low degree of structural similarity to specific instance, respectively), combined and separately, for an abstract relational family domain and a visual relational arches domain. Subjects had to evaluate the usefulness of the explanations. In the abstract domain, Rule based explanations were preferred while in the paired version, Rule-Example based explanations were preferred. In the visual domain, Examples were preferred while in the paired exposure, Example-Nearest were most preferred. Some differences in the results could be due to the character of the explanation, that is, if they were presented as a natural language structure or as a visual illustration.

Another study of Hase & Bansal (2020) tested users' understandability of machine's functionality and they found that Local Interpretable Model-agnostic Explanation (LIME) for tabular data and Prototype method (selection of representative instances of a concept or category) in counterfactual tests, helps subjects more to predict machine's algorithmic behavior. Lastly, Ferrario & Loi (2020) and Verma, Dickerson & Hines (2020) suggested ways to adapt counterfactual explanations in real-life situations. However, Slack, Hilgard, Lakkaraju & Singh (2021) and Rosenfeld (2021) propose some metrics that should be considered for evaluating XAI explanations before being tested by users, in order to avoid confirmation biases.

These metrics suggest to compare the performance of the transparent model with the performance of a "black box" model, to simplify the model by using less features and rules and, find stable features by adding noise to avoid overfitting which reduces the model's flexibility. On the other side, the quality of an explanation depends not only on relevance but also on persuasiveness and comprehension that provides (Burkart & Huber, 2021). For this reason, we should avoid to use one-size fits all approach, since different questions, target-groups and contexts demand different kind of explanations (Mothilal, Sharma & Tan, 2021; Amann, et al., 2022; Burkart & Huber, 2021; Schelenz, Segal & Gal, 2020).

Personalization that depends on different cognitive styles of people, could also be a relevant option in the future, for example if a person prefers an explanation to be presented visually or in a text.

There are a lot of researches concerning this matter, namely, adaptive and interactive systems which deal with the cognitive compatibility between humans and machines (Angeli, Valanides & Kirschner, 2009; Ugwitz et al., 2019; Belk, Germanakos, Constantinides & Samaras, 2015; Belk, Fidas, Germanakos & Samaras, 2017; Chen & Liu, 2011; Katsini et al., 2018; Alves et al., 2020; Herbig et al., 2020; Raptis, Fidas & Avouris, 2018).

1.3 Research aims and theoretical approaches

Putting it all together, we have summarily discussed the current XAI trends by examining both algorithmic parameters and cognitive requirements. As we have seen, there is a lack of researching people's perspective in an argumentation context with XAI. There are some surveys with conflicting or different results due to different usage of explanations, stimuli and experimental settings. The aim is neither to make people conform to the machine's reasoning capacities nor for the machines to comply to the multi-dimensional needs of humans, at least not yet, but to create an argumentation environment that is cognitively and algorithmically compatible for both parts. This being said, the purpose of the study is to investigate human's cognitive functions in an argumentative context with an AI system. This study will not use ML tools for the explanations as it exclusively and explicitly intends to explore human's cognitive needs, nevertheless the subjects will be aware that they are interacting with a machine.

From our theory report, we selected three matters. We saw a research exploration of choosing the best type of explanation for humans and that is our first topic of investigation. Then, we saw a new approach called Machine Coaching which is based on XIXO principle. This is an interactive approach which simulates, to the greatest extent, a real-case dialogue scenario between humans. As Michael (2019, 2020a, 2020b, n.d.) has pointed out, Machine Coaching aims to counterbalance both cognitive and algorithmic resources. Thus, our second topic is to measure cognitive load in a Machine Coaching context. To proceed, we also saw that XAI Symbolic knowledge can be biased. Particularly, Michael (2019, 2020a, 2020b, n.d.) said that we should test whether specific argumentation conditions facilitate a constructive and progressing dialogue, or people tend to be attached to their cognitive fallacies regardless the case. And so, our third point is to measure how different dialogue conditions, such as those of Machine Coaching, affect problem solving.

Before referring further to our experimental variables, we should present the theoretical background in which we will base our hypotheses. We will start with mental categorization which leads to human perception and attention and by extension to decisions, choices and problem solving. So, a fundamental part of our perceptual system is the organization of the external information. In order for the mind to interpret a complex formation, it needs to organize the perceptual elements in a simpler structure (Arnheim, 1997). The perceptual coding through basic general simple forms is ecumenical, from insects to primates and, it starts at a very young age (Arnheim, 1974). This organization of knowledge is flexible and allows multi-dimensional arrangements (Πόθος & Οικονόμου, 2010; Arnheim, 1974). The Laws of Gestalts are some grouping principles that affect our perceptual organization (Eysenck, 2006; Arnheim, 1974). These laws are related to Proximity, Simplicity-Pragnaz, Similarity, Good continuation, Common fate, Figure-ground, Symmetry/Order, Closure, Common region, Focal points, Parallelism and Past experiences. Although, the level of dominance of each of these principles or the kind of similarity that is preferred each time, it's not quite clear yet (Arnheim, R, 1974).

Consequently, categories are the basic elements that structure our thinking (Πόθος & Οικονόμου, 2010). It has to be noted that, in categorization, an object equals to a cognitive experience and a category is equivalent with a concept. The three most popular theories of cognitive categorization are the classical view theory, the conceptual clustering and the prototype theory (Eysenck, 2006; Πόθος & Οικονόμου, 2010). In the first theory, a category can be defined as a set of attributes which are singly necessary and jointly sufficient and it is expressed in a form of rule/definition/theory. For this reason, it can be related to Rule based ML explanations. In the second theory, a category is defined as a set of examples which have attributes that are based on rules and thus, it can be related to Example based ML explanations. Lastly, in the third theory, according to Posner & Keele (1968) and Smith & Minda (2000) as referred to Πόθος & Οικονόμου (2010), a category is expressed through a prototype which equals the average of all the examples of this category and therefore, it can be related to Prototype based ML explanation.

The memory of previous objects affects not only the codification of new information but also our perception (Arnheim, 1974; Hogg & Vaughan, 2008). Stored information can guide attention in an unconscious level and thus, former perceptual representations can lead to targeted attentional behavior which leads to new perceptual information (Spitz et al., 2016; Noah & Mangun, 2019).

Hence, categorization not only allows us to relate old experiences with new and but also to perceptually operate in a higher level of creativity and mental abstraction (Πόθος & Οικονόμου, 2010). The most categorization processes are automated and unconscious (Eysenck, 2006; Πόθος & Οικονόμου, 2010). Shank & Abelson (1977) as referred to Πόθος & Οικονόμου (2010) call these perceptual organizations "schemata and scripts". We tend to fill the gaps in our perception with schemata that have been acquired through repetitive correlations of stimuli which are stored in our memory (Hogg & Vaughan, 2008). For example, Tompkins, Woods & Aimola Davies, (2016) showed that people can perceive a non-existing object due to their expectation. So, the external information we receive is framed through a multiple number of assumptions.

In addition to this, as explained by Van de Cruys & Wagemans (2011), this process of categorization and storing of information, facilitates learning. More specifically, our brain does not register external stimuli through bottom-up processes, but it makes predictions based on top-down calculations. Our sense organs transfer the signal to the brain and this signal is modified to a conditioned mental representation which is aligned to the stored and activated data. Then, our perception organize these external data as gestalts. If a gestalt is proved to be successful in the past, is preferred because extensive sensory calculating is costly. Similarly, Cheng & Holyoak (1985) as referred to Πόθος & Οικονόμου (2010), supported that our perceptual system has adopted "permission schemata" that were proved to be successful for resolving ordinary and important situations in the past. Prediction errors that leads to learning happen when the activated schema/gestalt receives negative feedback from the environment.

We can meet insights of this notion on investigations of the developmental psychology of children as well (Piaget, 1928). Schemata is the basic element that structure children's perception as their behavior or choices cannot be orally explained. The child makes a right decision but without conscious awareness. According to Chaplin, John & Goldberg (1988) as referred to Hogg & Vaughan (2008) in the context of categorization, there are a lot of types of schemata which are fuzzy sets centered on an archetype. The archetype could be the mean object/idea of a category or an extreme member of a category. Therefore, categorization can produce stereotyping. Even though the categorization is an ecumenical process, the interpretation of an idea or concept is not. These differences on categorization processes can be explained through individual variations based on cognitive functions, previous experiences and background knowledge (Eysenck, 2006; Πόθος

& Οικονόμου, 2010). In general the categorizations that are related to perceptual characteristics or profound functions appear to be more similar among people than those which are related to complex or abstract concepts (Πόθος & Οικονόμου, 2010).

Predictably, human's reasoning could be biased because it is based on these scripts (Miller, 2019). Reasoning can differ due to the same reasons, that is, prior knowledge, cognitive skills, expertise, culture and age (Miller, 2019), while persuasive explanations are linked with cognitive functions, user preferences and expertise (Burkart & Huber, 2021). More specifically, humans tend to create mental models by quoting counter-examples to test assumptions which are not aligned with these initial mental models (Eysenck, 2006; Miller, 2019). Categorization is based on these causal calculations and a subset of the necessary causes is chosen for creating an explanation (Miller, 2019). People with limited working memory can only create one mental model, which is not enough to examine these models through a lot of different perspectives (Eysenck, 2006). That is why people tend to be better at evaluating the arguments of others than their own, especially when trying to investigate the truth instead of attempting to win a debate (Mercier & Sperber, 2011; Mercier, 2016). As Mercier (2016) has stated "Contexts in which argument production strongly dominates over argument evaluation – that is, in the absence of dialog and of conflict between views – often lead to epistemically and practically deleterious outcomes".

Despite the fact that people are better at evaluation than production of explanations, they use the same mental metrics on both of these processes. In a causal chaining people prefer generality (Miller, 2019), simplicity (Miller, 2019), completeness (Miller, 2019), abnormality (Burkart & Huber, 2021; Miller, Howe & Sonenberg, 2017; Miller, 2019), stereotypical examples (Miller, 2019), prioritization of features (Miller, 2019), proximal and controllable events (Miller, Howe & Sonenberg, 2017). As we can see, people use similar laws for both perceptual categorization and reasoning (see p. 18: Laws of Gestalt). To sum up, people search for the cause behind a fact (Miller, Howe & Sonenberg, 2017), "they explain the cause of an event relative to some other event that did not occur" and prefer contrastive and selective explanations, while probability and truth are less important variables (Miller, 2019).

The way we organize our knowledge plays also a determinable role to the degree we will use this knowledge to problem solving processes (Πόθος & Οικονόμου, 2010). Hence, as it has already been mentioned, people make assumptions based on memory and not based on on-line processing

(Hogg & Vaughan, 2008). Particularly, our cognitive system has developed strategies such as heuristics and schemata that can produce rapid responses and can be applied regardless the quantity of the available information (Eysenck, 2006; Πόθος & Οικονόμου, 2010). An accurate prediction contributes to homeostasis because a shorter reaction time can increase the ability of an individual to survive (Van de Cruys & Wagemans, 2011). To the opposite, extensive cognitive processes are useful, when the quantity of the available information is large but, they are time-consuming (Eysenck, 2006).

Some of these strategies are the confirmation biases, the preference of using the symbols that are already used in the formulation of a presented rule, the selection of the most believable case, anchoring and adjustment (Hogg & Vaughan, 2008; Πόθος & Οικονόμου, 2010). The believability is affected by two variables, the availability and the representativeness (Πόθος & Οικονόμου, 2010). The former is referring to the convenience of retrieving examples (easy examples) and the latter to the level of the typicality of the example (high level of modality). Along those lines, a lot of researches have shown that people tend to choose the most credible choice rather than the most probable (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974; Πόθος & Οικονόμου, 2010).

In general, for ordinary problem solving tasks, people do not follow logical rules with the exception of those who are scientific or mathematical trained (Πόθος & Οικονόμου, 2010). Humans cannot process data that are not provided with and therefore it's difficult to base their thinking on probabilities and percentages (Eysenck, 2006). Although according to Cosmides & Tooby (1996) as referred to Eysenck (2006) we are better on judging frequencies compared to probabilities. Some biased judgments can lead to severe impacts and thus, Frog et al. (1982) as referred to Hogg & Vaughan (2008) proposed that these systematic false thinking errors can be improved through official training on scientific and rational thinking with statistical techniques.

Kahneman & Fredrick (2002) and Kahneman (2003) as referred to Eysenck (2006) summarize these cognitive processes with the Dual processing model. This model consists of two systems. System type I is intuitive, automate and direct. The judgments which are based on this, are rapid, automate, effortless, associative and implicit. Despite the fact that most of heuristics are being produced through this filter, these strategies are not creative, that is, they are difficult to be adapted to changes. System type II, is analytical, well controlled and being governed by rules. The responses which are based on this system, are slow, serial, demanding, conscious and flexible. The same idea is supported by other researchers as well. Novick & Sherman (2003) as referred to Eysenck (2006), stated that intuitive solutions are a result of parallel unconscious calculations while not intuitive solutions are the outcome of conscious serial calculations (Novick & Sherman, 2003). Some other similar theoretical models are the Elaboration Likelihood Model of Petty & Cacioppo (1986b) and the Heuristic-Systematic Model of information processing of Chaiken which claim that the exportation of conclusions can be filter through a top-down or a bottom-up procedure (Hogg & Vaughan, 2008). The former is based on schemata and the latter is relied on the voluntary process of the direct external stimuli.

Some factors that affect the system which will be used are, the importance/necessity of the task, the cognitive skills of the individual, the specialized knowledge, the time pressure and, the nature of the task (multi-tasking or not) (Eysenck, 2006; Πόθος & Οικονόμου, 2010). It has to be noted that our ordinary decisions are mostly affected by emotions and motives. For example we are more sensitive to potential loss than potential gains. To be more specific, emotion interacts with attention and thus, modify perception (Stanley, Ferneyhough & Phelps, 2009). When attentional resources are limited, attention is focused on emotional stimuli and if those are negative then attention is harder to disengage. Emotion can alter all stages of perception, that is, encoding, storage and retrieval. Apparently, if we are motivated to focus on a task, we use our central conscious mechanisms, otherwise we activate our heuristics-simple rules of decision making (Hogg & Vaughan, 2008).

Let's see some conditions which can trigger the system type II. Van Stekelenburg, Schaap, Veling & Buijzen (2021) found that communicating scientific consensus can be effective in eliciting scientifically accurate beliefs, while Altay et al. (2022) discovered that people can update their beliefs if they are exposed to good arguments and counter-arguments. Likewise, the consider-the-opposite technique is proved to be effective for eliminating biases (Lee et al., 2016). Furthermore, McGuire (1964) as referred to Hogg & Vaughan (2008), introduced the term "inoculation" where people that have been exposed to a blunt counter-argument, can develop more efficient rebuttals for subsequent stronger arguments.

On the other hand, let's see conditions that can trigger the system type I. Sometimes, familiar scenarios may be in fact different, nevertheless they activate heuristics and make people provide incorrect answers. This is what happened in the study of Lee et al. (2016), where experts made the wrong decisions by using their heuristics while non-experts who did not have existing heuristics, were more cautious and had better performance. Additionally, some theories such as the Consistency and the Balance theory claim that people tend to maintain an internal balance by constructing coping mechanisms in order to avoid inconsistencies in their cognitive and behavioral attitudes (Hogg & Vaughan, 2008). Lastly, the theory of Brehm speaks about Reactance, which is a motivational reaction of people feeling threatened because of their eliminating freedom. Cialdini & Petty (1979) and Johnson (1994) as referred to Hogg & Vaughan (2008), add "warning-induced, resistance arise" because people will resist if they aware that they will be the target of a persuasion effort.

In short, we have talked about categorization processes which lead to schemata and heuristics, we have seen that counter-arguments are preferred both in categorization and reasoning, we have analyzed the perceptual processing systems and now we will analyze more how cognitive load interferes with these systems. We have already mentioned that limited working memory prevents people to counter-fact a lot of different scenarios at the same time. For this reason, debates help people consider aspects that were previously invisible. In this manner, as explained by Eysenck (2006), multi-tasking, that is the performance of two or more tasks simultaneously, is characterized by a Psychological Refractory Period (PRP) which affects the performance on the secondary task. The measurement of cognitive load usually consists of two tasks, that is, the primary and the secondary task. We measure the performance on the second task.

According to Lavie, Beck & Konstantinou (2014) as referred to Noah & Mangun (2019), on low cognitive load conditions, the perceptual processing is automatic. On high cognitive load conditions, attention becomes more selective and filters out irrelevant information. Thus, cognitive-perceptual load determines the threshold for conscious perception of visual stimuli. Moreover, Cohen, Alvarez & Nakayama (2011) say that under demanding tasks, inattentional blindness can occur. For example, people cannot perceive existing objects by the prevention of the interaction of the three stages of attention (Demacheva et al., 2012). Under this context, Park & Brünken (2014) used the precision of the performance on a secondary task, as an indicator for cognitive load: the higher the precision, the lower the cognitive load. These results provide evidence that rhythm precision allows for a precise and continuous measurement of cognitive load during learning environments.

As interpreted by Eysenck (2006), practicing improves the multi-tasking execution because they become automate scripts that are retrieved from procedural memory. This can lead to an absolute optimization if the tasks directly relate the nature of their stimulus with the type of response which they demand. Alternatively, it’s easier for two tasks to be executed together if they are different on the stimulus sensor and/or the kind of response that is needed. Nevertheless, multi-tasking can affect cognitive load regardless the difficulty of the combined tasks. This has been shown in researches of D’esposito et al. (1995) as referred to Eysenck (2006), where the functional activity of Prefrontal Cortex and the front part of the Cingulate Cortex which are related to attention processes, were significant. In addition to that the complexity of the tasks is an important variable as well.

Moving on, let’s see further our experimental variables now. We will start with our first topic which is the types of arguments that are best fitting for people. We define “best fitting” in terms of persuasiveness, trust and comprehension. The dimensions of relevance and performance are concepts that should be examined through algorithmic procedures. Therefore, we want to measure how different types of arguments affect the decision making process of participants. Here, it has to be noted that according to Hastie (2001) as referred to Eysenck (2006), estimation is the first stage of the decision making process. Our hypothesis is that if a specific type of explanation is more persuasive and clear, it will direct to a greater degree both the choices and the confidence level of participants throughout the process.

Regarding the types of explanations, we used rule based, example based and prototype based, being inspired by the classical view theory of cognitive categorization, the conceptual clustering and the prototype theory. We also created a paired variation where each of these types was combined with Near miss and Far miss versions and thus, we had 3x3=9 types of explanation. As we have seen the categorization plays significant role in reasoning and comprehension. We have also referred to human’s reasoning which is based on counter-factual logic. For this reason we created the above-mentioned combinations. For the argumentation context we used probability quizzes which are based on heuristics and biases. Even though, this research is explorative, we hypothesize that experts that is, people whose their field is related to probabilities, will prefer prototype-Near miss explanations while non-experts will prefer rule based-Near or Far miss explanations.

Hence, we will measure the consistency and the variance of confidence level throughout the different types of explanation to see if a certain type outweighs the others. More specifically, we only used trust and persuasiveness, as indicators of decision making, because we have excluded the comprehension. We had to create a self-report interview and choose a subgroup of our sample for the application due to the complexity of the experiment. The lack of time was the main reason for the exclusion. Therefore, for trust, we measured their confidence level, such as, if the confidence for their decisions were minimized, then the trust in the machine is increased. And for persuasiveness we measured the changes in their decisions (consistency).

We will continue with our second topic which is the measurement of cognitive load in Machine Coaching context. We chose three different kind of interactions for this. The first condition is a control condition where the interaction is unilateral. The participant receives feedback (explanations) concerning his decisions within the task. Here, subjects interact with the machine without using intellectual means of communication (argumentation). In the second condition the interaction is bilateral and the participant initiates the interaction by using arguments to explain his/her decisions. The last condition includes a bilateral interaction as well but here the participant reacts either congruently or incongruently to the explanations of the machine. So, in this condition, the subject is using arguments to explain their decision to the machine as a response to an argument/explanation of the machine. We want to examine the performance on cognitive load between these three conditions.

The Cognitive load which is expressed as an intellectual level of performance, is divided into three aspects: intrinsic (number of elements and element interactivity), extraneous (instructional presentation of any kind of material) and, germane (construction, automation and storage of schemata). Angeli, Valanides & Kirschner (2009) argue that when the complexity of intrinsic cognitive load is high, we have to minimize the extraneous cognitive load, that is, to be more affordable and understandable for users, in order for them to be able to use their germane load properly. Germane cognitive load is referring to our ability to restore our already established schemata in our working memory in order to relate them with the external stimuli while recreating our direct representations/thoughts.

Similarly, in our case we do not care about the intrinsic cognitive load which is related to the type of the information that is communicated but we are focused on the different types of interaction that can be related to the extraneous cognitive load. So the extraneous cognitive load (different types of interaction) is the variable that chose to manipulate in order to measure the germane cognitive load. For our specific case, we utilized a secondary task where the participants will be hearing either the letter D or B. They will have to press the spacebar while hearing the letter D while responding to a confidence scale and after or before having received the feedback of the machine for third and second condition respectively.

In this way, we will see how their cognitive load is affected after the different types of interaction of each of the conditions. As we can see the first task is visual and the second aural but they have to use their hand in both cases. The level of complexity in the first task is low while in the second task is medium. For these reasons the combined complexity of these two tasks is medium and we should see different variations of performance within the conditions. We will specifically measure reaction time for both correct and incorrect responses and both the percentage of correct answers (success rate) and incorrect responses (failure rate) within the trials. We hypothesized that in the third condition where the participants react to an explanation, their cognitive load will be lower while responding to the confidence scale because their cognitive resources will be lower as well, as a result of the previous shared problem solving task. We also used a control task to measure the short-term memory capacity of each of the participants.

We proceed to our third and last topic which is how different types of intellectual interaction within Machine Coaching context, affect problem solving. So, specifically, we have again three conditions: in the first condition the participants receive feedback/explanation for their choices, in the second condition the participants explain their decisions and are provided with their own previous arguments after they have received the feedback of the machine and, in the third condition the participants react to the explanation of the machine by providing their own counter-arguments. We will measure the consistency in their choices and the decrease or increase of incorrect answers throughout the three different experimental conditions of argumentation. Again, this task is explorative but we assume that either of the last two conditions should help participants, especially experts, to eliminate their fallacies. Our hypothesis is that in the second and third condition, participants and especially experts, will be more consistent in their choices and will have better performance as well.

Chapter 2
Research

The aim of the study was to investigate the effects of argumentation on humans' cognitive functions at Machine Coaching. More specifically, we wanted to examine how different types of machines' arguments, through different kinds of interaction, influence humans' cognitive functions such as decision making, cognitive load and problem solving, at argumentation based XAI context. Our main research questions were: 1) What types of arguments are best fitting for people? 2) What is the cognitive load of humans at different types of interaction? 3) Do we get better consistency as a result of using arguments? More specifically, our main hypothesis was that argumentation affects positively the cognitive functions at human-machine interaction. Even though our research was explorative, we created some more focused hypotheses: 1) Prototype+ Near miss based explanations are most suitable types for experts, 2) Rule+ Near miss based explanations are most suitable for non-experts, 3) Machine Coaching affects positively the cognitive load of humans, 4) Argumentation helps people eliminate their biases and, 5) Argumentation promotes consistency. The research was carried out in one phase but was divided into three sub-experiments. The segmentation was applied for experimental and practical reasons to help participants stay cognitively detached within the different conditions of the experiment by simultaneously maintaining a short and less complex process. For this reason, these three experiments followed both the same methodology and sampling. They were aimed to measure the effect of nine types of arguments (Rule based, Example based, Prototype based, R+ Near miss, E+ Near miss, P+ Near miss, R+ Far miss, E+ Far Miss, P+ Far miss) on cognitive functions (decision making, cognitive load and, problem solving) at three different versions of human-machine interaction (unilateral, bilateral with human initiation and bilateral with human reaction). All statistical analyses were performed using the statistical program of SPSS.21.

Methods

Participants

N=258 people (172 Male; 86 Female) ranging in age from 18 to 68 (M = 35.9, SD = 13.5) participated in this research. Participants have been selected randomly. They either have been volunteered or been asked to participate in the experiment. For the invitation process, it was used a convenience sampling where we tried to collect participants of all the different degrees of specialization. An a priori power analysis was conducted using G*Power3 (Faul, Erdfelder, Lang, & Buchner, 2007) to test the difference between three independent group means using a two-tailed test, a small effect size (d= .25), and an alpha of .05. Result showed that a total sample of 259 participants with three equal sized groups of n= 86.3 was required to achieve a power of .80. Therefore, we used a total number of 261 which was divided to three groups of 87 people. Three people were excluded from the statistical analyses, as they gave incomplete results. The only requirement was the normal or corrected to normal vision and having access to the audio sound of their devices. We can see the descriptives of the sample on both Table 1 and Table 2.

Table 1. Descriptive statistics of 258 participants in current study.

		n	**%**
Gender	*Men / Women*	172 / 86	66.7 / 33.3
Age, *years*	*mean age (stand. dev.)[min, max]*	35.9 (13.5) [18, 68]	
Level of education	*less than HS diploma*	10	3.9
	HS	17	6.6
	BSc	108	41.9
	MSc	98	38.0
	PhD	15	5.8
	Other	10	3.9

Table 2. Frequency distribution of profession of 258 participants in current study.

	n	%
Agriculture/Geology	6	2.3
Archeology	1	0.4
Art/Music/Literature	5	2.0
Artificial Intelligence	2	0.8
Astrophysics	1	0.4
Biology	2	0.8
Biomedical Engineering	1	0.4
Business Administration/Project management	4	1.6
Chemical Engineering	1	0.4
Chemistry	2	0.8
Civil Engineering	4	1.6
Computer Science	10	3.9
Economics/Innovation/Accounting/Finance/Marketing	16	6.3
Education science	17	6.6
Electrical-Computer Engineering	6	2.4
Electrical-Mechanical Engineering	10	3.9
Engineering	4	1.6
English/Italian Literature	5	2.0
Health Science	4	1.6
Legal Science	3	1.2
Linguistics	1	0.4
Mathematical Physics	1	0.4
Mathematics	13	5.0
Mathematics-Computer Science	1	0.4
Medical school/Pharmacy	3	1.2
Medicine-Physics	1	0.4
Philology/Philosophy/Theoretical sciences	11	4.2
Physics	13	5.0
Political Science	3	1.2
Psychology	22	8.5
Social Sciences	13	5.0
Speech Therapy	4	1.6
Sports Science/Physiotherapy	6	2.4
Statistics	2	0.8
STEM	10	3.9
Unknown	50	19.4
Total	258	100.0

Statistical Analyses

The analyses were conducted using the SPSS software (IBM Corp. Released 2021, IBM SPSS Statistics for Windows, v.28.0, Armonk, NY: IBM Corp.). Frequency distributions of descriptive characteristics of the 258 participants were calculated. The normality of confidence rates (initial values and Δ-changes) was assessed based on Blom's method (Q-Q plot). Repeated-measures analysis of variance was used to assess the Δ-changes of self-confidence levels. As covariates were used the age and gender. Chi-square method was also performed to assess relationships between answers (correct, incorrect), types of explanation (9 types), consistency, specialization, digit span (short term memory capacity) or conditions as well as the Kruskal-Wallis method in continues variables. The accepted critical level was set to 0.05.

2.1 First Experiment

Method

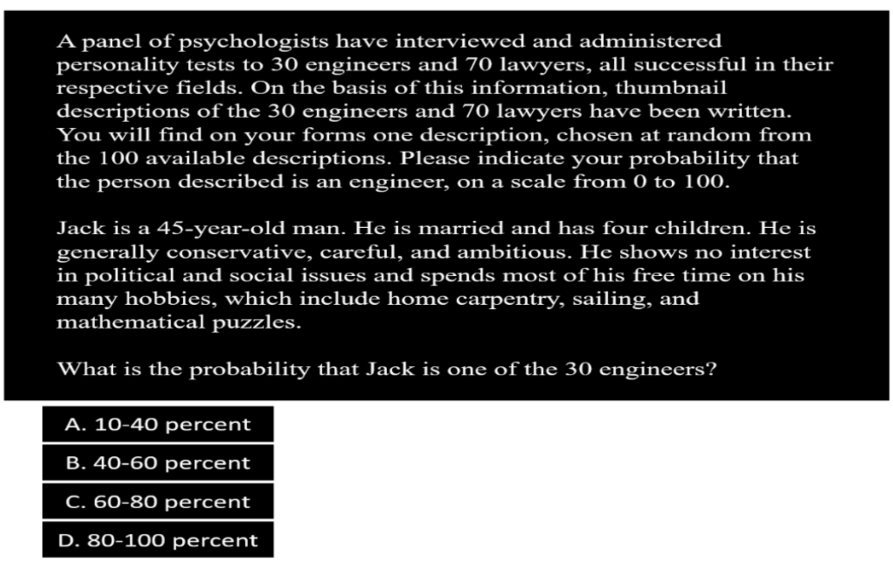
Participants

N=86 people (61 Male; 25 Female) ranging in age from 18 to 68 ($M = 35.9$, $SD = 13.5$) participated in this experiment. Participants have been selected randomly. They either have been volunteered or been asked to participate in the experiment. The only requirement was the normal or corrected to normal vision and having access to the audio sound of their devices.

Materials

Nine probability quizzes were used in this research. An example of the quizzes can be seen in Figure 1. These quizzes were presented via the screen of a computer in a randomized fashion. After the completion of each of the quizzes, participants were provided with feedback. Each feedback was corresponded to a different type of explanation. The types of explanations was a merge of the classical view theory, the conceptual clustering and the prototype theory of cognitive categorization and Near miss and Far miss of ML explanations (Rule based, Example based, Prototype based, R+ Near miss, E+ Near miss, P+ Near miss, R+ Far miss, E+ Far Miss, P+ Far miss). The presentation of explanations for each of the quizzes was randomized as well. Each type of explanation was conditioned to appear only once to each participant and thus, for each of the quizzes, the participants received different type of explanation (counterbalancing). With this way, we created all the combinations of quizzes and explanations. The quizzes, as it has already been mentioned, are based on heuristics and biases. Because all types of biases are a result of previous experiences which lead to categorization which creates heuristics that reinforce biases, we thought that these quizzes were a representative example to test behavior and cognitive biases in an argumentative context of HMI. The selected quizzes had different degrees of difficulty and the mathematical equations were meant for people who can understand. Last but not least, the quizzes can be placed on the abstract domain. The difference in this experiment compared to the existed empirical studies, is that the level of abstraction here is high. They have only used low level of abstraction domains such as pedigrees.

Figure 1. Quiz 1.



More specifically, the classical view theory category was consisted of a simple definition as an explanation, the definition compared to a contrastive definition with high degree of structural similarity (similar definition) to that and the definition compared to a contrastive definition with low degree of structural similarity to that. The conceptual clustering category was consisted of examples of a theory/definition, examples of the right theory/definition in comparison to examples of a contrastive theory/definition with high degree of structural similarity to that and examples of the right theory/definition in comparison to examples of a contrastive definition/theory with low degree of structural similarity to that. Finally, the prototype theory was consisted of a prototype example of the right theory/definition, a prototype example of the right theory/definition compared to a prototype example of a contrastive definition with high degree of structural similarity to that and, a prototype example of the right theory in contrast to a prototype example of a contrastive definition with low degree of structural similarity to that. As we have already analyzed in the introduction, these combination of theories were used because mental categorization plays dominant role to internal representation, information retrieval, reasoning, comprehension and problem solving. Additionally, Near and Far miss types have been used because comparisons have shown to help humans reason, create mental models and comprehend data. Some examples of explanations can be seen in Figures 2-6.

Figure 2. Rule based Type.

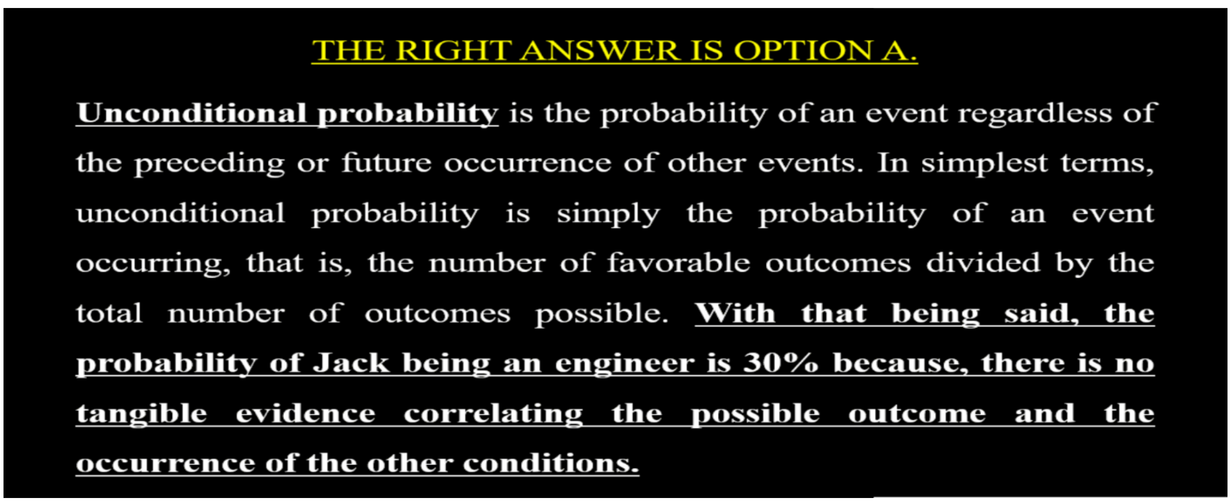


Figure 3. R+ Far miss.

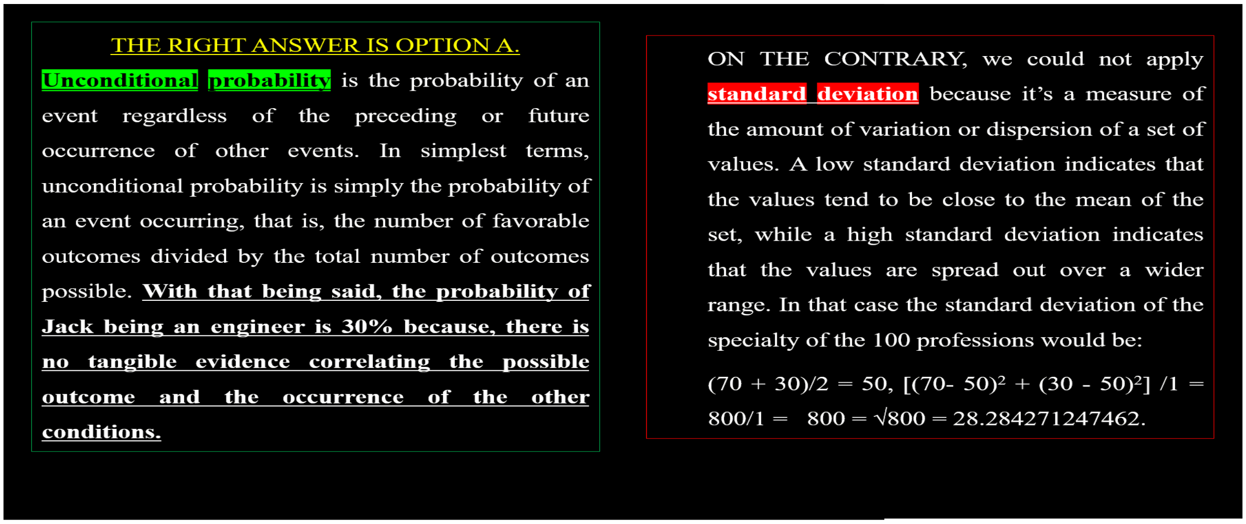


Figure 4. R+ Near miss.

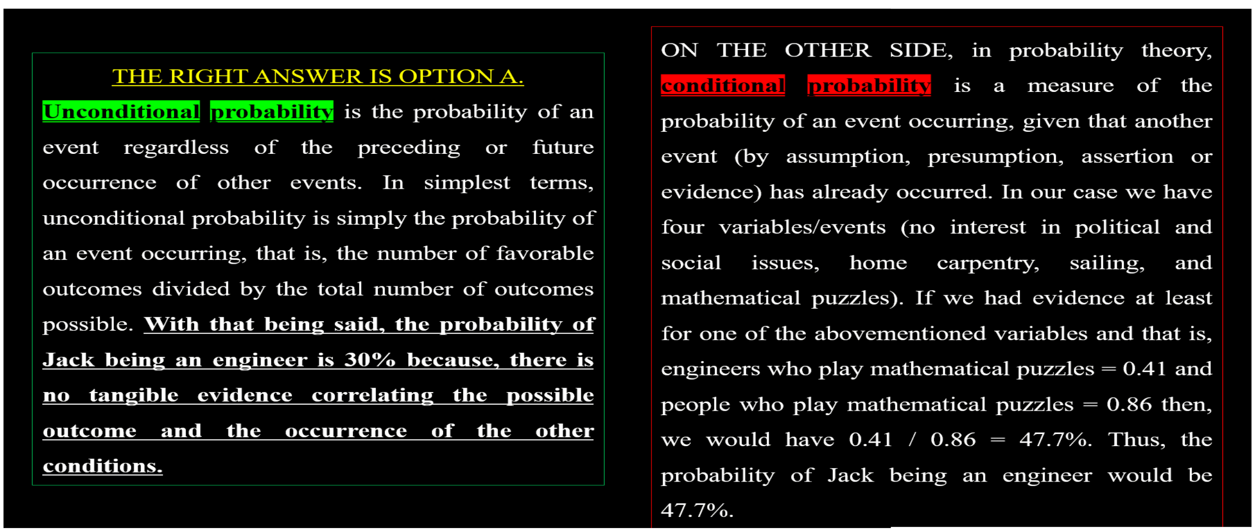


Figure 5. Example based type.

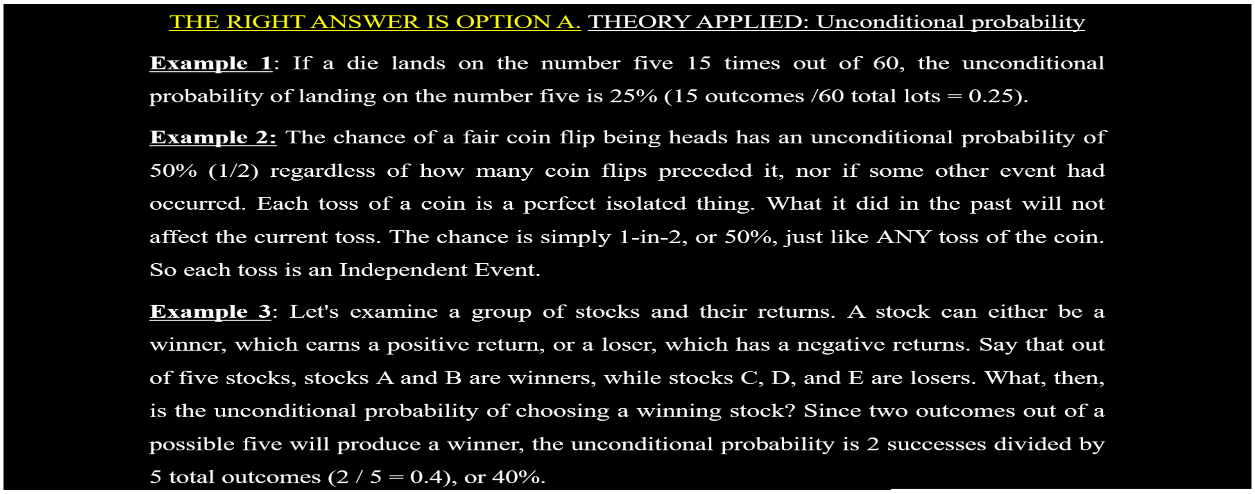
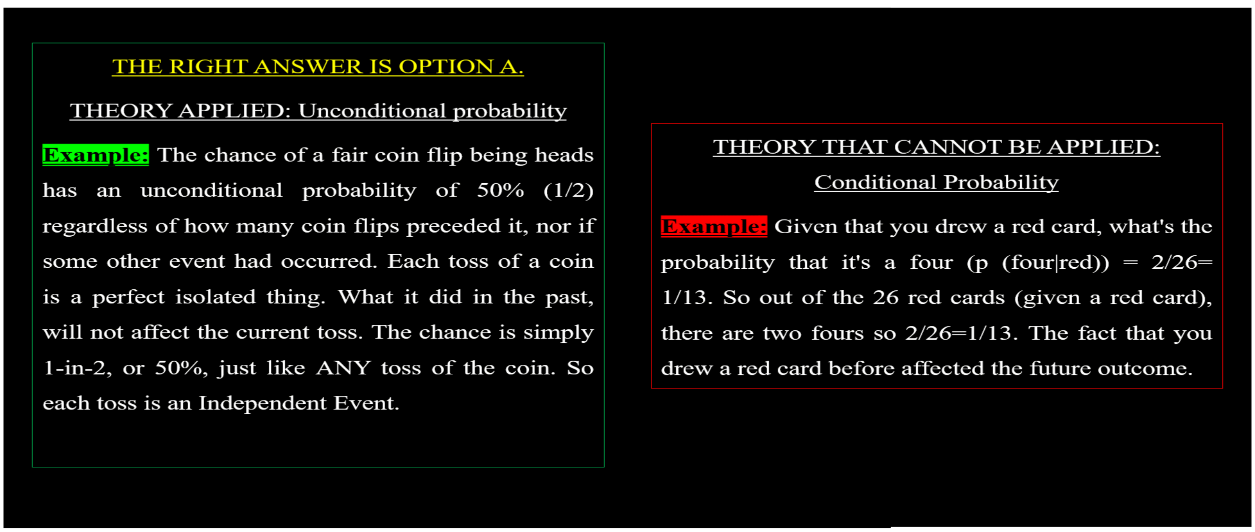


Figure 6. P+ Near miss.



A confidence rating scale was used to measure participants' confidence level of their responses, before and after the feedback of the machine (10-25%=Not confident at all, 30-45%= Slightly confident, 50-65%= Somewhat confident, 70-75%= Fairly confident, 90-100%= Completely confident). As it has already been explained before, the confidence rating was aimed to reveal trust levels of participants for the machine responses.

Nine random digits were initially used in each condition for testing the typical short-term memory capacity of the participants, in a digit span task. Subjects were asked to recall and type in the screen of the computer, using the keyboard, each of the nine digits, in the order that were previously displayed. This task was useful in order to compare potential differences in responses and the cognitive load among people with high and low short term memory capacity. Participants' levels of short-term memory capacity were divided to low, medium and high. This task was repeated three times in order to counterbalance factors such as lack of attention/distraction, lack of comprehension for the first trials and any other contributed factor.

Within this procedure of the experiment and after the digit memory task, they were used aural stimuli such as the sound of the letter b and the sound of the letter d, interchangeably and in a random manner, while participants were interacting with the machine. This choice was made to avoid the familiarization of participants with the sounds. The participants had to tap the spacebar whenever the sound was the letter d. This was a simultaneous secondary task in order to measure the cognitive load of the participants while interacting with the machine through the different conditions. The sound effects appeared two times: in the beginning of the first quiz only once as a reminder note and within the presentation of the second confidence scale. The point of the presentation of the aural stimuli could be applied during the process of typing their explanations for both second and third condition (our initial plan), nevertheless due to some reasons, it was applied during the rating of their confidence level. The experimental reason for this choice was that we wanted to see if the intellectual process or the different types of interaction of each of the three different conditions affected with different way their mental alertness. For example, if their mind was clearer or more organized depending on the condition (because x type of interaction helped them more), their alertness during rating their confidence level would be better.

Levels of specialization were divided to low 684, medium 855, and high 333 (subjects 76, 95 and 37 respectively, with unclassified data 50 = 258). The missing values were acquired due to lack of responses in the demographic section by some participants. The above mentioned disjunction was aimed to distinguish the sample to people that are more familiar with probabilistic thinking, due to their field of studies, and to those that are less familiar. Such separation is in alignment with Chase and Simon (1973), Gobet and Clarkson (2004), Charness (1981) and deliberate practice theory of Ericsson et al., (1993) as referred to Eysenck (2006). For example, people with degrees on mathematics, business administration and economics, as they also reported on their own, were already familiar with similar or identical quizzes. Hence, people with such degrees were assigned to high level of specialization, people from other STEM field who are in general familiar with mathematics were assigned to the medium level of specialization while people from theoretical sciences were assigned to low level of specialization.

Design and procedure

In the research was utilized 9x3 factorial design with manipulation in the second IV and thus, the first experiment was manipulated as a single factor design with nine levels [1.Definition (Rule based), 2.Contradictive definitions (R+ Far miss), 3.Similar definitions (R+ Near miss), 4.Examples (Example based), 5.Contrastive examples (E+ Far Miss), 6.Similar examples (E+ Near miss), 7.Prototypes (Prototype based), 8.Contradictive Prototypes (P+ Far miss), 9. Similar Prototypes (P+ Near miss)]. This condition is labeled as a unilateral interaction.

The experiment was conducted online with the PsychoPy survey software using Vertical Enhancement of Statistics and Psychology Research software tool for assigning randomly each participant to each of the three conditions. The whole process lasted for about 20-25 minutes for each participant. At the beginning, participants were receiving written instructions in their screen. When participants were ready, they could press the spacebar in their keyboard and, one sequence of nine digits were displayed consecutively in their screen. Each digit of the sequence was followed by the next every one second. After that, participants had fifteen seconds to recall and write the sequence of digits in their screen via their keyboard. This process was repeated 3 times. In the next stage, nine quizzes were displayed in the screen one by one at a random fashion for each participant. Every time a quiz appeared in the screen, the participant had to choose a predefined answer and the percentage of their confidence level for this answer. At this stage, machine had to provide the right answer to the participant by offering an argument/explanation as well. The types of explanations in this research were nine and thus, each quiz had nine different types of explanations attached to it that were presented to each participant randomly. Each participant was given all the versions of explanations by receiving a different type for each quiz. Then the participant was provided with the explanation of the machine, he/she was asked if he/she wants to change his/her answer and choose again his/her confidence level for his previous answer. Within this last procedure, sound effects such as the sound of the letter b and the sound of the letter d were perceptible, while participants were scaling their confidence level. The participants had to tap the spacebar whenever the sound was the letter d. Both the reaction time and the correct/false responses were recorded and calculated. It has to be noted that guideline notes were presented between the tasks in order for the participants to stay tuned.

Results

Between Incorrect and Correct answers was found a significant difference in the mean change (Δ) of the Confidence Rate in the 4th type of explanation (Ancova, F=6.14, d.f.=2301; -0.16 vs. 0.17, respectively, p = 0.017) or in general in all the mean changes regardless of the type of explanation (Ancova, F=5.67, d.f.=2301; -0.07 vs. 0.04, p = 0.013) (Table 3; Figure 7). This means that the Example based type of explanation decreased significantly the confidence level of the participants for their responses when they had provided incorrect answers and increased in a significant level their confidence, when they had provided correct answers.

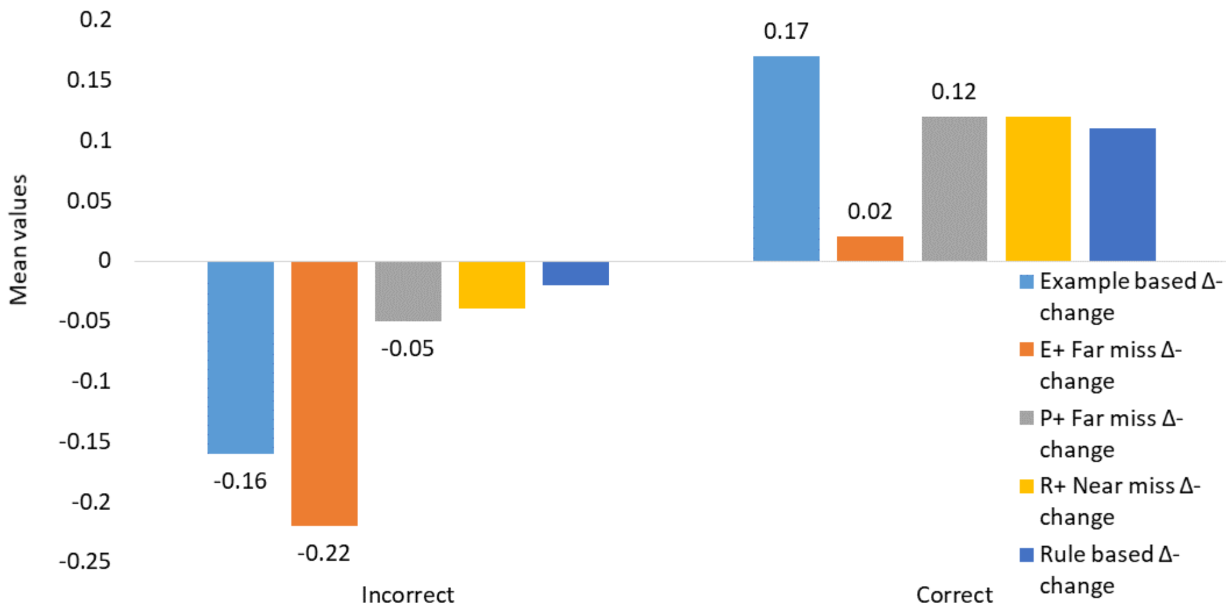
As we can see in Table 3, in general, machine's explanations affected the confidence level of participants for their answers. Even though only the example based explanation has statistical significance, it is quite interesting to see the priority sequence of the types that affected the confidence of participants on correct and incorrect answers: Example based > E+ Far miss > P+ Far miss > R+ Near miss > Rule based.

Table 3. Comparisons of changes in self-confidence levels between types of explanation and answers of 258 participants (2,322 measurements).

Type Of Explanation	Confidence Rate	Incorrect		Correct		p-value
		Mean	Std. Error	Mean	Std. Error	
Rule based	Before	3.67	0.08	3.83	0.12	
	After	3.65	0.09	3.94	0.13	
	Δ-change	-0.02		0.11		0.339
R+ Far miss	Before	3.47	0.08	3.72	0.12	
	After	3.47	0.09	3.76	0.13	
	Δ-change	0.01		0.04		0.825
R+ Near miss	Before	3.61	0.09	3.43	0.11	
	After	3.58	0.09	3.55	0.12	
	Δ-change	-0.04		0.12		0.220
Example based	Before	3.61	0.08	3.62	0.12	
	After	3.46	0.09	3.79	0.13	
	Δ-change	-0.16		0.17		**0.017**
E+ Far miss	Before	3.62	0.09	3.69	0.11	
	After	3.39	0.09	3.71	0.12	
	Δ-change	-0.22		0.02		0.066
E+ Near miss	Before	3.65	0.09	3.56	0.11	
	After	3.54	0.09	3.50	0.12	
	Δ-change	-0.11		-0.05		0.644
Prototype based	Before	3.79	0.09	3.75	0.11	
	After	3.70	0.09	3.66	0.12	
	Δ-change	-0.09		-0.10		0.939
P+ Far miss	Before	3.59	0.08	3.67	0.12	
	After	3.54	0.09	3.78	0.13	
	Δ-change	-0.05		0.12		0.211
P+ Near miss	Before	3.55	0.08	3.71	0.12	
	After	3.64	0.09	3.69	0.13	
	Δ-change	0.09		-0.03		0.413
Total	Before	3.62	0.03	3.67	0.04	
	After	3.55	0.03	3.71	0.04	
	Δ-change	-0.07		0.04		**0.013**

Repeated measures analysis (ancova). As covariates were used gender and age.

Figure 7. The effect of the types of explanation on confidence level of the total sample.



Repeated measures analysis (ancova), p=0.017.

On table 4 and Figure 8-12, we can see that there was a statistically significant difference in consistency[1] percentages in correct answers on 7th type of explanation (Chi-square test, Pearson $\chi^2 = 4.337$, d.f. = 1; 38.4% *vs.* 12.5%, p = 0.037). This means that the Prototype based explanation mostly affected the responses of participants, that is, the change of their answer, being incorrect, based on the feedback of the machine. Despite the fact that only the effect of Prototype based type was statistically significant, we should ranking the types that made participants change their incorrect answers: Prototype based > Rule based > R+ Near miss > P+ Far miss.

Table 4. Frequencies of answers between types of explanation and consistency of 258 participants (2,322 measurements).

Type Of Explanation	Consistency	**Incorrect**		**Correct**		p-value
		n	%	n	%	
Rule based	no	154	63.9	87	36.1	0.123
	yes	14	82.4	3	17.6	
R+ Far miss	no	169	67.6	81	32.4	0.075
	yes	3	37.5	5	62.5	
R+ Near miss	no	150	61.5	94	38.5	0.199
	yes	11	78.6	3	21.4	
Example based	no	166	67.5	80	32.5	0.953
	yes	8	66.7	4	33.3	
E+ Far miss	no	154	63.6	88	36.4	0.957
	yes	10	62.5	6	37.5	
E + Near miss	no	153	63.2	89	36.8	0.120
	yes	7	43.8	9	56.3	
Prototype based	no	149	61.6	93	38.4	**0.037**
	yes	14	87.5	2	12.5	
P+ Far miss	no	165	66.8	82	33.2	0.682
	yes	8	72.7	3	27.3	
P+ Near miss	no	167	68.2	78	31.8	0.477
	yes	7	58.3	5	41.7	
Total	no	1427	64.9	772	35.1	0.601
	yes	82	67.2	40	32.8	

Chi-square tests (χ^2)

[1] Consistency: no = they did not change their answer, yes= they changed their answer

Figure 8. The effect of the types of explanation on persuasiveness of the total sample.

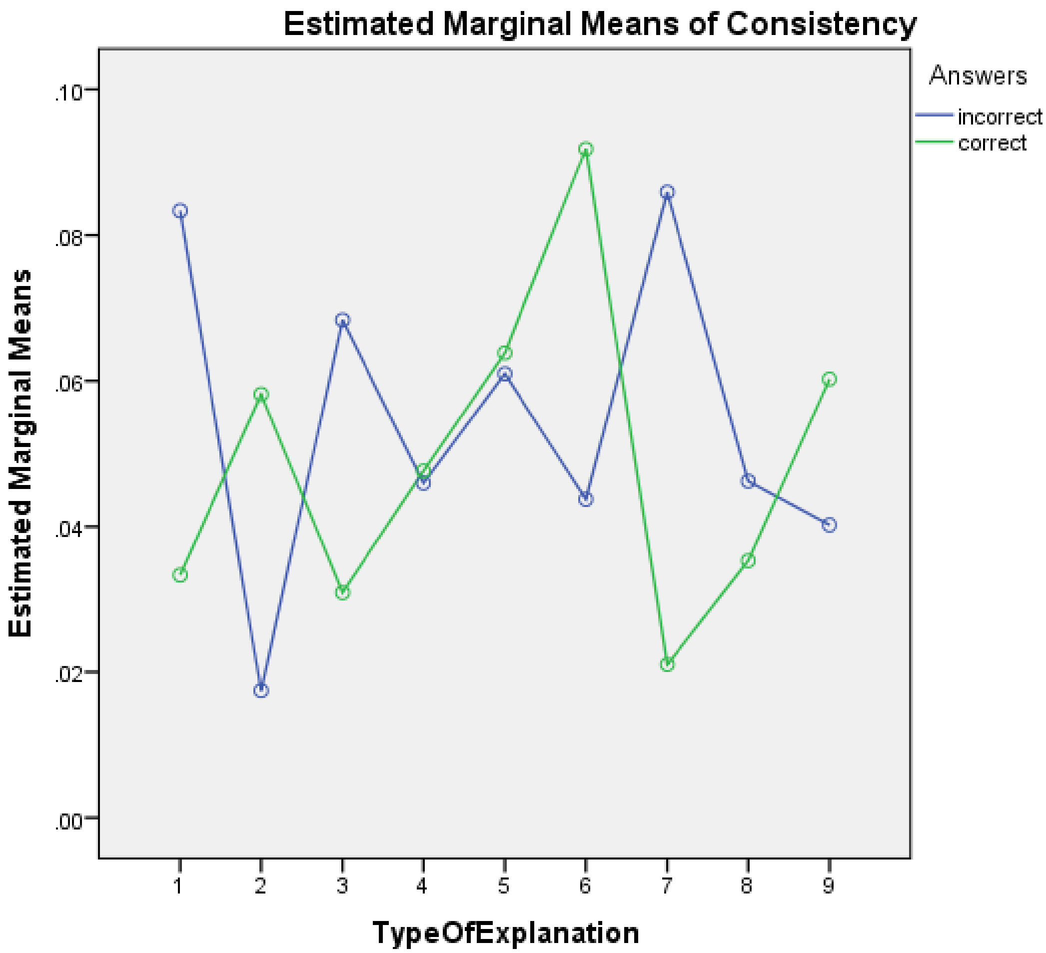
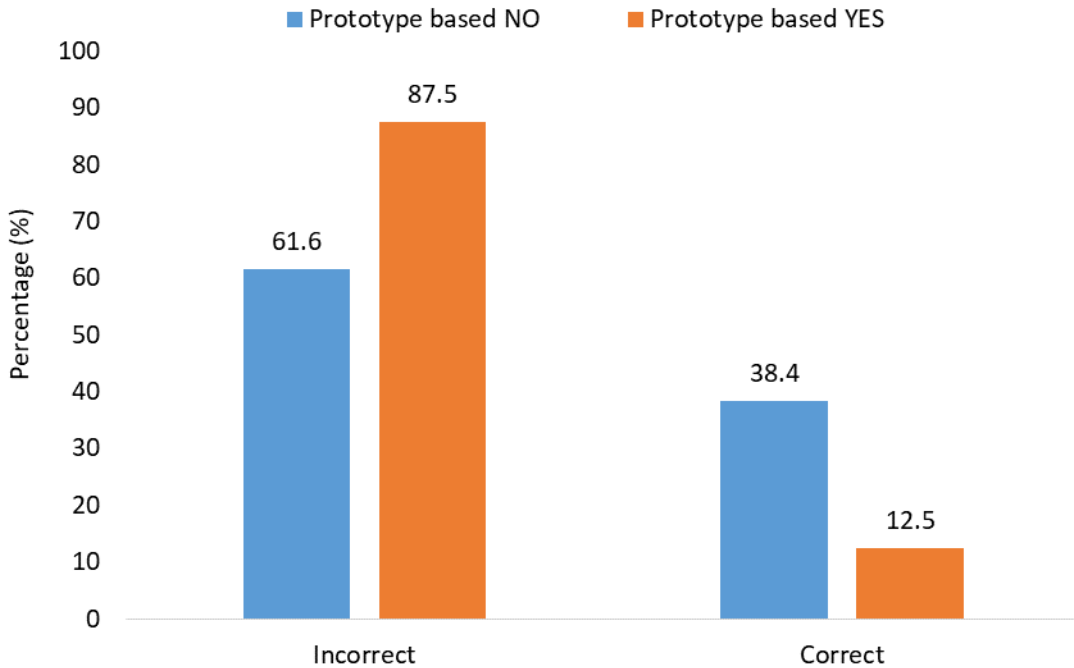
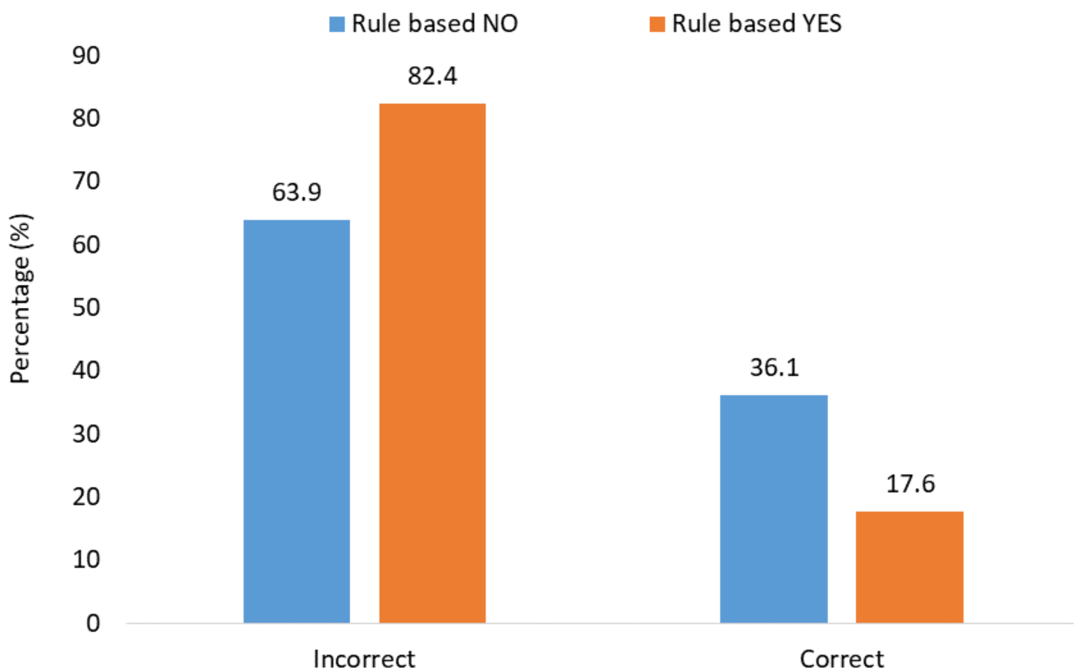


Figure 9. The effect of the Prototype based type on consistency.



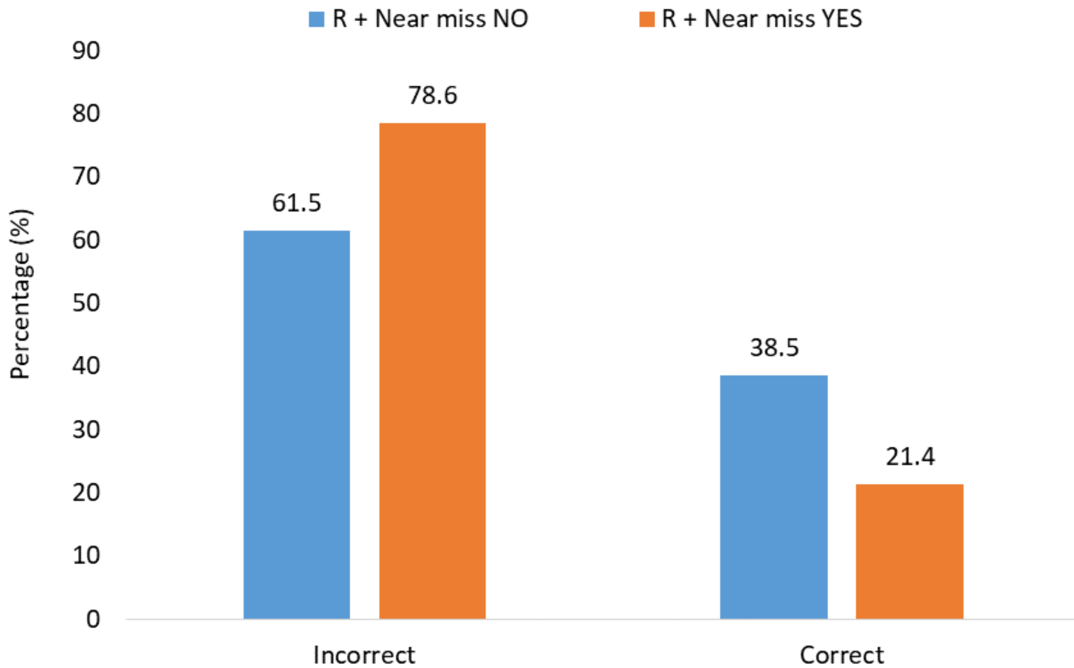
Chi-square test, p=0.037

Figure 10. The effect of the Rule based type on consistency.



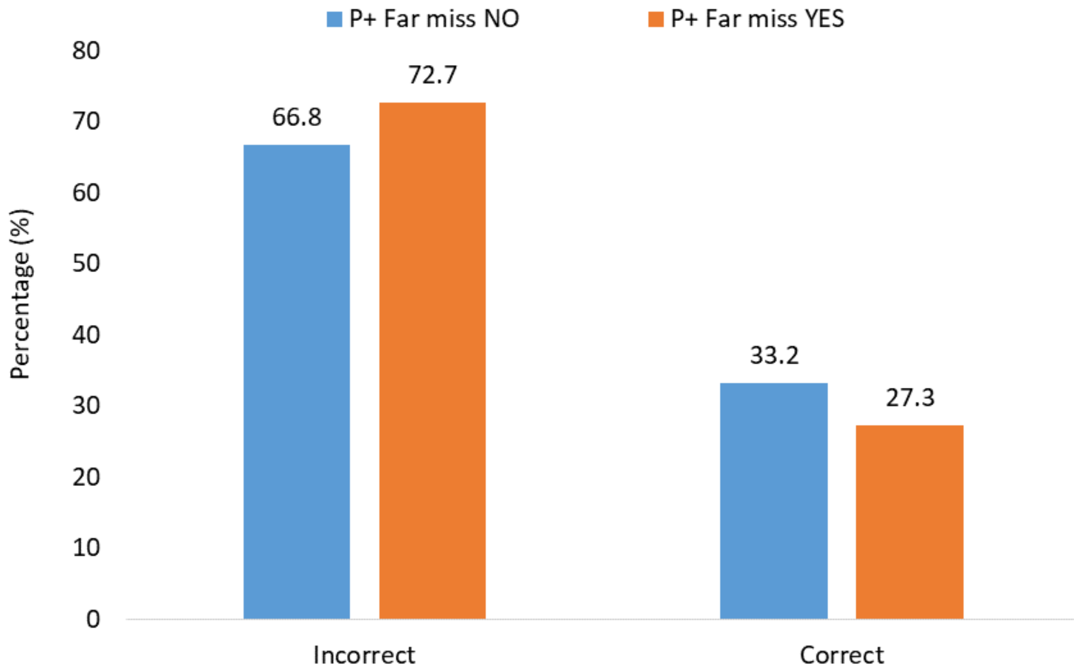
Chi-square test,p=0.123.

Figure 11. The effect of the Rule+ Near miss type on consistency.



Chi-square test,p=0.199.

Figure 12. The effect of the Prototype+ Far miss type on consistency.



Chi-square test,p=0.682.

Between Incorrect and Correct answers was found a significant difference in the mean change (D) of the Confidence Rate in the 4th type of explanation for individuals with high level of expertise (Ancova, F=4.64, d.f.=295; -0.38 vs. 0.46, respectively, p = 0.027). This means that the Example based type of explanation, decreased significantly the confidence level of participants, with high level of expertise, for their responses when they had provided incorrect answers and increased in a significant level their confidence, when they had provided correct answers (Table 5; Figure 13-15).

Regardless the results, it is interesting to rank again the effect of the different types of explanation on self-confidence throughout the different levels of expertise/specialization[2]:

<u>High level</u>

Example based > R+ Near miss > Rule based > Prototype based > E+ Near miss > R+ Far miss.

<u>Medium level</u>

Example based > P+ Far miss.

<u>Low level</u>

E+ Far miss > Example based > E+ Near miss > R+ Near miss > R+ Far miss.

[2] Specialization: high= economics/business administration/statistics/mathematics, medium= other STEM fields/exact sciences, low= theoretical sciences

Table 5. Comparisons of mean changes of self-confidence levels between types of explanation, specialization and answers of 258 participants (2,322 measurements).

Type Of Explanation	Specialization	Incorrect	Correct	p-value
		Δ mean-change		
Rule based	low	0.01	0.19	0.486
	medium	0.08	-0.04	0.591
	high	-0.05	0.41	0.216
R+ Far miss	low	-0.06	0.06	0.646
	medium	0.10	0.08	0.936
	high	-0.23	0.08	0.401
R+ Near miss	low	-0.06	0.13	0.439
	medium	0.04	0.08	0.870
	high	-0.23	0.33	0.110
Example based	low	-0.19	0.04	0.371
	medium	-0.07	0.27	0.127
	high	-0.38	0.46	**0.027**
E+ Far miss	low	-0.23	0.08	0.208
	medium	-0.29	-0.09	0.358
	high	-0.26	-0.10	0.651
E+ Near miss	low	-0.10	0.11	0.383
	medium	-0.10	-0.18	0.735
	high	-0.09	0.31	0.273
Prototype based	low	-0.11	-0.06	0.843
	medium	0.00	-0.29	0.192
	high	-0.18	0.24	0.236
P+ Far miss	low	-0.07	-0.07	0.998
	medium	-0.13	0.16	0.204
	high	0.09	0.07	0.967
P+ Near miss	low	0.00	0.17	0.514
	medium	0.09	-0.28	0.111
	high	0.08	0.08	0.999

Repeated measures analysis (ancova). As covariates were used gender and age.

Figure 13. The effect of the types of explanation on confidence level of people with high level of specialization.

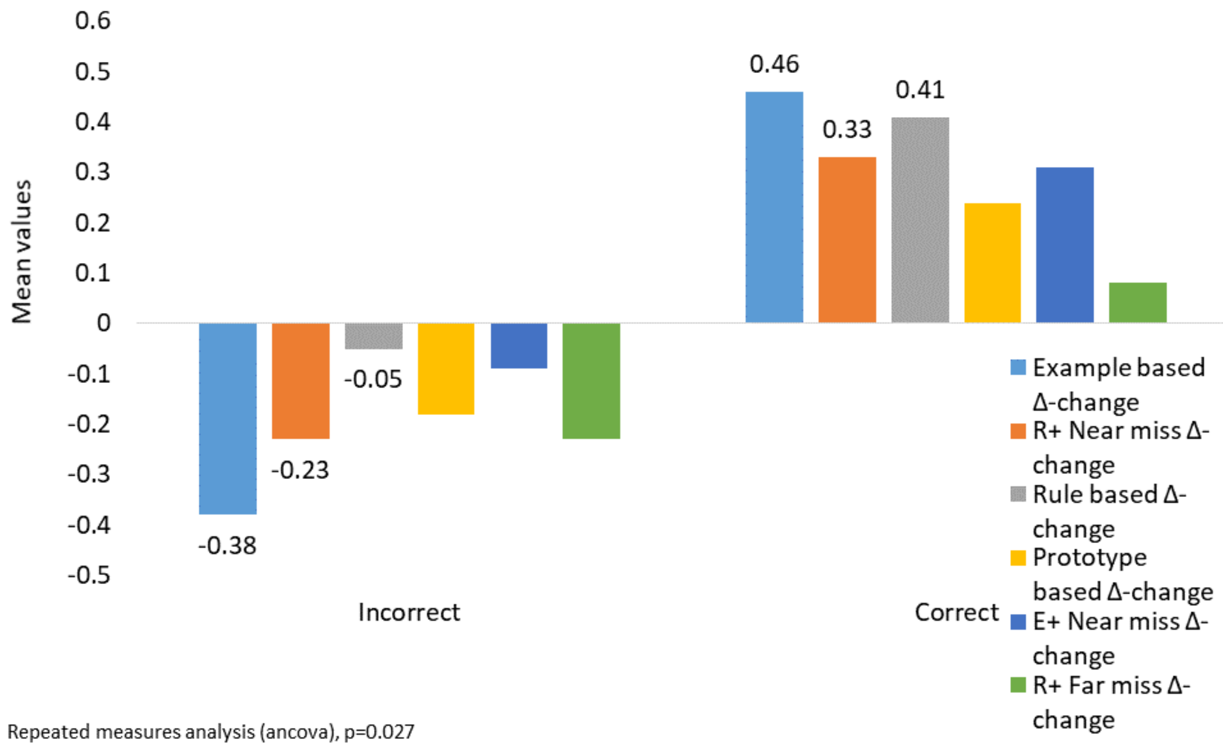
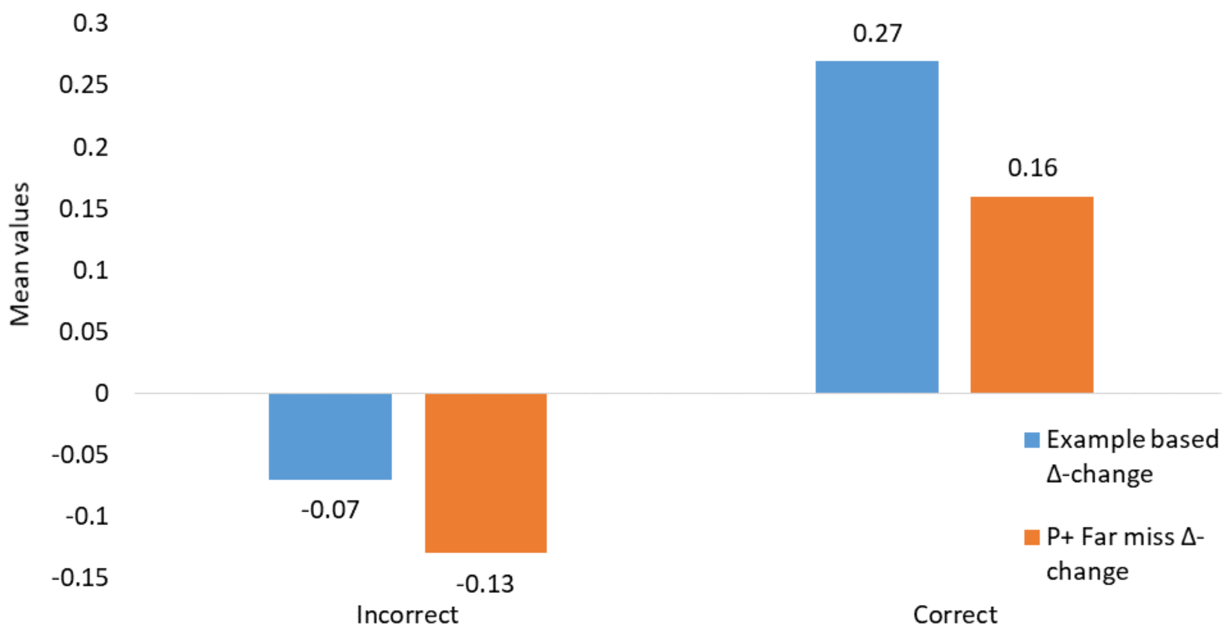
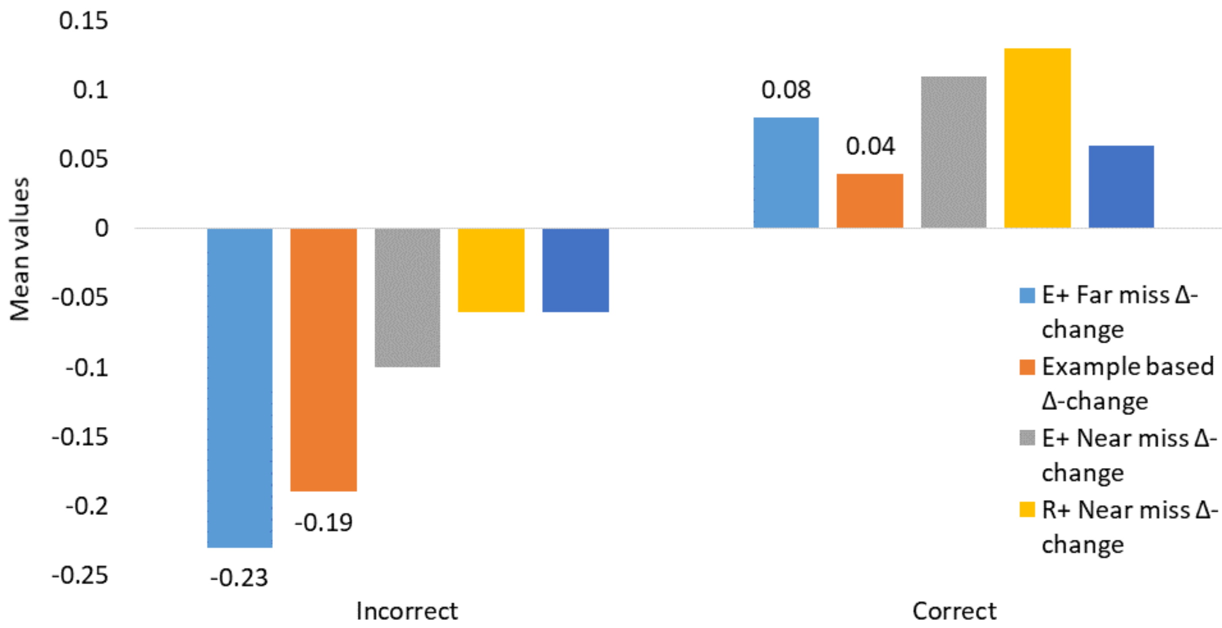


Figure 14. The effect of the types of explanation on confidence level of people with medium level of specialization.



Repeated measures analysis (ancova), $p > 0.05$.

Figure 15. The effect of the types of explanation on confidence level of people with low level of specialization.



Repeated measures analysis (ancova), $p > 0.05$.

2.2 Second Experiment

Method

Participants

N=89 people (59 Male; 30 Female) ranging in age from 18 to 68 (M = 35.9, SD = 13.5) participated in this experiment. Participants have been selected randomly. They either have been volunteered or been asked to participate in the experiment. The only requirement was the normal or corrected to normal vision and having access to the audio sound of their devices.

Materials

They were used the same tools as the experiment 1.

Design and procedure

In the research was utilized 9x3 factorial design with manipulation in the second IV and thus, this second experiment was manipulated as a single factor design with nine levels [1.Definition (Rule based), 2.Contradictive definitions (R+ Far miss), 3.Similar definitions (R+ Near miss), 4.Examples (Example based), 5.Contrastive examples (E+ Far Miss), 6.Similar examples (E+ Near miss), 7.Prototypes (Prototype based), 8.Contradictive Prototypes (P+ Far miss), 9. Similar Prototypes (P+ Near miss)]. This condition is labeled as a bilateral interaction with human initiation.

The experiment was conducted online with the PsychoPy survey software using Vertical Enhancement of Statistics and Psychology Research software tool for assigning randomly each participant to each of the three conditions. The whole process lasted for about 20-25 minutes for each participant. Initially, participants received instructions in writing, in their screen. When participants were ready, one sequence of nine digits were displayed consecutively in their screen. Each digit of the sequence was followed by the next every 1 second. After that, participants had 15 seconds to recall and write the sequence of digits in their screen via their keyboard. This process was repeated 3 times. In the next stage, nine quizzes were displayed in the screen one by one at a random manner to each of the participants. Every time a quiz was displayed in the screen, the participant had to choose a predefined answer, give an explanation for this answer and choose the percentage of their confidence level for the same answer. Later, the machine had to provide the right answer to the participant by offering at the same time an argument/explanation. The types of explanations in this research were nine and thus, each quiz had nine different types of explanations attached to it that were presented to each participant randomly. Each participant was given all the versions of explanations by receiving a different type for each quiz. Then, the machine had to show the explanation that the participant gave for this answer of the quiz and ask the participant if he/she would like to change the answer by offering again the option of scaling the confidence level. Within this last procedure, sound effects such as the sound of the letter b and the sound of the letter d were perceptible, while participants were scaling their confidence level. The participants had to tap the spacebar whenever the sound was the letter d. Both the reaction time and the correct/false responses were recorded and calculated. It has to be noted that guideline notes were presented between the tasks in order for the participants to stay tuned.

Results

No significant differences were found on mean reaction time in 1^{st}, 2^{nd} and 3^{rd} conditions for tone D, B, percentage of false responses and percentage of correct responses (Table 6; Figure 16, 17). This means that the performance of cognitive load did not significantly differ between the three different conditions.

Nevertheless, we can see a slightly better mean performance on reaction time and success rating on the condition 3 where participants had to react to an explanation, compared to the condition 2 where they had to initiate an explanation or the control condition, that is, condition 1.

Table 6. Comparisons of reaction time and correct/false responses between conditions in 258 participants (2,322 measurements).

	Total		Condition						
			1st		2nd		3rd		
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	p-value
Number	2321		774		801		747		
ToneSpeedD	0.28	0.44	0.26	0.42	0.27	0.43	0.31	0.47	0.248
ToneSpeedB	0.09	0.28	0.09	0.28	0.08	0.25	0.10	0.30	0.893
ToneFailRate (%)	6.6	21.3	6.6	21.2	6.4	21.1	7.0	21.8	0.943
ToneSuccessRate (%)	26.7	40.4	26.1	40.1	25.8	39.7	28.5	41.5	0.408

Kruskal-Wallis tests

Figure 16. The effect of condition on cognitive load, that is, reaction time for both tones.

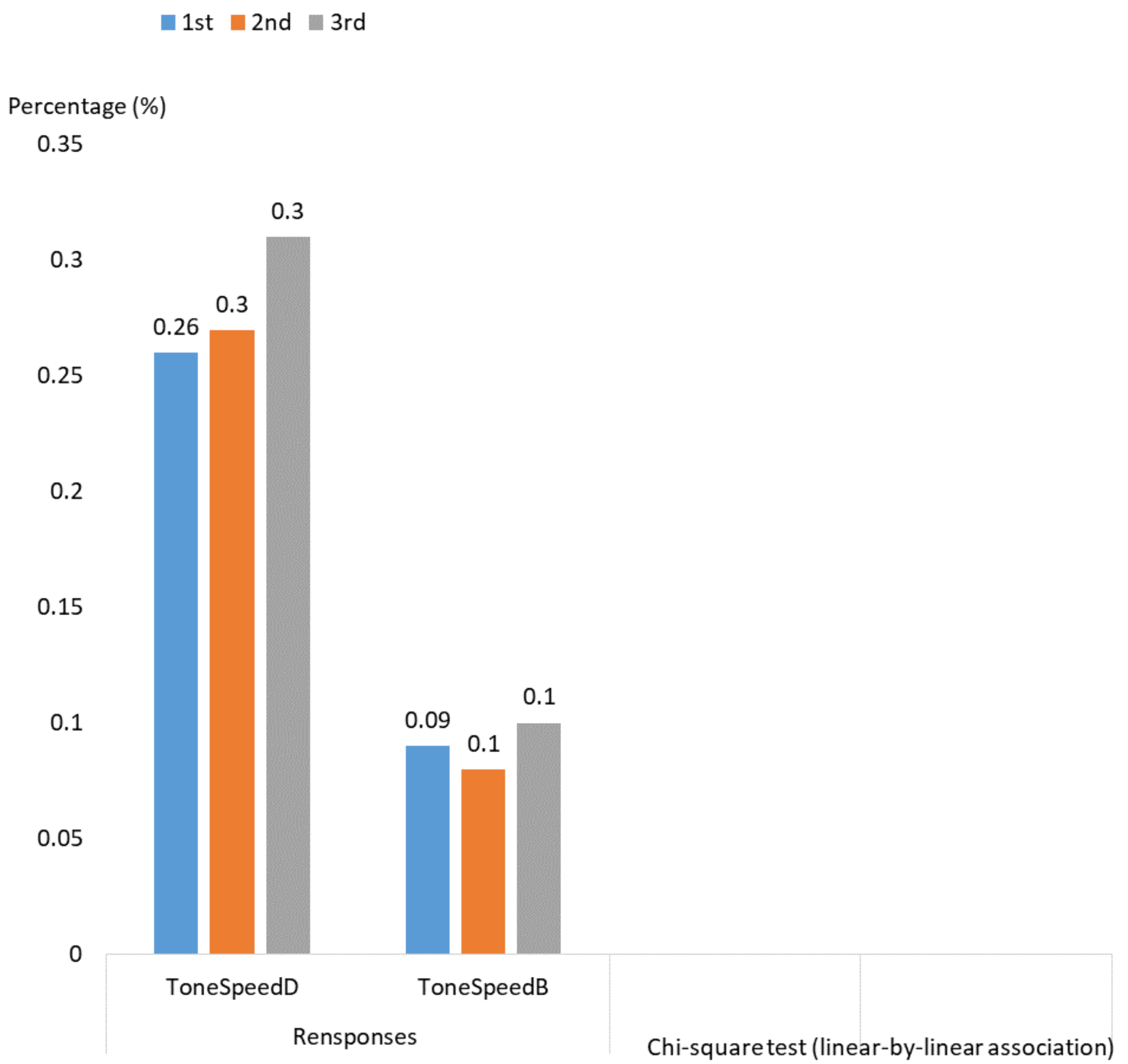
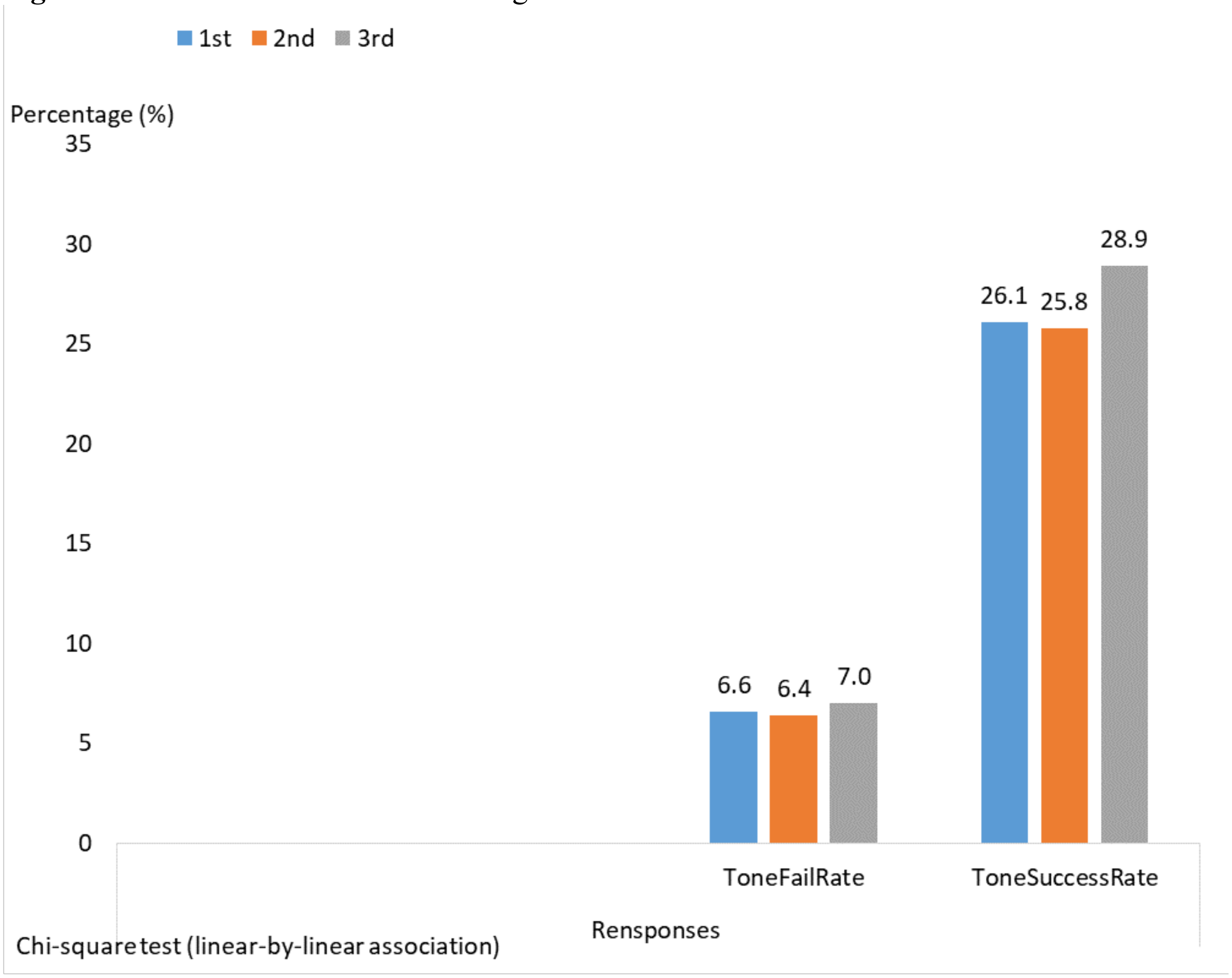


Figure 17. The effect of condition on cognitive load for both success and failure rate.



On table 7 and Figure 18-19, we can see a significant difference between the levels of working memory capacity on both consistency behavior (Chi-square test, Pearson $\chi^2 = 13.45$, df = 2; 10.3%, 4.3% and 4.9%, $p = 0.028$) and correct performance (Chi-square test, Pearson $\chi^2 = 9.83$, df = 2; 25.8%, 35.6% and 36.3%, $p = 0.013$). More specifically, the level of working memory capacity of participants affected their consistency behavior, with those who have higher capacity being more consistent compared to those with lower short term memory capacity. In the same manner, participants with higher working memory capacity had better performance in the quizzes in comparison to those with lower capacity.

Table 7. Frequencies of different levels of working memory between levels of consistency and answers in 258 participants (2,322 measurements).

			Digit span			p-value
			low	**medium**	**high**	
Consistency	*no*	*n*	210	775	1215	**0.028**
		%	89.7	95.7	95.1	
	yes	*n*	24	35	63	
		%	10.3	4.3	4.9	
Answers	*Incorrect*	*n*	173	522	814	**0.013**
		%	74.2	64.4	63.7	
	Correct	*n*	60	288	464	
		%	25.8	35.6	36.3	

Chi-square tests (χ^2) (linear-by-linear association)

Figure 18. The effect of working memory capacity on consistency.

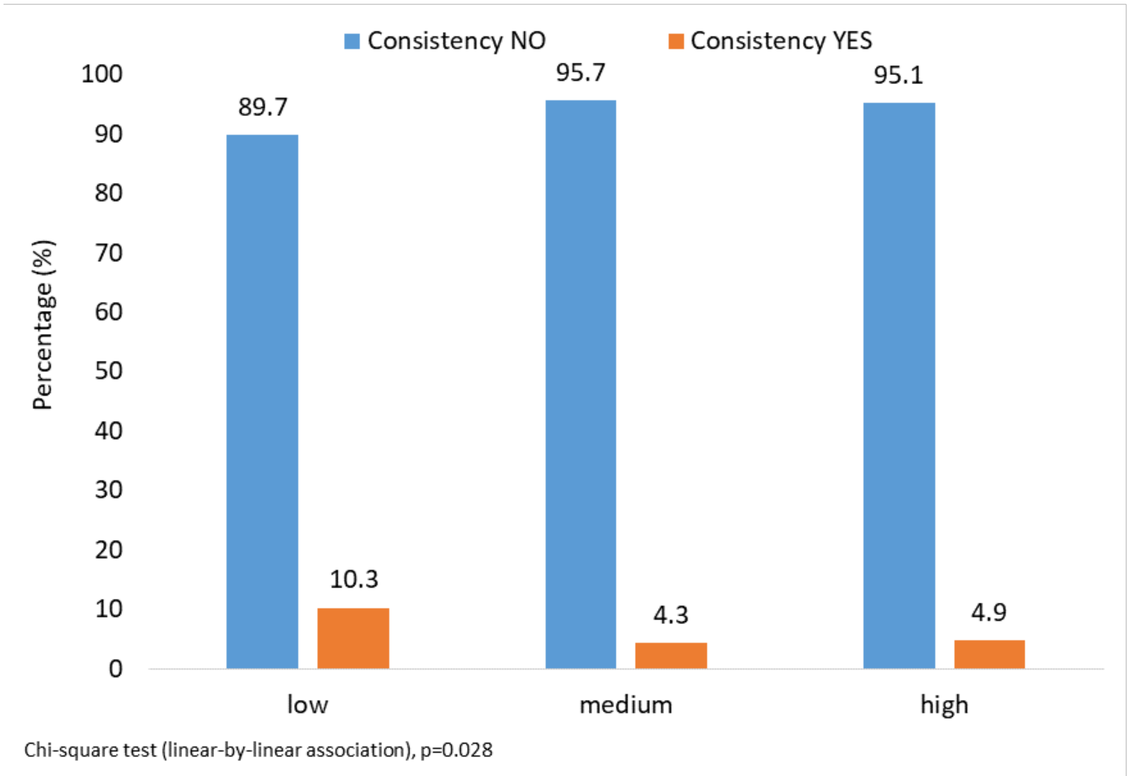
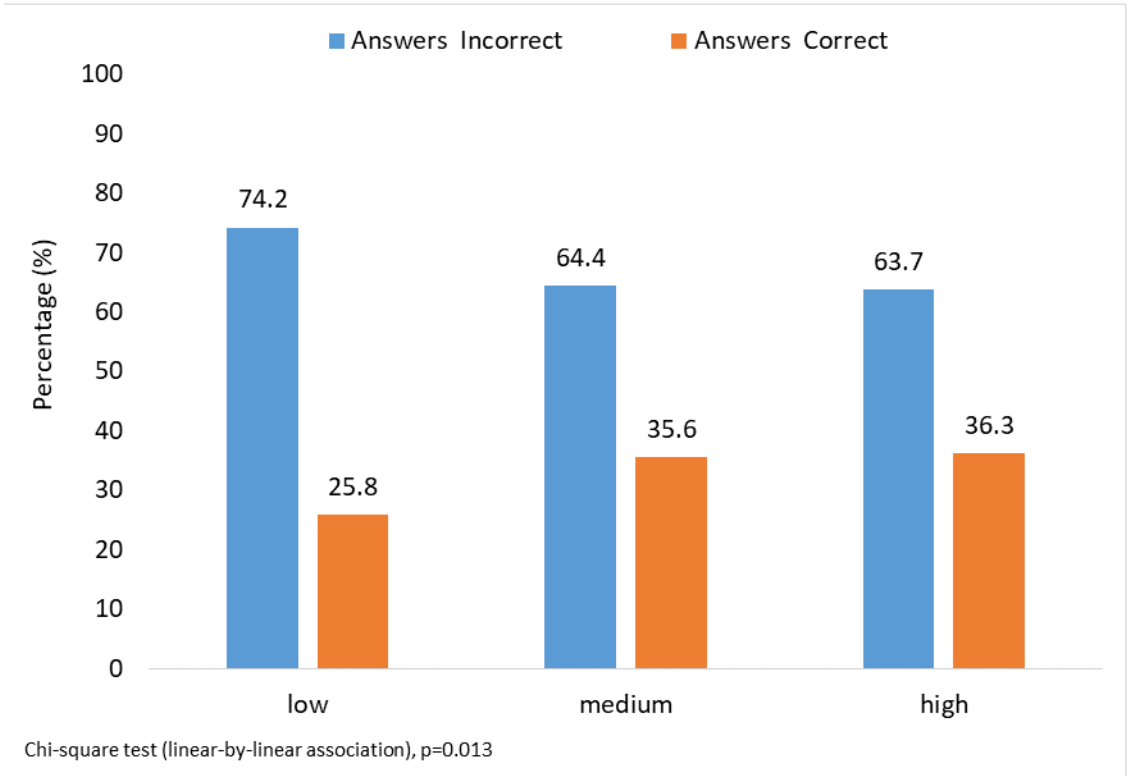


Figure 19. The effect of working memory capacity on performance.



Alternatively, as we can see on table 8 and Figure 20-21, there is no significant difference between the levels of specialization on both consistency behavior and performance ($p > 0.05$). With other words, the performance and behavioral consistency of participants did not affected by the level of their expertise, that is, if their field was related to the quizzes or not.

Table 8. Frequencies of different levels of specialization between levels of consistency and answers in 258 participants (2,322 measurements).

			Specialization			p-value
			low	**medium**	**high**	
Consistency	*no*	*n*	647	823	305	0.158
		%	36.5	46.4	17.2	
	yes	*n*	37	32	10	
		%	46.8	40.5	12.7	
Answers	*Incorrect*	*n*	445	558	194	0.468
		%	37.2	46.6	16.2	
	Correct	*n*	239	296	121	
		%	36.4	45.1	18.4	

Chi-square tests (χ^2) (linear-by-linear association)

Figure 20. The effect of specialization on consistency.

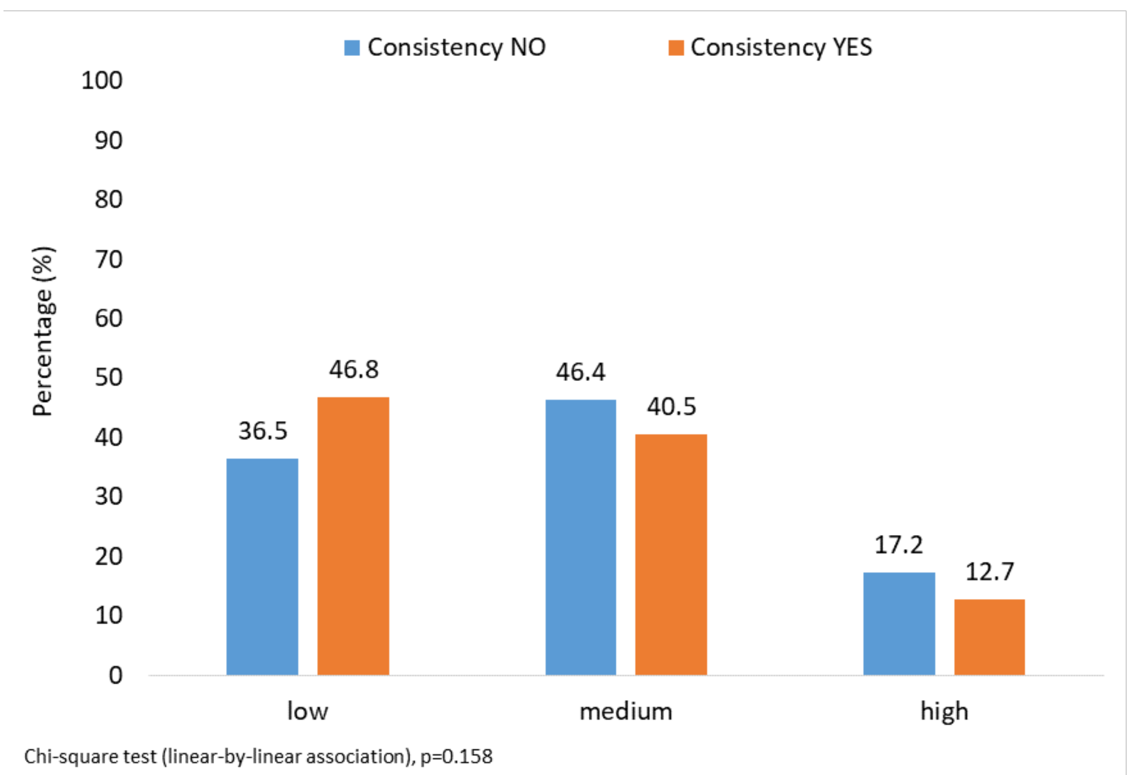
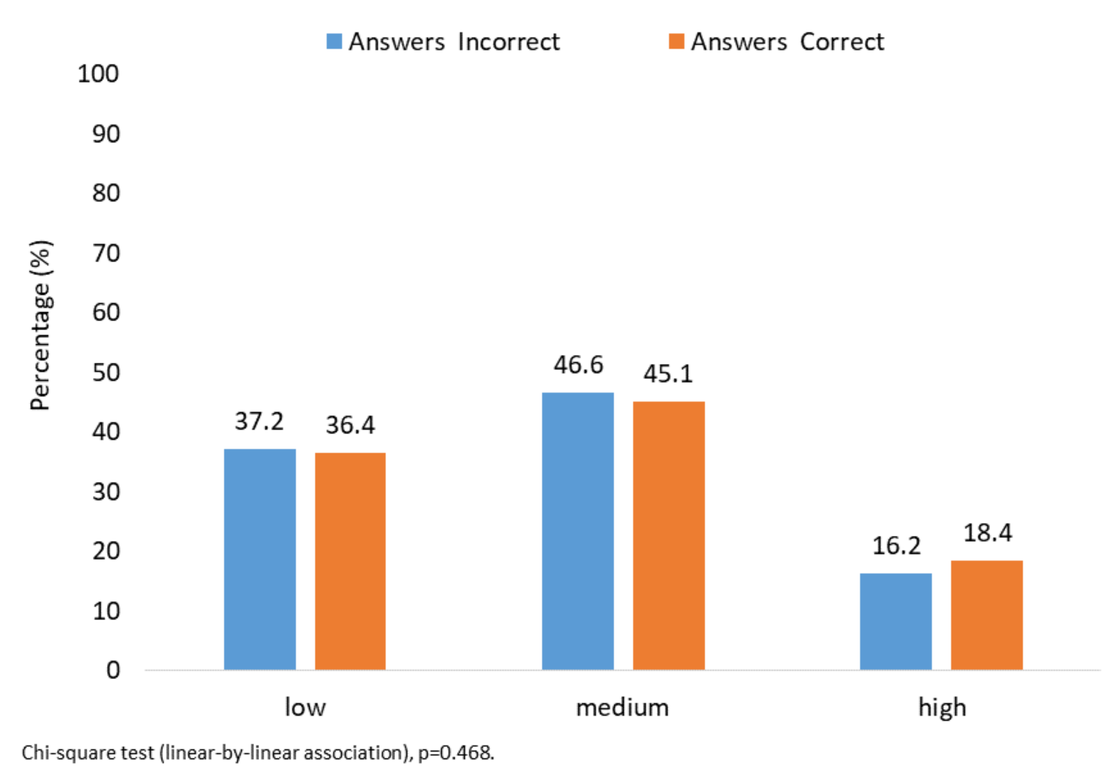


Figure 21. The effect of specialization on the performance.



2.3 Third Experiment

Method

Participants

N=83 people (52 Male; 31 Female) ranging in age from 18 to 68 (M = 35.9, SD = 13.5) participated in this experiment. Participants have been selected randomly. They either have been volunteered or been asked to participate in the experiment. The only requirement was the normal or corrected to normal vision and having access to the audio sound of their devices.

Materials

They were used the same tools as the experiment 1.

Design and procedure

In the research was utilized 9x3 factorial design with manipulation in the second IV and thus, this third experiment was manipulated as a single factor design with nine levels [1.Definition (Rule based), 2.Contradictive definitions (R+ Far miss), 3.Similar definitions (R+ Near miss), 4.Examples (Example based), 5.Contrastive examples (E+ Far Miss), 6.Similar examples (E+ Near miss), 7.Prototypes (Prototype based), 8.Contradictive Prototypes (P+ Far miss), 9. Similar Prototypes (P+ Near miss)]. This condition is labeled as a bilateral interaction with human reaction.

The experiment was conducted online with the PsychoPy survey software using Vertical Enhancement of Statistics and Psychology Research software tool for assigning randomly each participant to each of the three conditions. The whole process lasted for about 20-25 minutes for each participant. At first, participants were provided with written instructions in their screen. When participants were ready, one sequence of nine digits were displayed consecutively in their screen. Each digit of the sequence was followed by the next every 1 second. This process was repeated 3 times .After that, participants had 15 seconds to recall and write the sequence of digits in their screen via their keyboard. In the next stage, nine quizzes were displayed in the screen one by one at a random manner to each of the participants. Every time a quiz was displayed in the screen, the participant had to choose a predefined answer and the percentage of their confidence level for this answer. Then, the machine had to provide the right answer to the participant by offering an argument/explanation as well. The types of explanations in this research were nine and thus, each quiz had nine different types of explanations attached to it that were presented to each participant randomly. Each participant was given all the versions of explanations by receiving a different type for each quiz. After the participant was provided with the explanation of the machine, he/she had to provide an explanation of their previous choice. Then, the machine had to provide the option to the participant to change his/her answer and ask him/her to rate again his/her confidence level for his/her previous answer. Within this last procedure, sound effects such as the sound of the letter b and the sound of the letter d were perceptible, while participants were scaling their confidence level. The participants had to tap the spacebar whenever the sound was the letter d. Both the reaction time and the correct/false responses were recorded and calculated. It has to be noted that guideline notes were presented between the tasks in order for the participants to stay tuned.

Results

Between the three different conditions was found a significant difference in the performance of the high level of expertise category (Chi-square test, Pearson $\chi^2 = 7.64$, df = 2; 28.9%, 48.9% and 37.8%, $p = 0.022$) (Table 9; Figure 22, 23). Specifically, people whose their field is strongly related to the context of the quizzes, increased significantly their performance in both the second and third condition. In general people seem to increased their correct answers and decreased their incorrect answers in these last conditions and especially in the second condition.

Again, between the three different conditions was found a significant difference in the consistency behavior regardless of the performance, that is, being incorrect (Chi-square test, Pearson $\chi^2 = 268.3$, df = 2; $p < 0.001$), correct (Chi-square test, Pearson $\chi^2 = 91.4$, df = 2; $p < 0.001$) or in total (Chi-square test, Pearson $\chi^2 = 246.5$, df = 2; $p < 0.001$) (Table 10; Figure 24). Particularly, the second and fourth condition equated the consistency fluctuations of participants to zero.

Table 9. Frequencies of answers between levels of specialization and conditions in 258 participants (2,322 measurements).

Specialization	Answers		Condition			p-value
			1st	**2nd**	**3rd**	
Low	*Incorrect*	*n*	151	148	146	0.614
		%	64.5	63.2	67.6	
	Correct	*n*	83	86	70	
		%	35.5	36.8	32.4	
Medium	*Incorrect*	*n*	153	202	203	0.319
		%	68.3	62.3	66.3	
	Correct	*n*	71	122	103	
		%	31.7	37.7	33.7	
High	*Incorrect*	*n*	64	46	84	**0.022**
		%	71.1	51.1	62.2	
	Correct	*n*	26	44	51	
		%	28.9	48.9	37.8	
Total	*Incorrect*	*n*	368	396	433	0.064
		%	67.2	61.1	65.9	
	Correct	*n*	180	252	224	
		%	32.8	38.9	34.1	

Chi-square tests (χ^2)

Figure 22. The effect of condition on the performance for the total sample.

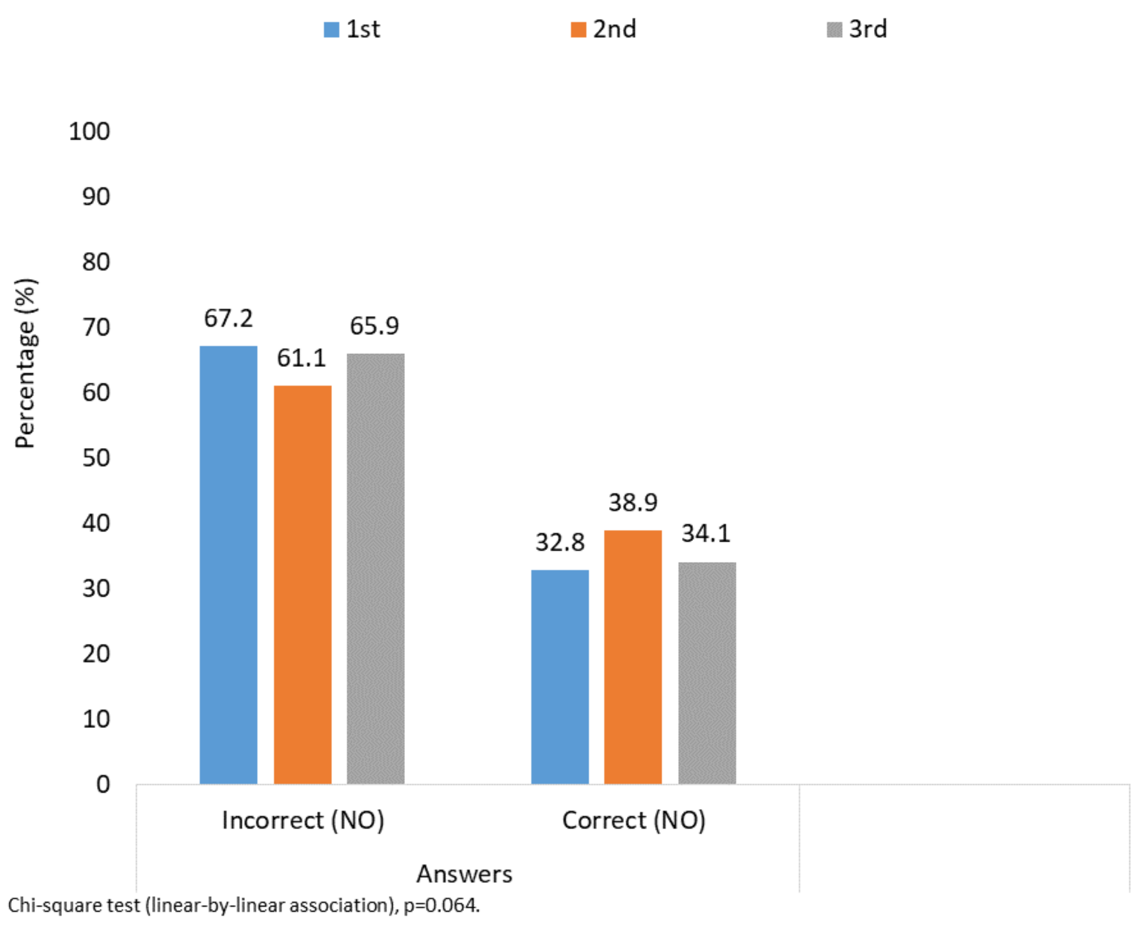


Figure 23. The effect of condition on the performance of people with high level of specialization.

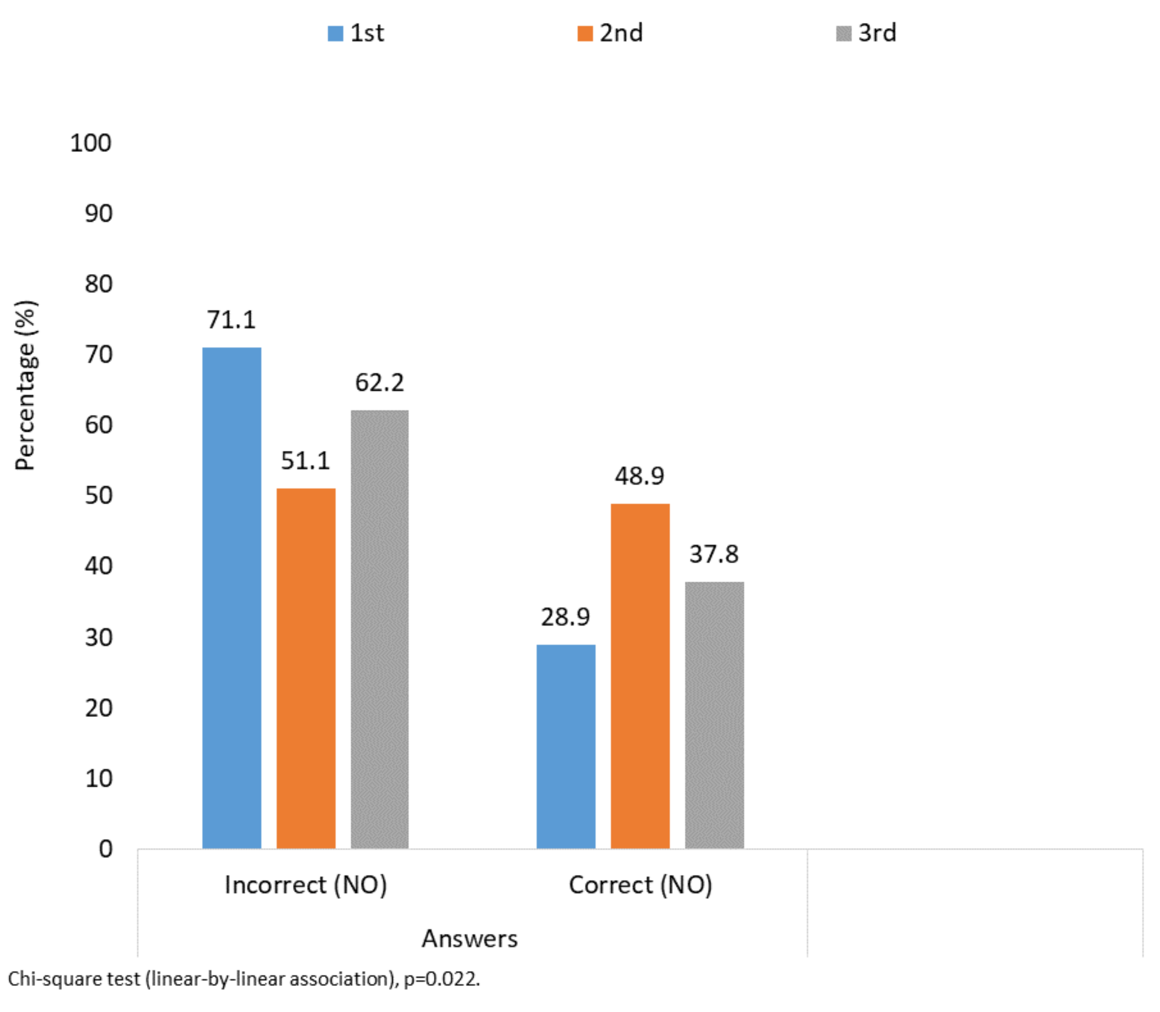
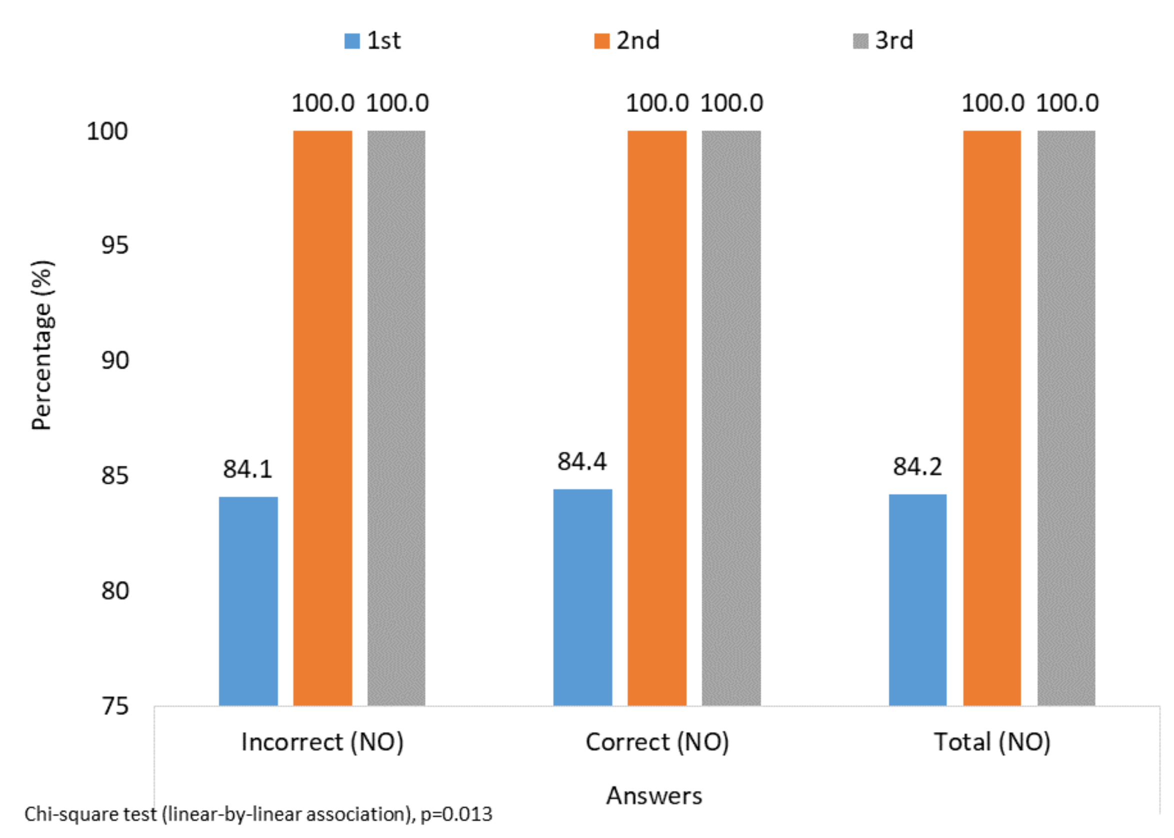


Table 10. Frequencies of answers between levels of consistency and conditions in 258 participants (2,322 measurements).

Answers	Consistency		Condition			p-value
			1st	**2nd**	**3rd**	
Incorrect	*no*	*n*	435	499	493	**<0.001**
		%	84.1	100.0	100.0	
	yes	*n*	82	0	0	
		%	15.9			
Correct	*no*	*n*	216	302	254	**<0.001**
		%	84.4	100.0	100.0	
	yes	*n*	40	0	0	
		%	15.6			
Total	*no*	*n*	651	801	747	**<0.001**
		%	84.2	100.0	100.0	
	yes	*n*	122	0	0	
		%	15.8			

Chi-square tests (χ^2)

Figure 24. The effect of condition on the consistency.



Chapter 3
Conclusion and discussion

To summarize, our goals for this research were divided to three sections: the first was to investigate the effects of different types of explanation on trust, persuasiveness and comprehensibility, the second was to measure cognitive load in three different HMI, and the third was to explore the effect of argumentation on problem solving such as consistency and performance, in three different conditions. Let's start with the first experiment. Our first plan was to measure comprehensibility by using self-report techniques, for instance, we had integrated a scale where participants had to choose characteristics to describe the different types of machine's explanations. Nevertheless, we excluded this scale from the experiment because in combination with the other tasks, the process was very time-consuming for the participants. For this reason, it will be more accurate to clarify that the findings of the first experiment can be interpreted in terms of trust and persuasiveness and not of comprehensibility.

We have seen that providing explanations can boost acceptance and trust (Ribeiro, Singh & Guestrin, 2016; Teso & Kersting, 2019) and this is also verified by our results, as machine's explanations affected the confidence level of participants. In general, explanations reduced the confidence level of participants for their answers, if they had given incorrect answers and raised the confidence level of participants for their answers, if they had given correct answers. Adhikari, Tax, Satta & Faeth (2019) and Rabold, Siebers & Schmid (2021) have shown that example based explanations are being considered more beneficial by users. Particularly, after examining the results of Rabold, Siebers & Schmid (2021), Rule based is most preferred in the abstract domain while Example based is most preferred in the visual domain. Nevertheless, in the abstract domain, rule based are preferred combined with the Example based explanations. Overall, near miss are preferred over far miss and E + far miss are preferred over R+ far miss. In visual domain, E+ far miss are preferred over R+ near miss.

Even though we did not use a visual domain or a combination of E+ R based types, our results are in alignment with the preference in Example based explanations, as they affected the most the confidence level of participants regardless their level of expertise. The difference here, is that the level of complexity/abstraction in our abstract domain is really high while in the other studies is really low and this aspect, should be considered as well. It has to be noted that our example based explanations were consisting of three different examples while our Prototype based type was chosen as the simplest and most representative of these examples. People with high level of expertise were affected in a significant degree by Example based explanations. An interesting pattern, though not significant, was that Near miss examples influenced mostly people with high level of expertise, while Far miss types influenced mostly people with medium and low level of expertise. This could be explained by the fact that due to limited working memory, our cognitive system uses schemata (Hogg & Vaughan, 2008; Kuhn, Amlani & Rensink, 2008; Kuhn & Rensink, 2016; Van de Cruys & Wagemans, 2011; Eysenck, 2006; Πόθος & Οικονόμου, 2010), these scripts are not available to people with lower level of expertise and thus, it is difficult for them to process a lot of unknown information at the same time. Far miss types are clearer and less complex for non-experts in order to spot contrasts and differences. Therefore, although proximity is an important variable as Miller, Howe & Sonenberg (2017) have stated, it seems that knowledge background determines this preference. We need further researches in order to explore if this pattern exists or not.

Another interesting finding is that the Prototype based explanation affected the most, in a significant degree, the participants regarding the changing of their responses. By and large, we could say that Example based explanations influence the trust level while Prototype based influence the behavior (most persuasive). Admittedly, there is a contradictive finding here. Why Example based explanations affect the confidence level and not the behavior of the participants and why Prototype based types affect the behavior and not the confidence level of participants. This may be due to the Reactance (theory of Brehm) where people try to resist if feeling threatened because of their eliminating freedom or feeling that they are the target of a persuasion effort (Hogg & Vaughan, 2008). As it has already been mentioned, example based types were consisting of three different examples which were presented to participants, while the prototype based was the simplest and shortest one of them. For this reason, participants might felt as if the machine was trying to persuade them to change their mind.

Seemingly, contrastive explanations reinforce the interpretability and understanding of XAI explanations. According to surveys, prototype counter-factual types (Hase & Bansal, 2020), Sensitivity based explanations (this could be interpreted as a Near miss type) (Shulner-Tal, Kuflik & Kliger, 2022), contrastive versions (Kobayashi, 2019) and consider the opposite technique (Lee et al., 2016), enhance the understanding and affect the beliefs of users. In these experiments, we did not use this type of measurement but we suggest that the integration of measurements such as trust, persuasiveness and comprehension at the same time, will give clearer results.

As we have already mentioned in the introduction, counter-arguments are the basic structure element of mental reasoning in humans (Eysenck, 2006; Miller, 2019). People with limited working memory cannot create a lot of contrastive mental models at the same time (Eysenck, 2006). This cognitive phenomenon might explain the fact that people with higher working memory had better performance on the quizzes compared to people with lower working memory. The most impressive result here is that people with high level of expertise did not have better performance than those with the lowest degree of specialization. This can be linked to the findings of Lee et al. (2016) where experts had lower performance than non-experts due to heuristics. Familiar patterns even if wrong, are preferred by our perceptual system due to reasons we have already explained in the introduction.

People with high level of working memory capacity not only did they have better performance but also were more consistent in their choices as well, in contrast to people with high specialization level that did not have the same results. Therefore, short term memory capacity affects consistency and performance while specialization did not show any effect on these two. Nevertheless, the cognitive load of participants did not show any significant difference between the three different conditions. This could be due to three reasons, the first is that the right point to measure their cognitive load was during giving explanations or counter-arguments to the machine in the second and third condition respectively, the second is that the complexity of the secondary task needed to be higher and, the third is that Electroencephalography (EEG) is the only way to have accurate results. The former originally seemed very confusing because participants would have to type via their keyboard while trying to respond to the secondary task through the keyboard as well. Additionally, it has to be reported that a large number of the sample did not react to this task at all. This might say that they did not acknowledge it as a part of the experimental process.

Even though different conditions did not affect the cognitive load, they considerably affected the consistency variable. In conditions where participants had to explain their answers, consistency reached one hundred percent. More precisely, in the condition two, participants had to explain their answer and the machine presented again this explanation to them in a later stage. In condition three, participants had to explain their answer by providing a counter-argument to the explanation of the machine. It is quite impressive that in these two versions, nobody changed their answers regardless of being correct or incorrect even though there were variations in their confidence level. Therefore, we do get better consistency as a result of using arguments. Another important finding is that people of high level of expertise increased their performance significantly on both condition two and three as well. In general, all participants seemed to increase their performance on these last two conditions and especially in the second condition. Subsequently, argumentation might promote consistency but reduces biases as well.

The effect of better performance, mainly in condition two, was expected, as participants were challenged to use the system type II, in order to resolve the quizzes. The condition one mostly triggered the heuristics of the participants and, for this reason the performance was lower. As Eysenck (2006) and Πόθος & Οικονόμου (2010) have stated, the importance/necessity of the task, the cognitive skills of the individual, the specialized knowledge, emotional state, motives, the time pressure and, the nature of the task (multi-tasking or not) are some factors that determine which system will be used. Hence, participants with higher working memory and participants with higher specialization who have been motivated to explain their answer had better performance by using the system type II. On the other side, the perfect degree of consistency on both condition two and three, could be explained by Consistency and Balance theory which support that people tend to maintain an internal balance by constructing coping mechanisms in order to avoid incongruences in their cognitions (Hogg & Vaughan, 2008).

Overall, as other researchers have reported, explanations should be adapted based on the context (Amann et al., 2022), target group, that is, experts/non-experts (Burkart & Huber, 2021; Amann et al., 2022) and questions (Mothilal, Sharma & Tan 2021). We saw that our results differed depending on the short-term memory capacity, the level of expertise and the condition. Similarly, Burkart & Huber (2021) argued that persuasive explanations are related to cognitive functions, user preferences and expertise. Thus, personalized transparency seems very reasonable (Schelenz, Segal & Gal, 2020). For instance, if argumentation reduces biases which are based on heuristics, personalized XAI would reinforce this productive dialogue in HMI. Apparently, as Kakas (2020) has argued, “weak fallacies will not survive within a context of reasoning”. Even if prejudices could not be completely eradicated in this context, they are certain to be minimized.

To conclude: 1) Example based explanations enhanced trust while prototype influenced the behavior of participants. Nevertheless these findings are not very clear and thus, we need to incorporate psychometrics for comprehension, trust and persuasiveness together as well as self-report tools, 2) Example based explanations affected in a significant degree the confidence level of participants with high level of expertise. We need further researches to explore if non-experts prefer far miss types and experts prefer near types because this was an observed pattern but not a significant one, 3) Argumentation/mutual explainability promotes consistency, boosts better performance and eliminates biases, 4) The effect of argumentation had a really high significant effect on people with high level of expertise, that is better performance and, 5) High working memory capacity increases both performance and consistency while high level of expertise did not show any significant linear association.

3.1 Limitations and future work

This study was conducted via online means and thus, even though the variables were counterbalanced, the experimental environment was not well controlled. According to Christensen (2001) sometimes, an increased ecological validity can minimize the internal validity. For this reason, we believe that an online experiment can simulate in a better degree real-life case scenarios but a repeated measurement in a laboratory can confirm or contradict the results. Therefore, a combination of both cases, would be the ideal scenario to verify the results of the study.

Some other factors that could direct the results could be the interpolation of multiple experimental conditions, the Hawthorne effect, the impact of innovation/distraction, the effect of the researcher, the effect of preliminary examination and the multi-layered temporal validity (Christensen, 2001). Let’s consider each of these potential influencing factors separately. In this study, we used three different conditions in order to decrease the learning and transfer across the conditions and thus interpolation could not affect the results. Nevertheless, due to the nature of the experiment, some participants could leave the experiment and start from the beginning. If this was the case, some participants in their second trial, could be assigned by the system in a different condition.

The Hawthorne effect, that is, when participants are aware that they participate in a study, was tried to be controlled by informing participants the aims of the study which were very far from measuring their IQ or testing their knowledge. In a laboratory experiment, it's easier for the researcher to present the experiment in a beneficial way and give more details after the experimental process. Furthermore, in this study the researcher could not affect the participants nor expose the subjects in a preliminary examination. Last but not least, the temporal validity could be verified through other researches.

In general, the required number of participants for this study was 261 but for some of the statistical analyses, our sample was divided to three different levels of expertise. Nevertheless a large number of participants did not reveal their field of studies. It would be more precise if the sample was consisting of 261 participants for each of these categories. In addition to that, levels of specialization could also be divided based on the level of education, that is, if participants had a bachelor, master or doctorate degree on fields related to the quizzes.

However, research on this area should focus on exploring more criteria in order to build personalized systems based on the cognitive needs of users. We have shown that variables such as working memory capacity, level of expertise and argumentative conditions have significant effects. More specifically, we need to consider variables such as elementary cognitive factors such as working memory capacity and high level cognitive factors such as cognitive styles (Belk, 2021). Depending on the factor, we could present the types of explanation with multiple ways. For example, for low working memory capacity, in the explanation screen, we would provide a storage tool where participants could see again information of the quizzes.

For cognitive styles such as imager or verbal, we would present the explanations with diagrammatic information combined with textual information and textual presentation, respectively. Moreover, for wholist/field dependent or analyst/field independent people, we would present the explanations with the quizzes in guided steps with expandable content and a whole preview of quiz and explanation in a single one-page, respectively. Nevertheless, to establish ecological validity, systems that uses human factors to adapt their content should be adaptive and refresh their clustering from time to time because people can change.

Another limitation of our work is that we excluded the scale of understanding, trust and satisfaction for the answers of the machine. Therefore, we also suggest for future researches to integrate psychometrics for trust, persuasiveness and comprehension, as well as self-report tools in order to have more complete results. The internal validity of the study will be increased if behavioral metrics converge with the self-report measurements. Some supplementary factors that could also be considered in future researches, are the gender counterbalancing and the inclusion of E+ R based type of explanation as these coefficients were absent in this study.

Last but not least, the measurement of cognitive load throughout the conditions was not effective. It's possible that a concrete number of participants did not pay enough attention to this task as we have a lot of zeros reactions in the data. For this reason, an EEG measurement of cognitive load should be more appropriate. Although, we acknowledge that the precisely correct point for this measurement should be during the typing stage where participants were called to explain their decisions. Future researches will have to use both these types of measurements in order to be able to conclude to hypotheses.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). *Trends and Trajectories for Explainable, Accountable and Intelligible Systems* [Conference Session]. CHI '18: CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3173574.3174156.

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160.

Adhikari, A., Tax, D. M., Satta, R., & Faeth, M. (2019, Aug 16). *Leafage: Example-based and feature importance-based explanations for black-box ML models* [Conference session]. Fuzzy Systems (FUZZ-IEEE).). https://doi.org/10.1109/fuzz-ieee.2019.8858846.

Alikhademi, K., Richardson, B., Drobina, E., & Gilbert, J. E. (2021, Jun 14). *Can Explainable AI Explain Unfairness? A Framework for Evaluating Explainable AI* [Conference session]. https://arxiv.org/pdf/2106.07483.pdf.

Altay, S., Schwartz, M., Hacquin, A. S., Allard, A., Blancke, S., & Mercier, H. (2022). Scaling up interactive argumentation by providing counterarguments with a chatbot. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01271-w.

Alves, P., Saraiva, P., Carneiro, J., Campos, P., Martins, H., Novais, P., & Marreiros, G. (2020, Jul 14-17). Modeling Tourists' Personality in Recommender Systems [Conference Session]. User Modeling, Adaptation and Personalization. ACM. https://doi.org/10.1145/3340631.3394843.

Amann, J., Vetter, D., Blomberg, S. N., Christensen, H. C., Coffee, M., Gerke, S., Gilbert, T. K., Hagendorff, T., Holm, S., Livne, M., Spezzatti, A., Strümke, I., Zicari, R. V., & Madai, V. I. (2022). To explain or not to explain? Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health*, *1*(2), Article e0000016. https://doi.org/10.1371/journal.pdig.0000016.

Angeli, C., Valanides, N., & Kirschner, P. (2009). Field dependence–independence and instructional-design effects on learners' performance with a computer-modeling tool. *Computers in Human Behavior*, *25*(6), 1355-1366.

Arnheim, R. (1974). *Τέχνη και οπτική αντίληψη: Η ψυχολογία της δημιουργικής όρασης*. Μτφρ. Ι. Ποταμιανός. Αθήνα: Θεμέλιο. 2005.

Arnheim, R. (1997). *Οπτική σκέψη*. Μτφρ. Ι. Ποταμιανός. Θεσσαλονίκη: University Studio Press. 2007.

Belk, M. (2021). *Adaptive and Interactive Systems* [Class handout]. Open University of Cyprus, Cyprus, COS614.

Belk, M., Germanakos, P., Constantinides, A., & Samaras, G. (2015). A Human Cognitive Processing Perspective in Designing E-Commerce Checkout Processes. In *Human-Computer Interaction – INTERACT 2015* (pp. 523–530). Springer International Publishing.

Belk, M., Fidas, C., Germanakos, P., & Samaras, G. (2017). The interplay between humans, technology and user authentication: A cognitive processing perspective. *Computers in Human Behavior, 76*, 184–200.

Bodria, F., Giannotti, F., Guidotti, R., Naretto, F., Pedreschi, D., & Rinzivillo, S. (2021, Feb 25). *Benchmarking and survey of explanation methods for black box models* [Conference session]. https://arxiv.org/abs/2102.13076.

Bruckert, S., Finzel, B., & Schmid, U. (2020). The Next Generation of Medical Decision Support: A Roadmap Toward Transparent Expert Companions. *Frontiers in Artificial Intelligence*, 3.

Burkart, N., & Huber, M. F. (2021). A Survey on the Explainability of Supervised Machine Learning. *Journal of Artificial Intelligence Research*, *70*, 245–317.

Chazette L, Brunotte W & Speith T. (2021) *Exploring explainability: a definition, a model, and a knowledge catalogue* [Conference session]. Engineering conference (RE). https://arxiv.org/abs/2108.03012.

Chen, S. Y., & Liu, X. (2011). Mining students' learning patterns and performance in Web-based instruction: a cognitive style approach. *Interactive Learning Environments, 19*(2), 179–192.

Christensen, L. B. (2001). *Η πειραματική μέθοδος στην επιστημονική έρευνα*. Μτφρ. Α. Γιαννακουλόπουλος & Ν. Παπασταύρου. Αθήνα: Παπαζήση. 2007.

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-Scene Perception Requires Attention. *Psychological Science, 22*(9), 1165-1172.

Cyras, K., Badrinath, R., Mohalik, S. K., Mujumdar, A., Nikou, A., Previti, A., Sundararajan, V., & Feljan, A. V. (2020, Dec 2). *Machine Reasoning Explainability* [Conference session]. https://arxiv.org/abs/2009.00418.

Demacheva, I., Ladouceur, M., Steinberg, E., Pogossova, G., & Raz, A. (2012). The Applied Cognitive Psychology of Attention: A Step Closer to Understanding Magic Tricks. *Applied Cognitive Psychology*, *26*(4), 541-549.

Doshi-Velez, F. & Kim, B. (2018, Feb 28*). Considerations for evaluation and generalization in interpretable machine learning* [Conference session]. Explainable and Interpretable Models in Computer Vision and Machine Learning, Berlin, Heidelberg, Germany. https://link.springer.com/chapter/10.1007/978-3-319-98131-4_1.

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., Riener, A., & Riedl, M. O. (2021). *Operationalizing Human-Centered Perspectives in Explainable AI* [Conference Session]. CHI '21: CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3411763.3441342.

Eysenck, M. W. (2006). *Βασικές αρχές γνωστικής ψυχολογίας*. Μτφρ. Μ. Κουλεντιανού. Αθήνα: Gutenberg. 2010.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical poweranalysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Ferrario, A., & Loi. M. (2020). *A series of unfortunate counterfactual events: The role of time in counterfactual explanations* [Conference session]. https://arxiv.org/ftp/arxiv/papers/2010/2010.04687.pdf.

Galhotra, S., Pradhan, R., & Salimi, B. (2021, Jun 23). *Explaining Black-Box Algorithms Using Probabilistic Contrastive Counterfactuals* [Conference session]. https://arxiv.org/abs/2103.11972.

Guidotti, R. (2020). Explaining Explanation Methods. http://ceur-ws.org/Vol-2741/paper-04.pdf.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, *51*(5), 1–42.

Hase, P., & Bansal, M. (2020, May 4). *Evaluating explainable AI: Which algorithmic explanations help users predict model behavior?* [Conference session]. In Proceedings of the 58th Annual Association for Computational Linguistics. https://arxiv.org/abs/2005.01831.

Herbig, N., Düwel, T., Helali, M., Eckhart, L., Schuck, P., Choudhury, S., & Krüger, A. (2020, Jul 14-17). Investigating Multi-Modal Measures for Cognitive Load Detection in E-Learning [Conference Session]. User Modeling, Adaptation and Personalization. ACM. https://doi.org/10.1145/3340631.3394861.

Hogg, M. A., & Vaughan, G. M. (2008). *Κοινωνική ψυχολογία*. Μτφρ. Ε. Βασιλικός & Α. Αρβανίτης. Αθήνα: Gutenberg. 2010.

John-Mathews, J.-M. (2021). *A Critical Empirical Study of Black-box Explanations in AI* [Conference session]. Information Systems (ICIS). https://arxiv.org/abs/2109.15067.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341–350.

Kakas, A. (2020). *Computational Argumentation and Human-Centric AI* [Discussion]. International Workshop on Computational Argumentation and Cognition. http://cognition.ouc.ac.cy/cognitar/.

Karimi, H., Barthe, G., Schölkopf, B., & Valera, I. (2021, Mar 1). *A survey of algorithmic recourse: contrastive explanations and consequential recommendations* [Conference session]. https://arxiv.org/pdf/2010.04050.pdf.

Karimi, H., Schölkopf, B., & Valera, I. (2020, Oct 8). *Algorithmic recourse: from counterfactual explanations to interventions* [Conference session]. Fairness, accountability, and transparency. https://arxiv.org/abs/2002.06278.

Katsini, C., Fidas, C., Raptis, G. E., Belk, M., Samaras, G., & Avouris, N. (2018, April 21-26). *Influences of Human Cognition and Visual Behavior on Password Strength during Picture Password Composition* [Conference Session]. Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3173574.3173661.

Kobayashi, K. (2019). Communicating highly divergent levels of scientific and social consensus: its effects on people's scientific beliefs. *Social Influence*, *14*(3-4), 65–76.

Kuhn, G., Amlani, A., & Rensink, R. (2008). Towards a science of magic. *Trends In Cognitive Sciences, 12*(9), 349-354.

Kuhn, G. & Rensink, R. (2016). The Vanishing Ball Illusion: A new perspective on the perception of dynamic events. *Cognition*, *148*, 64-70.

Lampridis, O., Guidotti, R., & Ruggieri, S. (2020). Explaining Sentiment Classification with Synthetic Exemplars and Counter-Exemplars. *In Discovery Science* (pp. 357–373). Springer International Publishing. https://doi.org/10.1007/978-3-030-61527-7_24.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research, 10*(2), 141–160.

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1641), 20130205.

Lee, Y.-H., Dunbar, N. E., Miller, C. H., Lane, B. L., Jensen, M. L., Bessarabova, E., Burgoon, J. K., Adame, B. J., Valacich, J. J., Adame, E. A., Bostwick, E., Piercy, C. W., Elizondo, J., & Wilson, S. N. (2016). Training Anchoring and Representativeness Bias Mitigation Through a Digital Game. *Simulation & Gaming*, *47*(6), 751–779.

Liao, V. Q., Gruen, D., & Miller, S. (2020, Sep 3). *Questioning the AI: Informing Design Practices for Explainable AI User Experiences* [Conference session]. Human Factors in Computing

Systems: Association for Computing Machinery, New York, NY, USA. https://arxiv.org/abs/2104.03483.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*.

Mercier, H. (2016). The Argumentative Theory: Predictions and Empirical Evidence. *Trends in Cognitive Sciences*, *20*(9), 689–700.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57–74.

Michael, L. (2020a). Machine Coaching [Unpublished raw data].

Michael, L. (2019, Aug 11). *Machine Coaching* [Paper presentation]. Workshop on Explainable Artificial Intelligence, p. 80–86. https://doi.org/10.5281/zenodo.3931266.

Michael, L. (2020b, May 13). *Machine Ethics through Machine Coaching* [Paper presentation]. Workshop on Implementing Machine Ethics. https://www.internetofus.eu/wp-content/uploads/sites/38/2021/08/Loizos-Michael-ethics-through-coaching.pdf.

Michael, L. (in press). *Explainability and the Fourth AI Revolution*. Handbook Of Research On Artificial Intelligence, Innovation And Entrepreneurship. https://arxiv.org/ftp/arxiv/papers/2111/2111.06773.pdf.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38.

Miller, T., Howe, P., & Sonenberg, L. (2017, Dec 5). *Explainable AI: Beware of Inmates Running the Asylum. Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences* [Paper Presentation]. Workshop on Explainable Artificial Intelligence (XAI), IJCAI. https://arxiv.org/abs/1712.00547.

Mothilal, R. K., Sharma, A., & Tan, C. (2021, May 29). *Explaining machine learning classifiers through diverse counterfactual explanations* [Conference session]. https://arxiv.org/abs/2011.04917.

Noah, S., & Mangun, G. R. (2019). Recent evidence that attention is necessary, but not sufficient, for conscious perception. *Annals of the New York Academy of Sciences*.

Πόθος, Ε., & Οικονόμου, Η. (2010). *Θέματα γνωσιακής ψυχολογίας* (Γ. Μαμάης, Επίμ). Gutenberg.

Park, B., & Brünken, R. (2014). The Rhythm Method: A New Method for Measuring Cognitive Load-An Experimental Dual-Task Study. *Applied Cognitive Psychology*, *29*(2), 232–243.

Piaget, J. (1928). *Judgment and reasoning in the child*. K. Paul, Trench, Trubner & Co. ltd.

Rabold, J., Siebers, M., & Schmid, U. (2021). *Generating contrastive explanations for inductive logic programming based on a near miss approach* [Conference session]. Machine Learning. https://arxiv.org/abs/2106.08064.

Ramon, Y., Martens, D., Provost, F., & Evgeniou, T. (2020, Dec 4). *A comparison of instance-level counterfactual explanation algorithms for behavioral and textual data: SEDC, LIME-C and SHAP-C* [Conference session]. Advances in Data Analysis and Classification. https://arxiv.org/abs/1912.01819.

Raptis, G. E., Fidas, C., & Avouris, N. (2018). Effects of mixed-reality on players' behaviour and immersion in a cultural tourism game: A cognitive processing perspective. *International Journal of Human-Computer Studies, 114*, 69–79.

Rehse, J.-R., Mehdiyev, N., & Fettke, P. (2019). Towards Explainable Process Predictions for Industry 4.0 in the DFKI-Smart-Lego-Factory. *KI - Künstliche Intelligenz*, *33*(2), 181–187.

Ribeiro, M., Singh, S., & Guestrin, C. (2016). "*Why Should I Trust You?": Explaining the Predictions of Any Classifier* [Conference Session]. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. Association for Computational Linguistics. https://doi.org/10.18653/v1/n16-3020.

Rosenfeld, A. (2021, May 21). *Better metrics for evaluating explainable artificial intelligence* [Conference session]. Autonomous agents and multiagent systems. https://dl.acm.org/doi/abs/10.5555/3463952.3463962.

Schelenz, L., Segal, A., & Gal, K. (2020) *Best practices for transparency in machine generated personalization* [Conference session]. User Modeling, Adaptation and Personalization. https://doi.org/10.1145/3386392.3397593.

Schmid, U., & Finzel, B. (2020). Mutual Explanations for Cooperative Decision Making in Medicine. *KI - Künstliche Intelligenz*, *34*(2), 227–233.

Shulner-Tal, A., Kuflik, T., & Kliger, D. (2022). Fairness, explainability and in-between: understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics and Information Technology*, *24*(1).

Slack, D., Hilgard, S., Lakkaraju, H., and Singh, S. (2021, Nov 3). *Counterfactual explanations can be manipulated* [Conference session]. https://arxiv.org/abs/2106.02666.

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2016). Visual search behaviors of association football referees during assessment of foul play situations. *Cognitive Research: Principles and Implications, 1*(1).

Stanley, D., Ferneyhough, E., & Phelps, E. A. (2009). *Neural Perspectives on Emotion: Impact on Perception, Attention, and Memory*. Handbook of Neuroscience for the Behavioral Sciences.

Teso, S., & Kersting, K. (2019, Jan 27-28). *Explanatory interactive machine learning* [Conference session]. AI, Ethics, and Society. https://dl.acm.org/doi/10.1145/3306618.3314293.

Tompkins, M., Woods, A., & Aimola Davies, A. (2016). The Phantom Vanish Magic Trick: Investigating the Disappearance of a Non-existent Object in a Dynamic Scene. *Frontiers In Psychology*, *7*.

Tung, Y. W., & Chan, K. C. C. (2009). A Unified Human–Computer Interaction Requirements Analysis Framework for Complex Socio-technical Systems. *International Journal of Human-Computer Interaction*, *26*(1), 1–21.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: *Heuristics and Biases. Science, 185*(4157), 1124–1131.

Ugwitz, P., Juřík, V., Herman, L., Stachoň, Z., Kubíček, P., & Šašinka, Č. (2019). Spatial Analysis of Navigation in Virtual Geographic Environments. *Applied Sciences, 9*(9), 1873.

Van de Cruys, S., & Wagemans, J. (2011). Gestalts as predictions: Some reflections and an application to art. *Gestalt Theory, 33*(3-4), 325–344.

Van Stekelenburg, A., Schaap, G., Veling, H., & Buijzen, M. (2021). Boosting Understanding and Identification of Scientific Consensus Can Help to Correct False Beliefs. *Psychological Science*, 095679762110077.

Vassiliades, A., Bassiliades, N., & Patkos, T. (2021). Argumentation and explainable artificial intelligence: a survey. *The Knowledge Engineering Review*, *36*.

Verma, S., Dickerson, J., & Hines, K. (2020, Oct 20). *Counterfactual Explanations for Machine Learning: A Review* [Conference session]. https://arxiv.org/pdf/2204.06917.pdf.

Vernon, D., Metta, G., & Sandini, G. (2007). A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents. *IEEE Transactions on Evolutionary Computation, 11*(2), 151–180.

Vultureanu-Albişi, A., & Bădică, C. (2022). A survey on effects of adding explanations to recommender systems. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.6834.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3063289.

Zhong, J., & Negre, E. (2021, Jun). *AI: To interpret or to explain?* [Conference Session]. Congrès Inforsid (INFormatique des ORganisations et Systèmes d'Information et de Décision), Dijon, France. http://inforsid.fr/actes/2021/INFORSID_2021_p149-164.pdf.

Appendices

Appendix A

A. Table of Acronyms

Acronym	Meaning
AI	Artificial Intelligence
ML	Machine Learning
DNNs	Deep Neural Networks
RL	Reinforcement Learning
XAI	Explainable Artificial Intelligence
HMI	Human-Machine Interaction
ME	Mutual Explainability
SL	Supervised Learning
XIXO	Explanation In, Explanation Out
PAC	Probably Approximately Correct
MR	Machine Reasoning
LIME	Interpretable Model-agnostic Explanation
PRP	Psychological Refractory Period
R+ Far miss	Rule and Far miss combination type of explanation
R+ Near miss	Rule and Near miss combination type of explanation
E+ Far miss	Example and Far miss combination type of explanation
E+ Near miss	Example and Near miss combination type of explanation
P+ Far miss	Prototype and Far miss combination type of explanation
P+ Near miss	Prototype and Near miss combination type of explanation
E+ R based	Example and Rule based combination type of explanation
EEG	Electroencephalography

Appendix B

B. Quizzes

First Quiz

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms one description, chosen at random from the 100 available descriptions. Please indicate your probability that the person described is an engineer, on a scale from 0 to 100.

Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies, which include home carpentry, sailing, and mathematical puzzles.

What is the probability that Jack is one of the 30 engineers?

A. 10-40 percent

B. 40-60 percent

C. 60-80 percent

D. 80-100 percent

(Kahneman & Tversky, 1973, p. 241)

Second Quiz

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

A. The larger hospital

B. The smaller hospital

C. About the same (that is, within 5% of each other)

(Kahneman & Tversky, 1972, p. 443)

Third Quiz

Imagine an urn filled with balls, of which 2/3 are of one color and 1/3 of another. One individual has drawn 5 balls from the urn, and found that 4 were red and 1 was white. Another individual has drawn 20 balls and found that 12 were red and 8 were white. Which of the two individuals should feel more confident that the urn contains 2/3 red balls and 1/3 white balls, rather than the opposite?

A. First person

B. Second person

(Tversky & Kahneman, 1974, p. 1125)

Forth Quiz

Case 1: Imagine that you have decided to watch a movie tonight and you already purchased a ticket for $10. As you enter the theater, you discover that you have lost the ticket. The ticket office tells you they keep no records so you will have to purchase a new ticket to see the movie. Would you still pay $10 for another ticket?

Case 2: Imagine that you have decided to watch a movie tonight and the ticket is $10. As you enter the theater, you discover that have lost $10 bill on the way. Would you still pay $10 for a ticket for the movie?

Case 1: A. Yes

Case 1: B. No

Case 2: A. Yes

Case 2: B. No

(Kahneman & Tversky, 1984, p. 347)

Fifth Quiz

Choose between getting $900 for sure or a 90 percent chance of getting $1,000.

A. Getting $900

B. 90 percent chance of getting $1,000

(Kahneman, 2011, p. 271)

Sixth Quiz

Choose between losing $900 for sure or a 90 percent chance of losing $1,000.

A. Losing $900

B. 90 percent chance of losing $1,000

C. Whatever

(Kahneman, 2011, p. 271)

Seventh Quiz

There are two programs in a high school. Boys are a majority (65%) in program A, and a minority (45%) in program B. There is an equal number of classes in each of the two programs. You enter a class at random, and observe that 55% of the students are boys. What is your best guess-does the class belong to program A or to program B?

A. A program

B. B program

(Kahneman & Tversky, 1972, p. 433)

Eighth Quiz

On each round of a game, 20 marbles are distributed at random among five children: Alan, Ben, Carl, Dan, and Ed. Consider the following distributions:

I	II
- -	
Alan 4	Alan 4
Ben 4	Ben 4
Carl 5	Carl 4
Dan 4	Dan 4
Ed 3	Ed 4

In many rounds of the game, will there be more results of type I or of type II?

A. TYPE I

B. TYPE II

(Kahneman & Tversky, 1972, p. 434)

Ninth Quiz

All the families having exactly six children in a particular city were surveyed. In 72 families, the exact order of births of boys (B) and girls (G) was G B G B B G. What is your estimate of the number of families surveyed in which the exact order of births was B G B B B B? Choose the closest to your answer.

A. 18

B. 36

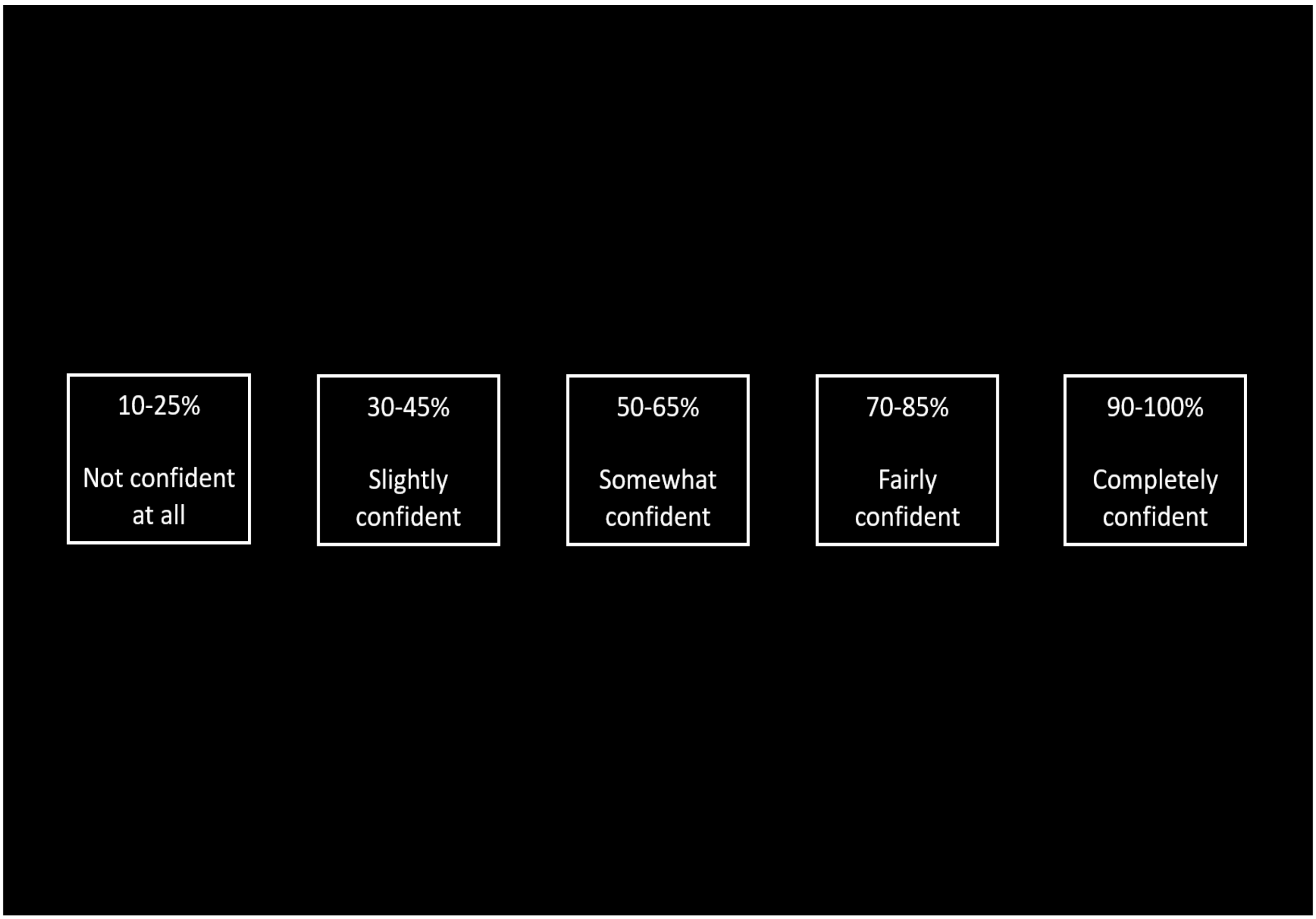
C. 72

D. 102

(Kahneman & Tversky, 1972, p. 432)

Appendix C

C. Confidence rating scale.



Appendix D

D. Answers of the machine.

FIRST QUIZ

Q1_A1

THE RIGHT ANSWER IS OPTION A.

Unconditional probability is the probability of an event regardless of the preceding or future occurrence of other events. In simplest terms, unconditional probability is simply the probability of an event occurring, that is, the number of favorable outcomes divided by the total number of outcomes possible. **With that being said, the probability of Jack being an engineer is 30% because, there is no tangible evidence correlating the possible outcome and the occurrence of the other conditions.**

Q1_A2

THE RIGHT ANSWER IS OPTION A.

Unconditional probability is the probability of an event regardless of the preceding or future occurrence of other events. In simplest terms, unconditional probability is simply the probability of an event occurring, that is, the number of favorable outcomes divided by the total number of outcomes possible. **With that being said, the probability of Jack being an engineer is 30% because, there is no tangible evidence correlating the possible outcome and the occurrence of the other conditions.**

ON THE CONTRARY, we could not apply **standard deviation** because it's a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range. In that case the standard deviation of the specialty of the 100 professions would be:

$(70 + 30)/2 = 50$, $[(70- 50)^2 + (30 - 50)^2] /1 = 800/1 = 800 = \sqrt{800} = 28.284271247462$.

Q1_A3

<u>THE RIGHT ANSWER IS OPTION A.</u>

<u>Unconditional</u> **<u>probability</u>** is the probability of an event regardless of the preceding or future occurrence of other events. In simplest terms, unconditional probability is simply the probability of an event occurring, that is, the number of favorable outcomes divided by the total number of outcomes possible. **<u>With that being said, the probability of Jack being an engineer is 30% because, there is no tangible evidence correlating the possible outcome and the occurrence of the other conditions.</u>**

ON THE OTHER SIDE, in probability theory, **<u>conditional</u>** **<u>probability</u>** is a measure of the probability of an event occurring, given that another event (by assumption, presumption, assertion or evidence) has already occurred. In our case we have four variables/events (no interest in political and social issues, home carpentry, sailing, and mathematical puzzles). If we had evidence at least for one of the abovementioned variables and that is, engineers who play mathematical puzzles = 0.41 and people who play mathematical puzzles = 0.86 then, we would have 0.41 / 0.86 = 47.7%. Thus, the probability of Jack being an engineer would be 47.7%.

Q1_A4

THE RIGHT ANSWER IS OPTION A. THEORY APPLIED: Unconditional probability

Example 1: If a die lands on the number five 15 times out of 60, the unconditional probability of landing on the number five is 25% (15 outcomes /60 total lots = 0.25).

Example 2: The chance of a fair coin flip being heads has an unconditional probability of 50% (1/2) regardless of how many coin flips preceded it, nor if some other event had occurred. Each toss of a coin is a perfect isolated thing. What it did in the past will not affect the current toss. The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

Example 3: Let's examine a group of stocks and their returns. A stock can either be a winner, which earns a positive return, or a loser, which has a negative returns. Say that out of five stocks, stocks A and B are winners, while stocks C, D, and E are losers. What, then, is the unconditional probability of choosing a winning stock? Since two outcomes out of a possible five will produce a winner, the unconditional probability is 2 successes divided by 5 total outcomes (2 / 5 = 0.4), or 40%.

Q1_A5

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example 1: If a die lands on the number five 15 times out of 60, the unconditional probability of landing on the number five is 25% (15 outcomes /60 total lots = 0.25).

THEORY THAT CANNOT BE APPLIED: Standard deviation

Example 1: There are 39 plants in the garden. A few plants were selected randomly and their heights in cm were recorded as follows: 51, 38, 79, 46, 57. Calculate the variation (standard deviation) of their heights. Solution:

N = 5, Mean = (51+38+79+46+57)/5 = 54.2, Standard Deviation= √(51−54.2)2+(38−54.2)2+(79−54.2)2+(46−54.2)2+(57−54.2)24(51−54.2)2+(38−54.2)2+(79−54.2)2+(46−54.2)2+(57−54.2)24 = 15.5

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example 2: The chance of a fair coin flip being heads has an unconditional probability of 50% (1/2) regardless of how many coin flips preceded it, nor if some other event had occurred. Each toss of a coin is a perfect isolated thing. What it did in the past, will not affect the current toss. The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

THEORY THAT CANNOT BE APPLIED: Standard deviation

Example 2: In a class of 50, 4 students were selected at random and their total marks in the final assessments are recorded, which are: 812, 836, 982, 769. Find the variation (standard deviation) of their marks. Solution:

N = 4, Sample Mean = (812+836+982+769)/4 = 849.75, Variance = $[(812 - 849.75)^2 + (836 - 849.75)^2 + (982 - 849.75)^2 + (769 - 849.75)^2] /3 = 92.4$, Standard Deviation = $\sqrt{92.4} = 2\sqrt{23.1} = 9.6$

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example 3: Let's examine a group of stocks and their returns. A stock can either be a winner, which earns a positive return, or a loser, which has a negative returns. Say that out of five stocks, stocks A and B are winners, while stocks C, D, and E are losers. What, then, is the unconditional probability of choosing a winning stock? Since two outcomes out of a possible five will produce a winner, the unconditional probability is 2 successes divided by 5 total outcomes (2 / 5 = 0.4), or 40%.

THEORY THAT CANNOT BE APPLIED: Standard deviation

Example 3: A market researcher is analyzing the results of a recent customer survey that ranks a product from 1 to 10. He wants to have some measure of the reliability of the answers received in the survey in order to predict how a larger group of people might answer the same questions.

Solution:

Scores for the survey: 9, 7, 10, 8, 9, 7, 8, and 9. Mean: 8.4. Variance: 1.12. Standard Deviation =1.06. The standard deviation is 1.06, which is somewhat low. The researcher now knows that the results of the sample size are probably reliable.

Q1_A6

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example 1: If a die lands on the number five 15 times out of 60, the unconditional probability of landing on the number five is 25% (15 outcomes /60 total lots = 0.25).

THEORY THAT CANNOT BE APPLIED: Conditional Probability

Example 1: Given that you drew a red card, what's the probability that it's a four (four|red) = 2/26= 1/13. So out of the 26 red cards (given a red card), there are two fours so 2/26=1/13.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example 2: The chance of a fair coin flip being heads has an unconditional probability of 50% (1/2) regardless of how many coin flips preceded it, nor if some other event had occurred. Each toss of a coin is a perfect isolated thing. What it did in the past, will not affect the current toss. The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

THEORY THAT CANNOT BE APPLIED: Conditional Probability

Example 2: Event A is that an individual applying for college will be accepted. There is an 80% chance that this individual will be accepted to college. Event B is that this individual will be given dormitory housing. Dormitory housing will only be provided for 60% of all of the accepted students. (Dormitory Housing | Accepted) = (0.60)*(0.80) = 0.48.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example 3: Let's examine a group of stocks and their returns. A stock can either be a winner, which earns a positive return, or a loser, which has a negative returns. Say that out of five stocks, stocks A and B are winners, while stocks C, D, and E are losers. What, then, is the unconditional probability of choosing a winning stock? Since two outcomes out of a possible five will produce a winner, the unconditional probability is 2 successes divided by 5 total outcomes (2 / 5 = 0.4), or 40%.

THEORY THAT CANNOT BE APPLIED: Conditional Probability

Example 3: There are 2 blue and 3 red marbles in a bag. What are the chances of getting a blue marble? The chance is 2 in 5. But after taking one out, the chances change! So the next time: probability marbles if we got a red marble before, then the chance of a blue marble next is 2 in 4 probability marbles if we got a blue marble before, then the chance of a blue marble next is 1 in 4.

Q1_A7

THE RIGHT ANSWER IS OPTION A. THEORY APPLIED: Unconditional probability

Example: The chance of a fair coin flip being heads has an unconditional probability of 50% (1/2) regardless of how many coin flips preceded it, nor if some other event had occurred. Each toss of a coin is a perfect isolated thing. What it did in the past, will not affect the current toss. The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

Q1_A8

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example: The chance of a fair coin flip being heads has an unconditional probability of 50% (1/2) regardless of how many coin flips preceded it, nor if some other event had occurred. Each toss of a coin is a perfect isolated thing. What it did in the past, will not affect the current toss. The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

THEORY THAT CANNOT BE APPLIED: Standard deviation

Example: **A market researcher is analyzing the results of a recent customer survey that ranks a product from 1 to 10. He wants to have some measure of the reliability of the answers received in the survey in order to predict how a larger group of people might answer the same questions. Solution:**

Scores for the survey: 9, 7, 10, 8, 9, 7, 8, and 9.

Mean: 8.4.

Variance: 1.12

Standard Deviation =1.06

The standard deviation is 1.06, which is somewhat low. The researcher now knows that the results of the sample size are probably reliable.

Q1_A9

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Unconditional probability

Example: The chance of a fair coin flip being heads has an unconditional probability of 50% (1/2) regardless of how many coin flips preceded it, nor if some other event had occurred. Each toss of a coin is a perfect isolated thing. What it did in the past, will not affect the current toss. The chance is simply 1-in-2, or 50%, just like ANY toss of the coin. So each toss is an Independent Event.

THEORY THAT CANNOT BE APPLIED: Conditional Probability

Example: Given that you drew a red card, what's the probability that it's a four (p (four|red)) = 2/26= 1/13. So out of the 26 red cards (given a red card), there are two fours so 2/26=1/13. The fact that you drew a red card before affected the future outcome.

SECOND QUIZ

Q2_A1

THE RIGHT ANSWER IS OPTION B.

The **Central Limit Theorem** states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. **Thus, the expected number of days in which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent.** For the small hospital we have: 15 x 365= 5.475, [(1/2), (1/4) x (1/5475) = (0.5, 0.00004566) and for the small hospital we have: 45 x 365=16.425, [(1/2),(1/4) x (1/16425) = (0.5, 0.0000152).

Q2_A2

THE RIGHT ANSWER IS OPTION B.

The Central Limit Theorem states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. **Thus, the expected number of days in which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent.** For the small hospital we have: 15 x 365= 5.475, [(1/2), (1/4) x (1/5475) = (0.5, 0.00004566) and for the small hospital we have: 45 x 365=16.425, [(1/2),(1/4) x (1/16425) = (0.5, 0.0000152).

CONTRALILY, **Joint probability** is a statistical measure that could not be applied here because it calculates the likelihood of two events occurring together and at the same point in time. Joint probability is the probability of event Y occurring at the same time that event X occurs. If we suppose that the small hospital tomorrow will has ½ probability to exceed the 60% then we also suppose that the big hospital will has ¼ probability to exceed the 60% at the same day. For this reason their joint probability for tomorrow would be 12.5%.

Q2_A3

THE RIGHT ANSWER IS OPTION B.

The **Central Limit Theorem** states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. **Thus, the expected number of days in which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent.** For the small hospital we have: 15 x 365= 5.475, [(1/2), (1/4) x (1/5475) = (0.5, 0.00004566) and for the small hospital we have: 45 x 365=16.425, [(1/2),(1/4) x (1/16425) = (0.5, 0.0000152).

ALTERNATIVELY, in probability theory, the **law of large numbers** is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value and tends to become closer to the expected value as more trials are performed. In our case, if we observe the number of boys that are born each day in an extended period of time, the average number will be closer to the expected value of 45 for the big hospital and 15 for the small hospital. Obviously, this is not what has been requested.

Q2_A4

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 1: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

Example 2: A population of community college students includes inner city students (p = .33). What is the probability that a random sample of 45 students from the population will have from 20% to 40% inner city students? The probability your random sample will have 20 – 40% inner city students is 85.71%. **Alternatively, If a certain group of welfare recipients receives SNAP benefits of \$110 per week with a standard deviation of \$20. If a random sample of 25 people is taken, what is the probability their mean benefit will be greater than \$120 per week?** The probability is 99.38%.

Example 3: Approximately 60% (p = 0.6.) of all part-time college students in the United States are female What would you expect to see in terms of the behavior of a sample proportion of females, if random samples of size 100 were taken from the population of all part-time college students? For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

Q2_A5

THE RIGHT ANSWER IS OPTION B.

THEORY APPLIED: Central Limit theorem

Example 1: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED:

Joint Probability

Example 1: What is the joint probability of rolling the number five twice in a fair six-sided dice?

Event "A" = The probability of rolling a 5 in the first roll is 1/6 = 0.1666.

Event "B" = The probability of rolling a 5 in the second roll is 1/6 = 0.1666.

Therefore, the joint probability of event "A" and "B" is (1/6) x (1/6) = 0.02777 = 2.8%.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Central Limit theorem

Example 2: **A population of community college students includes inner city students (p = .33). What is the probability that a random sample of 45 students from the population will have from 20% to 40% inner city students?** The probability your random sample will have 20 – 40% inner city students is 85.71%. **Alternatively, If a certain group of welfare recipients receives SNAP benefits of $110 per week with a standard deviation of $20. If a random sample of 25 people is taken, what is the probability their mean benefit will be greater than $120 per week?** The probability is 99.38%.

THEORY THAT CANNOT BE APPLIED:

Joint Probability

Example 2: **What is the joint probability of getting a head followed by a tail in a coin toss?**

Event "A" = The probability of getting a head in the first coin toss is 1/2 = 0.5.

Event "B" = The probability of getting a tail in the second coin toss is 1/2 = 0.5.

Therefore, the joint probability of event "A" and "B" is (1/2) x (1/2) = 0.25 = 25%.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Central Limit theorem

Example 3: **Approximately 60% (p = 0.6.) of all part-time college students in the United States are female. What would you expect to see in terms of the behavior of a sample proportion of females, if random samples of size 100 were taken from the population of all part-time college students?** For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

THEORY THAT CANNOT BE APPLIED:

Joint Probability

Example 3: **What is the joint probability of drawing a number ten card that is black?**

Event "A" = The probability of drawing a 10 = 4/52 = 0.0769

Event "B" = The probability of drawing a black card = 26/52 = 0.50

Therefore, the joint probability of event "A" and "B" is (4/52) x (26/52) = 0.0385 = 3.9%.

Q2_A6

THE RIGHT ANSWER IS OPTION B.

THEORY APPLIED: Central Limit theorem

Example 1: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED:

Law of large numbers

Example 1: An economist may collect a simple random sample of 50 individuals in a town and use the average annual income of the individuals in the sample to estimate the average annual income of individuals in the entire town. If the economist finds that the average annual income of the individuals in the sample is $58,000, then her best guess for the true average annual income of individuals in the entire town will be $58,000.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Central Limit theorem

Example 2: **A population of community college students includes inner city students (p = .33). What is the probability that a random sample of 45 students from the population will have from 20% to 40% inner city students?** The probability your random sample will have 20 – 40% inner city students is 85.71%. **Alternatively, If a certain group of welfare recipients receives SNAP benefits of $110 per week with a standard deviation of $20. If a random sample of 25 people is taken, what is the probability their mean benefit will be greater than $120 per week?** The probability is 99.38%.

THEORY THAT CANNOT BE APPLIED:

Law of large numbers

Example 2: A biologist may measure the height of 30 randomly selected plants and then use the sample mean height to estimate the population mean height. If the biologist finds that the sample mean height of the 30 plants is 10.3 inches, then her best guess for the population mean height will also be 10.3 inches.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Central Limit theorem

Example 3: **Approximately 60% (p = 0.6.) of all part-time college students in the United States are female. What would you expect to see in terms of the behavior of a sample proportion of females, if random samples of size 100 were taken from the population of all part-time college students?** For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

THEORY THAT CANNOT BE APPLIED:

Law of large numbers

Example 3: The HR department of some company may randomly select 50 employees to take a survey that assesses their overall satisfaction on a scale of 1 to 10. If it's found that the average satisfaction among employees in the survey is 8.5 then the best guess for the average satisfaction rating of all employees at the company is also 8.5.

Q2_A7

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

Q2_A8

THE RIGHT ANSWER IS OPTION B.

THEORY APPLIED: Central Limit theorem

Example: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED: Joint Probability

Example: **What is the joint probability of getting a head followed by a tail in a coin toss?**

Event "A" = The probability of getting a head in the first coin toss is 1/2 = 0.5.

Event "B" = The probability of getting a tail in the second coin toss is 1/2 = 0.5.

Therefore, the joint probability of event "A" and "B" is P(1/2) x P(1/2) = 0.25 = 25%.

Q2_A9

THE RIGHT ANSWER IS OPTION B.

THEORY APPLIED: Central Limit theorem

Example: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED:

Law of Large Numbers

Example: The HR department of some company may randomly select 50 employees to take a survey that assesses their overall satisfaction on a scale of 1 to 10. If it's found that the average satisfaction among employees in the survey is 8.5 then the best guess for the average satisfaction rating of all employees at the company is also 8.5.

THIRD QUIZ

Q3_A1

THE RIGHT ANSWER IS OPTION B.

The **law of likelihood**, that is, the notion that the extent to which the evidence supports one parameter value or hypothesis against another is indicated by the ratio of their likelihoods, their likelihood ratio. **By applying this rule in our case and by assuming equal prior probabilities, the first person's posterior would be 8:1, while the second person's would be 16:1. We can translate these into probabilities of 8/9 and 16/17. With that being said, the second person should be more confident.**

Q3_A2

THE RIGHT ANSWER IS OPTION B.

The law of likelihood, that is, the notion that the extent to which the evidence supports one parameter value or hypothesis against another is indicated by the ratio of their likelihoods, their likelihood ratio. **By applying this rule in our case and by assuming equal prior probabilities, the first person's posterior would be 8:1, while the second person's would be 16:1. We can translate these into probabilities of 8/9 and 16/17. With that being said, the second person should be more confident.**

ON CONTRAST, **correlation coefficient** is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate random variable with a known distribution. Obviously, this is not the right strategy of resolving this arithmetical problem because we don't search for the statistical correlation between the first and the second case scenario.

Q3_A3

THE RIGHT ANSWER IS OPTION B.

The law of likelihood, that is, the notion that the extent to which the evidence supports one parameter value or hypothesis against another is indicated by the ratio of their likelihoods, their likelihood ratio. **By applying this rule in our case and by assuming equal prior probabilities, the first person's posterior would be 8:1, while the second person's would be 16:1. We can translate these into probabilities of 8/9 and 16/17. With that being said, the second person should be more confident.**

ON THE FLIP SIDE, the probability of an event is the number of favorable outcomes divided by the total number of outcomes possible. In our example we could not apply this rule because we do not have enough data in the hypothesis stage and thus, we have to use our given data to create a hypothesis, for example which urn contains more red balls. Having said that, **in likelihood theorem**, the data are a given and the hypotheses vary.

Q3_A4

THE RIGHT ANSWER IS OPTION A. THEORY APPLIED: Likelihood theorem

Example 1: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

Example 2: **Assume that the chances of a person having a skin disease are 40%. Assuming that skin creams and drinking enough water reduces the risk of skin disease by 30% and prescription of a certain drug reduces its chance by 20%. At a time, a patient can choose any one of the two options with equal probabilities. It is given that after picking one of the options, the patient selected at random has the skin disease. Find the probability that the patient picked the option of skin screams and drinking enough water.** **A)** The patient uses skin creams and drinks enough water = ½, and B) the patient uses the drug = ½. Using the probabilities known to us, we have: The selected patient has the skin disease for the case A = 0.4 × (1-0.3) = 0.28 and for the case B = 0.4 × (1-0.2) = 0.32. Using Bayes Theorem, the probability that the selected patient uses skin creams and drinks enough water is (0.28 × 0.5)/(0.28 × 0.5 + 0.32 × 0.5) = 0.14/(0.14 + 0.16) = 0.47. Thus, the probability that the patient picked the first option is 0.47.

Example 3: **A man is known to speak the truth 3/4 times. He draws a card and reports it is king. Find the probability that it is actually a king**. The probability that king is drawn = ¼. The probability that the king is not drawn = ¾. The probability that the man says the truth that king is drawn when actually king is drawn = ¾. The probability that the man lies that king is drawn when actually king is drawn = ¼. By using Bayes Theorem, the probability that it is actually a king = [1/4 × 3/4] ÷ [(1/4 × 3/4) + (1/4 × 3/4)] = 3/16 ÷12/16 = 3/16 × 16/12 =1/2 = 0.5. Thus, the probability that the drawn card is actually a king = 0.5.

Q3_A5

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Likelihood theorem

Example 1: Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I. The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED: Correlation Coefficient

Example 1: A survey was conducted in your city. Given is the following sample data containing a person's age and their corresponding income. Find out whether the increase in age has an effect on income. The results showed that with the increase in age, a person's income increases as well, since the Pearson correlation coefficient between age and income is very close to 1.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Likelihood theorem

Example 2: Assume that the chances of a person having a skin disease are 40%. Assuming that skin creams and drinking enough water reduces the risk of skin disease by 30% and prescription of a certain drug reduces its chance by 20%. At a time, a patient can choose any one of the two options with equal probabilities. It is given that after picking one of the options, the patient selected at random has the skin disease. Find the probability that the patient picked the option of skin screams and drinking enough water. A) The patient uses skin creams and drinks enough water = ½, and B) the patient uses the drug = ½. Using the probabilities known to us, we have: The selected patient has the skin disease for the case A = 0.4 × (1-0.3) = 0.28 and for the case B = 0.4 × (1-0.2) = 0.32. Using Bayes Theorem, the probability that the selected patient uses skin creams and drinks enough water is (0.28 × 0.5)/(0.28 × 0.5 + 0.32 × 0.5) = 0.14/(0.14 + 0.16) = 0.47. Thus, the probability that the patient picked the first option is 0.47.

THEORY THAT CANNOT BE APPLIED:

Correlation Coefficient

Example 2: Suppose that the prices of coffee and computers are observed and found to have a correlation of +.0008. This means that there is no correlation, or relationship, between the two variables.

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Likelihood theorem

Example 3: **A man is known to speak the truth 3/4 times. He draws a card and reports it is king. Find the probability that it is actually a king**. The probability that king is drawn = ¼. The probability that the king is not drawn = ¾. The probability that the man says the truth that king is drawn when actually king is drawn = ¾. The probability that the man lies that king is drawn when actually king is drawn = ¼. By using Bayes Theorem, the probability that it is actually a king = [1/4 × 3/4] ÷ [(1/4 × 3/4) + (1/4 × 3/4)] = 3/16 ÷12/16 = 3/16 × 16/12 =1/2 = 0.5. Thus, the probability that the drawn card is actually a king = 0.5.

THEORY THAT CANNOT BE APPLIED:

Correlation Coefficient

Example 3: Suppose the value of oil prices is directly related to the prices of airplane tickets, with a correlation coefficient of +0.95. The relationship between oil prices and airfares has a very strong positive correlation since the value is close to +1. So, if the price of oil decreases, airfares also decrease, and if the price of oil increases, so do the prices of airplane tickets.

Q3_A7

THE RIGHT ANSWER IS OPTION A. THEORY APPLIED: Likelihood theorem

Example: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

Q3_A8

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Likelihood theorem

Example: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED: Correlation Coefficient

Example: Suppose the value of oil prices is directly related to the prices of airplane tickets, with a correlation coefficient of +0.95. The relationship between oil prices and airfares has a very strong positive correlation since the value is close to +1. So, if the price of oil decreases, airfares also decrease, and if the price of oil increases, so do the prices of airplane tickets.

Q3_A9

THE RIGHT ANSWER IS OPTION A.

THEORY APPLIED: Likelihood theorem

Example: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED:

Probability Theory

Example: Suppose the probability of obtaining a number 4 on rolling a fair dice needs to be established. The number of favorable outcomes is 1. The possible outcomes of the dice are {1, 2, 3, 4, 5, 6}. This implies that there are a total of 6 outcomes. Thus, the probability of obtaining 4 on a dice roll, using probability theory, can be computed as 1 / 6 = 0.167.

FORTH QUIZ

Q4_A1

THE RIGHT ANSWER IS A-A OR B-B.

The **expected utility** of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. **Thus, between both cases your answers should be equivalent because you've lost $10, either in the form of a $10 bill or a $10 ticket.**

Q4_A2

THE RIGHT ANSWER IS A-A OR B-B.

The **expected** **utility** of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. **Thus, between both cases your answers should be equivalent because you've lost $10, either in the form of a $10 bill or a $10 ticket.**

DIFFIRENTLY, a sequence of numbers that identifies items as belonging to mutually exclusive categories can be represented through a **Categorical Scale.** When the number sequence has a meaningful order a categorical scale is more precisely called an ordinal scale; when it is devoid of such meaningful order it is known as a nominal scale. Here, we are not interested in categorizing our data but to find the most valuable option.

Q4_A3

THE RIGHT ANSWER IS A-A OR B-B.

The expected utility of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. **Thus, between both cases your answers should be equivalent because you've lost $10, either in the form of a $10 bill or a $10 ticket.**

The prospect theory says that investors value gains and losses differently, placing more weight on perceived gains versus perceived losses. An investor presented with a choice, both equal, will choose the one presented in terms of potential gains. On the other hand, expected Utility theory assumes individuals will choose the outcome which gives maximum utility given the probability of outcomes. For this reason, Prospect theory is not the chosen one because, it allows for the fact that individuals may choose a decision which doesn't necessarily maximize utility because they place other considerations above utility.

Q4_A4

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket).
THEORY APPLIED: Expected Utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

Example 2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

Example 3: **Suppose: A lottery ticket costs $20. The probability of winning the $2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = $10.** The expected value of owning a lottery ticket is $10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs $20, it seems an illogical decision to buy – because the expected value of buying a ticket is $10 – a smaller figure than the cost of purchase $20.

Q4_A5

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 1: **After rendering service to customers, businesses like to get feedback from customers regarding their service to improve. For example:** Kindly rate your customer service experience with us:

Very poor

Poor

Neutral

Good

Very good

The above is an example of an ordinal data collection process. The responses have a specific order to them, listed in ascending order.

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example 2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 2: **The level of education of a respondent may be requested for when filling forms for job applications, admission, training etc. This is used to assess their qualification for a specific role. Consider the example below:** What is your highest level of education?

School SAT

High School

BSc.

MSc.

PhD

This is also a closed-ended nominal data example.

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example 3: **Suppose: A lottery ticket costs $20. The probability of winning the $2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = $10.** The expected value of owning a lottery ticket is $10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs $20, it seems an illogical decision to buy – because the expected value of buying a ticket is $10 – a smaller figure than the cost of purchase $20.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 3: **Respondents are asked for their gender when filling out a biodata. This is mostly categorized as male or female, but may also be nonbinary. For example:**

What is your gender?

Male

Female

This is a binary and closed-ended nominal data example.

Q4_A6

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example 1: There are two options. Option 1 gives the recipient $50. Option 2 gives the recipient $100 but then takes away $50. The end outcome is that both provide a net gain of $50. Yet option 1 is preferred as it provides the recipient with a perceived net gain that avoids the pain of a loss.

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example 2: One advisor may present a mutual fund to a potential investor, claiming that it has had good returns over the last decade but also that it has been declining in the last two years. Another advisor may present the same fund and report that it has had an average return of 15% over the past three years. Investors would be much more likely to invest with the second advisor, as the fund was presented only in terms of gains.

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example 3: **Suppose: A lottery ticket costs $20. The probability of winning the $2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = $10.** The expected value of owning a lottery ticket is $10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs $20, it seems an illogical decision to buy – because the expected value of buying a ticket is $10 – a smaller figure than the cost of purchase $20.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example 3: Assume that a lottery provides two options, A and B. Option A provides a guaranteed win of $100 while option B provides the possibility of winning $200, with a 70% chance of winning and 30% chance of losing. Most people will choose option A since it provides a guaranteed win, even though it offers a lower return compared to B.

Q4_A7

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket).
THEORY APPLIED: Expected Utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

Q4_A8

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example: **After rendering service to customers, businesses like to get feedback from customers regarding their service to improve. For example:** Kindly rate your customer service experience with us:

Very poor

Poor

Neutral

Good

Very good

The above is an example of an ordinal data collection process. The responses have a specific order to them, listed in ascending order.

Q4_A9

THE RIGHT ANSWER IS A-A OR B-B (you've lost $10, either in the form of a $10 bill or a $10 ticket). THEORY APPLIED: Expected Utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED:

Prospect theory

Example: There are two options. Option 1 gives the recipient $50. Option 2 gives the recipient $100 but then takes away $50. The end outcome is that both provide a net gain of $50. Yet option 1 is preferred as it provides the recipient with a perceived net gain that avoids the pain of a loss.

FIFTH QUIZ

Q5_A1

THE RIGHT ANSWER IS OPTION B.

The **expected utility** of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. **In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B.**

Sure loss (100%) → -$900 → Expected Value= -$900
[Possible loss (90%) → -$1000
Neutral (10%) → $0] → Expected Value = (-$1000x0.9) + ($0x0.1) = -$900

Q5_A2

THE RIGHT ANSWER IS OPTION B.

The **expected** **utility** of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. **In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B.**

Sure loss (100%) → -$900 → Expected Value= -$900
[Possible loss (90%) → -$1000
Neutral (10%) → $0] → Expected Value = (-$1000x0.9) + ($0x0.1) = -$900

ON CONTRAST, a sequence of numbers that identifies items as belonging to mutually exclusive categories. When the number sequence has a meaningful order a **categorical** **scale** is more precisely called an ordinal scale; when it is devoid of such meaningful order it is known as a nominal scale. Here, we are not interested in categorizing our data but to find the most valuable option.

Q5_A3

THE RIGHT ANSWER IS OPTION B.

The expected utility of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. **In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B.**

Sure loss (100%)→ -\$900 → (Expected Value= -\$900)
[Possible loss (90%) → -\$1000
Neutral (10%) → \$0] → Expected Value = (-\$1000x0.9) + (\$0x0.1) = -\$900

The prospect theory says that investors value gains and losses differently, placing more weight on perceived gains versus perceived losses. An investor presented with a choice, both equal, will choose the one presented in terms of potential gains. On the other half, expected utility theory assumes individuals will choose the outcome which gives maximum utility given the probability of outcomes. For this reason, prospect theory is not the chosen one because, it allows for the fact that individuals may choose a decision which doesn't necessarily maximize utility because they place other considerations above utility.

Q5_A4

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

Example 2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

Example 3: **Suppose: A lottery ticket costs $20. The probability of winning the $2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = $10.** The expected value of owning a lottery ticket is $10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs $20, it seems an illogical decision to buy – because the expected value of buying a ticket is $10 – a smaller figure than the cost of purchase $20.

Q5_A5

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 1: **After rendering service to customers, businesses like to get feedback from customers regarding their service to improve. For example:** Kindly rate your customer service experience with us:

Very poor

Poor

Neutral

Good

Very good

The above is an example of an ordinal data collection process. The responses have a specific order to them, listed in ascending order.

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 2: **The level of education of a respondent may be requested for when filling forms for job applications, admission, training etc. This is used to assess their qualification for a specific role. Consider the example below:** What is your highest level of education?

School SAT

High School

BSc.

MSc.

PhD

This is also a closed-ended nominal data example.

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 3: **Suppose: A lottery ticket costs $20. The probability of winning the $2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = $10.** The expected value of owning a lottery ticket is $10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs $20, it seems an illogical decision to buy – because the expected value of buying a ticket is $10 – a smaller figure than the cost of purchase $20.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 3: **Respondents are asked for their gender when filling out a biodata. This is mostly categorized as male or female, but may also be nonbinary. For example:**

What is your gender?

Male

Female

This is a binary and closed-ended nominal data example.

Q5_A6

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example 1: There are two options. Option 1 gives the recipient \$50. Option 2 gives the recipient \$100 but then takes away \$50. The end outcome is that both provide a net gain of \$50. Yet option 1 is preferred as it provides the recipient with a perceived net gain that avoids the pain of a loss.

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example 2: One advisor may present a mutual fund to a potential investor, claiming that it has had good returns over the last decade but also that it has been declining in the last two years. Another advisor may present the same fund and report that it has had an average return of 15% over the past three years. Investors would be much more likely to invest with the second advisor, as the fund was presented only in terms of gains.

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example 3: **Suppose: A lottery ticket costs $20. The probability of winning the $2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = $10.** The expected value of owning a lottery ticket is $10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs $20, it seems an illogical decision to buy – because the expected value of buying a ticket is $10 – a smaller figure than the cost of purchase $20.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example 3: Assume that a lottery provides two options, A and B. Option A provides a guaranteed win of $100 while option B provides the possibility of winning $200, with a 70% chance of winning and 30% chance of losing. Most people will choose option A since it provides a guaranteed win, even though it offers a lower return compared to B.

Q5_A7

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

Q5_A8

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example: **After rendering service to customers, businesses like to get feedback from customers regarding their service to improve. For example:** Kindly rate your customer service experience with us:

Very poor

Poor

Neutral

Good

Very good

The above is an example of an ordinal data collection process. The responses have a specific order to them, listed in ascending order.

Q5_A9

THE RIGHT ANSWER IS OPTION B. In our situation, the expected utility is the same for both cases but, there is a possibility of gaining a larger reward by choosing option B. THEORY APPLIED: Expected Utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THEORY THAT CANNOT BE APPLIED: Prospect Theory

Example: There are two options. Option 1 gives the recipient $50. Option 2 gives the recipient $100 but then takes away $50. The end outcome is that both provide a net gain of $50. Yet option 1 is preferred as it provides the recipient with a perceived net gain that avoids the pain of a loss.

SIXTH QUIZ

Q6_A1

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases.

Expected value is a concept used in situations in which it is desirable to establish the value of different options with uncertain outcomes. The expected value of an action is the sum of the value of each potential outcome multiplied by the probability of that outcome occurring. **In our case, the expected value is the same for both cases**.

Sure loss (100%) → -$900 → Expected Value= -$900

[Possible loss (90%) → -$1000

Neutral (10%) → $0] → Expected Value = (-$1000x0.9) + ($0x0.1) = -$900

Q6_A2

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases.

Expected value is a concept used in situations in which it is desirable to establish the value of different options with uncertain outcomes. The expected value of an action is the sum of the value of each potential outcome multiplied by the probability of that outcome occurring. **In our case, the expected value is the same for both cases**.

Sure loss (100%) → -$900 → (Expected Value= -$900)

[Possible loss (90%) → -$1000

Neutral (10%) → $0] → Expected Value = (-$1000x0.9) + ($0x0.1) = -$900

ON THE CONTRARY, a sequence of numbers that identifies items as belonging to mutually exclusive categories. When the number sequence has a meaningful order a categorical scale is more precisely called an ordinal scale; when it is devoid of such meaningful order it is known as a nominal scale. Here, we are not interested in categorizing our data but to find the most valuable option.

Q6_A3

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases.

Expected value is a concept used in situations in which it is desirable to establish the value of different options with uncertain outcomes. The expected value of an action is the sum of the value of each potential outcome multiplied by the probability of that outcome occurring. **In our case, the expected value is the same for both cases**.

Sure loss (100%) → -$900 → (Expected Value= -$900)

[Possible loss (90%) → -$1000

Neutral (10%) → $0] → Expected Value = (-$1000x0.9) + ($0x0.1) = -$900

The expected utility of an action to an agent, is calculated by multiplying the value to the agent of each possible outcome of the action by the probability of that outcome occurring and then summing those numbers. The concept of expected utility is used to elucidate decisions made under conditions of risk. According to standard decision theory, when comparing alternative courses of action, one should choose the action that has the greatest expected utility. On the other side, expected value is a concept used in situations in which it is desirable to establish the value of different options with uncertain outcomes. The expected value of an action is the sum of the value of each potential outcome multiplied by the probability of that outcome occurring. In our case, we are interested for the expected value of each option and not for the best outcome.

Q6_A4

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 1: If you toss a coin ten times, the probability of getting a heads in each trial is 1/2 so the expected value (the number of heads you can expect to get in 10 coin tosses) is: 0.5 * 10 = 5.

Example 2: You buy one $10 raffle ticket for a new car valued at $15,000. Two thousand tickets are sold. What is the expected value of your gain? If we do the right calculations the expected value would be: $7.495 + -$9.995 = -$2.5.

Example 3: Flipping a coin! You have two outcomes: heads or tails. The probabilities of both are 50%. Let's say that you play 100 rounds with your friend. You risk $1 in each round. If it's tails, you double your money, if it's heads, you lose your money. Using the expected value formula, we see the expected revenue from this game is $1. And you have to invest $1 in each round. So your expected value of your profit is $0. In other words, if you play this game long enough, you won't lose or win any money.

Q6_A5

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 1: If you toss a coin ten times, the probability of getting a heads in each trial is 1/2 so the expected value (the number of heads you can expect to get in 10 coin tosses) is: 0.5 * 10 = 5

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 1: After rendering service to customers, businesses like to get feedback from customers regarding their service to improve. For example: Kindly rate your customer service experience with us:

Very poor

Poor

Neutral

Good

Very good

The above is an example of an ordinal data collection process. The responses have a specific order to them, listed in ascending order.

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 2: You buy one $10 raffle ticket for a new car valued at $15,000. Two thousand tickets are sold. What is the expected value of your gain? If we do the right calculations the expected value would be: $7.495 + -$9.995 = -$2.5.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 2: **The level of education of a respondent may be requested for when filling forms for job applications, admission, training etc. This is used to assess their qualification for a specific role. Consider the example below:** What is your highest level of education?

School SAT

High School

BSc.

MSc.

PhD

This is also a closed-ended nominal data example.

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 3: Flipping a coin! You have two outcomes: heads or tails. The probabilities of both are 50%. Let's say that you play 100 rounds with your friend. You risk $1 in each round. If it's tails, you double your money, if it's heads, you lose your money. Using the expected value formula, we see the expected revenue from this game is $1. And you have to invest $1 in each round. So your expected value of your profit is $0. In other words, if you play this game long enough, you won't lose or win any money.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example 3: Respondents are asked for their gender when filling out a biodata. This is mostly categorized as male or female, but may also be nonbinary. For example:

What is your gender?

Male

Female

This is a binary and closed-ended nominal data example.

Q6_A6

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 1: If you toss a coin ten times, the probability of getting a heads in each trial is 1/2 so the expected value (the number of heads you can expect to get in 10 coin tosses) is: 0.5 * 10 = 5

THEORY THAT CANNOT BE APPLIED: Expected utility

Example 1: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 2: You buy one $10 raffle ticket for a new car valued at $15,000. Two thousand tickets are sold. What is the expected value of your gain? If we do the right calculations the expected value would be: $7.495 + -$9.995 = -$2.5.

THEORY THAT CANNOT BE APPLIED: Expected utility

Example 2: **Suppose we decide to study for three years to try and gain an economic degree. A good degree is likely to lead to a higher paying job but there is no guarantee. We may fail the degree or the jobs market may turn against a surplus of graduates.** Therefore, we may estimate we have a 0.7 chance of gaining an extra $250,000 earnings in our lifetime. In this case, the expected utility of an economics degree is $175,000.

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example 3: Flipping a coin! You have two outcomes: heads or tails. The probabilities of both are 50%. Let's say that you play 100 rounds with your friend. You risk \$1 in each round. If it's tails, you double your money, if it's heads, you lose your money. Using the expected value formula, we see the expected revenue from this game is \$1. And you have to invest \$1 in each round. So your expected value of your profit is \$0. In other words, if you play this game long enough, you won't lose or win any money.

THEORY THAT CANNOT BE APPLIED: Expected utility

Example 3: **Suppose: A lottery ticket costs \$20. The probability of winning the \$2000 prize is 0.5%. The likely value from having a lottery ticket will be the outcome x probability of the event occurring. Therefore, expected value = 0.005 x 2000 = \$10.** The expected value of owning a lottery ticket is \$10. With an infinite number of events, on average, this is the likely payout. Of course, we may be lucky or maybe unlucky if we play only once. Since the ticket costs \$20, it seems an illogical decision to buy – because the expected value of buying a ticket is \$10 – a smaller figure than the cost of purchase \$20.

Q6_A7

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example: If you toss a coin ten times, the probability of getting a heads in each trial is 1/2 so the expected value (the number of heads you can expect to get in 10 coin tosses) is: 0.5 * 10 = 5.

Q6_A8

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example: If you toss a coin ten times, the probability of getting a heads in each trial is 1/2 so the expected value (the number of heads you can expect to get in 10 coin tosses) is: 0.5 * 10 = 5.

THEORY THAT CANNOT BE APPLIED: Categorical Scale of measurement

Example: **After rendering service to customers, businesses like to get feedback from customers regarding their service to improve. For example:** Kindly rate your customer service experience with us:

Very poor

Poor

Neutral

Good

Very good

The above is an example of an ordinal data collection process. The responses have a specific order to them, listed in ascending order.

Q6_A9

THE RIGHT ANSWER IS OPTION C. The expected value is the same for both cases. THEORY APPLIED: Expected Value

Example: If you toss a coin ten times, the probability of getting a heads in each trial is 1/2 so the expected value (the number of heads you can expect to get in 10 coin tosses) is: 0.5 * 10 = 5.

THEORY THAT CANNOT BE APPLIED: Expected utility

Example: Purchasing a lottery ticket represents two possible outcomes for the buyer. They could end up losing the amount they invested in buying the ticket, or they could end up making a smart profit by winning either a portion of the entire lottery. Assigning probability values to the costs involved (in this case, the nominal purchase price of a lottery ticket), it is not difficult to see that the expected utility to be gained from purchasing a lottery ticket is greater than not buying it.

SEVENTH QUIZ

Q7_A1

THE RIGHT ANSWER IS OPTION B.

The **Central Limit Theorem** states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. **Here, the variance for p = .45 exceeds that for p = .65 therefore, it is more, it is more likely that the class belongs to program B.**

Q7_A2

THE RIGHT ANSWER IS OPTION B.

The **Central Limit Theorem** states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. **Here, the variance for p = .45 exceeds that for p = .65 therefore, it is more, it is more likely that the class belongs to program B.**

UNDER OTHER CONDITIONS, **Joint probability** is a statistical measure that calculates the likelihood of two events occurring together and at the same point in time. Joint probability is the probability of event Y occurring at the same time that event X occurs. If we suppose that the A program has 65/100 probability to be that class and we also suppose that the B program has 45/100 probability to be that class, their joint probability would be 29,3%. Thus, we could not apply this theory here.

Q7_A3

THE RIGHT ANSWER IS OPTION B.

The Central Limit Theorem states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger, no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. **Here, the variance for p = .45 exceeds that for p = .65 therefore, it is more, it is more likely that the class belongs to program B.**

ON THE OTHER HAND, in probability theory, the law of large numbers is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value and tends to become closer to the expected value as more trials are performed. In our case, there is no point to measure the variance of boys in a large number of trials for each class because the size remains the same.

Q7_A4

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 1: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

Example 2: A population of community college students includes inner city students (p = .33). What is the probability that a random sample of 45 students from the population will have from 20% to 40% inner city students? The probability your random sample will have 20 – 40% inner city students is 85.71%. **Alternatively, If a certain group of welfare recipients receives SNAP benefits of $110 per week with a standard deviation of $20. If a random sample of 25 people is taken, what is the probability their mean benefit will be greater than $120 per week?** The probability is 99.38%.

Example 3: Approximately 60% of all part-time college students in the United States are female. (In other words, the population proportion of females among part-time college students is p = 0.6.) What would you expect to see in terms of the behavior of a sample proportion of females (p-hat) if random samples of size 100 were taken from the population of all part-time college students? For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

Q7_A5

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 1: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED: Joint probability

Example 1: What is the joint probability of rolling the number five twice in a fair six-sided dice?

Event "A" = The probability of rolling a 5 in the first roll is 1/6 = 0.1666.

Event "B" = The probability of rolling a 5 in the second roll is 1/6 = 0.1666.

Therefore, the joint probability of event "A" and "B" is P(1/6) x P(1/6) = 0.02777 = 2.8%.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 2: **A population of community college students includes inner city students (p = .33). What is the probability that a random sample of 45 students from the population will have from 20% to 40% inner city students?** The probability your random sample will have 20 – 40% inner city students is 85.71%. **Alternatively, If a certain group of welfare recipients receives SNAP benefits of $110 per week with a standard deviation of $20. If a random sample of 25 people is taken, what is the probability their mean benefit will be greater than $120 per week?** The probability is 99.38%.

THEORY THAT CANNOT BE APPLIED: Joint probability

Example 2: **What is the joint probability of getting a head followed by a tail in a coin toss?**

Event "A" = The probability of getting a head in the first coin toss is 1/2 = 0.5.

Event "B" = The probability of getting a tail in the second coin toss is 1/2 = 0.5.

Therefore, the joint probability of event "A" and "B" is P(1/2) x P(1/2) = 0.25 = 25%.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 3: Approximately 60% of all part-time college students in the United States are female. (In other words, the population proportion of females among part-time college students is p = 0.6.) What would you expect to see in terms of the behavior of a sample proportion of females (p-hat) if random samples of size 100 were taken from the population of all part-time college students? For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

THEORY THAT CANNOT BE APPLIED: Joint probability

Example 3: What is the joint probability of drawing a number ten card that is black?

Event "A" = The probability of drawing a 10 = 4/52 = 0.0769

Event "B" = The probability of drawing a black card = 26/52 = 0.50

Therefore, the joint probability of event "A" and "B" is P(4/52) x P(26/52) = 0.0385 = 3.9%.

Q7_A6

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 1: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED: Law of large numbers

Example 1: An economist may collect a simple random sample of 50 individuals in a town and use the average annual income of the individuals in the sample to estimate the average annual income of individuals in the entire town. If the economist finds that the average annual income of the individuals in the sample is $58,000, then her best guess for the true average annual income of individuals in the entire town will be $58,000.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 2: **A population of community college students includes inner city students (p = .33). What is the probability that a random sample of 45 students from the population will have from 20% to 40% inner city students?** The probability your random sample will have 20 – 40% inner city students is 85.71%. **Alternatively, If a certain group of welfare recipients receives SNAP benefits of $110 per week with a standard deviation of $20. If a random sample of 25 people is taken, what is the probability their mean benefit will be greater than $120 per week?** The probability is 99.38%.

THEORY THAT CANNOT BE APPLIED: Law of large numbers

Example 2: A biologist may measure the height of 30 randomly selected plants and then use the sample mean height to estimate the population mean height. If the biologist finds that the sample mean height of the 30 plants is 10.3 inches, then her best guess for the population mean height will also be 10.3 inches.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example 3: **Approximately 60% of all part-time college students in the United States are female. (In other words, the population proportion of females among part-time college students is p = 0.6.) What would you expect to see in terms of the behavior of a sample proportion of females (p-hat) if random samples of size 100 were taken from the population of all part-time college students?** For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

THEORY THAT CANNOT BE APPLIED: Law of large numbers

Example 3: The HR department of some company may randomly select 50 employees to take a survey that assesses their overall satisfaction on a scale of 1 to 10. If it's found that the average satisfaction among employees in the survey is 8.5 then the best guess for the average satisfaction rating of all employees at the company is also 8.5.

Q7_A7

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

Q7_A8

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED: Joint probability

Example: **What is the joint probability of getting a head followed by a tail in a coin toss?**

Event "A" = The probability of getting a head in the first coin toss is 1/2 = 0.5.

Event "B" = The probability of getting a tail in the second coin toss is 1/2 = 0.5.

Therefore, the joint probability of event "A" and "B" is P(1/2) x P(1/2) = 0.25 = 25%.

Q7_A9

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Central Limit theorem

Example: Think about which is more likely: getting more than 60 percent heads in three flips of coin or getting more than 60 percent heads in 3,000 flips. Half of the time, three flips will produce more than 60% heads. However, 10 flips will only produce more than 60% heads about 17% of the time. Three thousands flips will produce more than 60% heads only .000001% of the time (odds of one in a million).

THEORY THAT CANNOT BE APPLIED: Law of large numbers

Example: The HR department of some company may randomly select 50 employees to take a survey that assesses their overall satisfaction on a scale of 1 to 10. If it's found that the average satisfaction among employees in the survey is 8.5 then the best guess for the average satisfaction rating of all employees at the company is also 8.5.

EIGHTH QUIZ

Q8_A1

THE RIGHT ANSWER IS OPTION B.

The **law of likelihood**, that is, the notion that the extent to which the evidence supports one parameter value or hypothesis against another is indicated by the ratio of their likelihoods, their likelihood ratio. **Applying this theorem here, we have $[20! / (4!)^3 5!\ 3!] / 5^{20}$ =0.25% for the first order and, $[20! / (4!)^5]/5^{20}$ = 0.32% for the second order. Therefore, the uniform distribution of marbles is more probable than the nonuniform distribution.**

Q8_A2

THE RIGHT ANSWER IS OPTION B.

The law of likelihood, that is, the notion that the extent to which the evidence supports one parameter value or hypothesis against another is indicated by the ratio of their likelihoods, their likelihood ratio. **Applying this theorem here, we have [20! / $(4!)^3$5! 3!] / 5^{20} =0.25% for the first order and, [20! / $(4!)^5$]/5^{20} = 0.32% for the second order. Therefore, the uniform distribution of marbles is more probable than the nonuniform distribution.**

IN DIFFIRENT CIRCUMSTANCES, a **correlation coefficient** is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate random variable with a known distribution. Obviously, this is not the right strategy of resolving this arithmetical problem because we don't search for the statistical correlation between the first and the second case scenario.

Q8_A3

THE RIGHT ANSWER IS OPTION B.

The law of likelihood, that is, the notion that the extent to which the evidence supports one parameter value or hypothesis against another is indicated by the ratio of their likelihoods, their likelihood ratio. **Applying this theorem here, we have [20! / $(4!)^3 5!$ 3!] / 5^{20} =0.25% for the first order and, [20! / $(4!)^5$]/5^{20} = 0.32% for the second order. Therefore, the uniform distribution of marbles is more probable than the nonuniform distribution.**

The probability of an event is the number of favorable outcomes divided by the total number of outcomes possible. In our example we could not apply this rule because we do not have enough data in the hypothesis stage and thus, we have to use our given data to create a hypothesis, for example which order distribution will appear the most. In law of likelihood theorem, the data are a given and the hypotheses vary.

Q8_A4

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 1: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

Example 2: **Assume that the chances of a person having a skin disease are 40%. Assuming that skin creams and drinking enough water reduces the risk of skin disease by 30% and prescription of a certain drug reduces its chance by 20%. At a time, a patient can choose any one of the two options with equal probabilities. It is given that after picking one of the options, the patient selected at random has the skin disease. Find the probability that the patient picked the option of skin screams and drinking enough water.** A) The patient uses skin creams and drinks enough water = ½, and B) the patient uses the drug = ½. Using the probabilities known to us, we have: The selected patient has the skin disease for the case A = 0.4 × (1-0.3) = 0.28 and for the case B = 0.4 × (1-0.2) = 0.32. Using Bayes Theorem, the probability that the selected patient uses skin creams and drinks enough water is (0.28 × 0.5)/(0.28 × 0.5 + 0.32 × 0.5) = 0.14/(0.14 + 0.16) = 0.47. Thus, the probability that the patient picked the first option is 0.47.

Example 3: **A man is known to speak the truth 3/4 times. He draws a card and reports it is king. Find the probability that it is actually a king**. The probability that king is drawn = ¼. The probability that the king is not drawn = ¾. The probability that the man says the truth that king is drawn when actually king is drawn = ¾. The probability that the man lies that king is drawn when actually king is drawn = ¼. By using Bayes Theorem, the probability that it is actually a king = [1/4 × 3/4] ÷ [(1/4 × 3/4) + (1/4 × 3/4)] = 3/16 ÷12/16 = 3/16 × 16/12 =1/2 = 0.5. Thus, the probability that the drawn card is actually a king = 0.5.

Q8_A5

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 1: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED: Correlation Coefficient

Example 1: A survey was conducted in your city. Given is the following sample data containing a person's age and their corresponding income. Find out whether the increase in age has an effect on income. The results showed that with the increase in age a person's income increases as well, since the Pearson correlation coefficient between age and income is very close to 1.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 2: Assume that the chances of a person having a skin disease are 40%. Assuming that skin creams and drinking enough water reduces the risk of skin disease by 30% and prescription of a certain drug reduces its chance by 20%. At a time, a patient can choose any one of the two options with equal probabilities. It is given that after picking one of the options, the patient selected at random has the skin disease. Find the probability that the patient picked the option of skin screams and drinking enough water. A) The patient uses skin creams and drinks enough water = ½, and B) the patient uses the drug = ½. Using the probabilities known to us, we have: The selected patient has the skin disease for the case A = 0.4 × (1-0.3) = 0.28 and for the case B = 0.4 × (1-0.2) = 0.32. Using Bayes Theorem, the probability that the selected patient uses skin creams and drinks enough water is (0.28 × 0.5)/(0.28 × 0.5 + 0.32 × 0.5) = 0.14/(0.14 + 0.16) = 0.47. Thus, the probability that the patient picked the first option is 0.47.

THEORY THAT CANNOT BE APPLIED: Correlation Coefficient

Example 2: Suppose that the prices of coffee and computers are observed and found to have a correlation of +.0008. This means that there is no correlation, or relationship, between the two variables.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 3: **A man is known to speak the truth 3/4 times. He draws a card and reports it is king. Find the probability that it is actually a king**. The probability that king is drawn = ¼. The probability that the king is not drawn = ¾. The probability that the man says the truth that king is drawn when actually king is drawn = ¾. The probability that the man lies that king is drawn when actually king is drawn = ¼. By using Bayes Theorem, the probability that it is actually a king = [1/4 × 3/4] ÷ [(1/4 × 3/4) + (1/4 × 3/4)] = 3/16 ÷12/16 = 3/16 × 16/12 =1/2 = 0.5. Thus, the probability that the drawn card is actually a king = 0.5.

THEORY THAT CANNOT BE APPLIED: Correlation Coefficient

Example 3: Suppose the value of oil prices is directly related to the prices of airplane tickets, with a correlation coefficient of +0.95. The relationship between oil prices and airfares has a very strong positive correlation since the value is close to +1. So, if the price of oil decreases, airfares also decrease, and if the price of oil increases, so do the prices of airplane tickets.

Q8_A6

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 1: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED: Probability Theorem

Example 1: Suppose the probability of obtaining a number 4 on rolling a fair dice needs to be established. The number of favorable outcomes is 1. The possible outcomes of the dice are {1, 2, 3, 4, 5, 6}. This implies that there are a total of 6 outcomes. Thus, the probability of obtaining 4 on a dice roll, using probability theory, can be computed as 1 / 6 = 0.167.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 2: Assume that the chances of a person having a skin disease are 40%. Assuming that skin creams and drinking enough water reduces the risk of skin disease by 30% and prescription of a certain drug reduces its chance by 20%. At a time, a patient can choose any one of the two options with equal probabilities. It is given that after picking one of the options, the patient selected at random has the skin disease. Find the probability that the patient picked the option of skin screams and drinking enough water. A) The patient uses skin creams and drinks enough water = ½, and B) the patient uses the drug = ½. Using the probabilities known to us, we have: The selected patient has the skin disease for the case A = 0.4 × (1-0.3) = 0.28 and for the case B = 0.4 × (1-0.2) = 0.32. Using Bayes Theorem, the probability that the selected patient uses skin creams and drinks enough water is (0.28 × 0.5)/(0.28 × 0.5 + 0.32 × 0.5) = 0.14/(0.14 + 0.16) = 0.47. Thus, the probability that the patient picked the first option is 0.47.

THEORY THAT CANNOT BE APPLIED: Probability Theorem

Example 2: What is the probability of drawing a queen from a deck of cards? A deck of cards has 4 suits. Each suit consists of 13 cards. Thus, the total number of possible outcomes = (4) (13) = 52. There can be 4 queens, one belonging to each suit. Hence, the number of favorable outcomes = 4. The card probability = 4/52 = 1/13. The probability of getting a queen from a deck of cards is 1/13.

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example 3: **A man is known to speak the truth 3/4 times. He draws a card and reports it is king. Find the probability that it is actually a king**. The probability that king is drawn = ¼. The probability that the king is not drawn = ¾. The probability that the man says the truth that king is drawn when actually king is drawn = ¾. The probability that the man lies that king is drawn when actually king is drawn = ¼. By using Bayes Theorem, the probability that it is actually a king = [1/4 × 3/4] ÷ [(1/4 × 3/4) + (1/4 × 3/4)] = 3/16 ÷12/16 = 3/16 × 16/12 =1/2 = 0.5. Thus, the probability that the drawn card is actually a king = 0.5.

THEORY THAT CANNOT BE APPLIED: Probability Theorem

Example 3: **When two dice are rolled what is the probability of getting a sum of 8?** When two dice are rolled there are 36 possible outcomes. To get the sum as 8 there are 5 favorable outcomes: [(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)]. Using probability theory formulas, Probability = Number of favorable outcomes / total number of possible outcomes. For this reason, the probability of getting the sum as 8 when two dice are rolled is 5 / 36.

Q8_A7

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

Q8_A8

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED: Correlation Coefficient

Example: Suppose the value of oil prices is directly related to the prices of airplane tickets, with a correlation coefficient of +0.95. The relationship between oil prices and airfares has a very strong positive correlation since the value is close to +1. So, if the price of oil decreases, airfares also decrease, and if the price of oil increases, so do the prices of airplane tickets.

Q8_A9

THE RIGHT ANSWER IS OPTION B. THEORY APPLIED: Likelihood theorem

Example: **Amy has two bags. Bag I has 7 red and 4 blue balls and bag II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the bag I.** The probability of choosing a bag for drawing a ball is ½. Since there are 7 red balls out of a total of 11 balls in the bag I, therefore, drawing a red ball from the bag I= 7/11. Similarly, drawing a red ball from bag II = 5/14. Then, we will use Bayes Theorem and we have the following: [((7/11) (1/2))/ (7/11) (1/2) + (5/14) (1/2)] = 0.64. Hence, the probability that the ball is drawn is from bag I is 0.64.

THEORY THAT CANNOT BE APPLIED: Probability Theorem

Example: Suppose the probability of obtaining a number 4 on rolling a fair dice needs to be established. The number of favorable outcomes is 1. The possible outcomes of the dice are {1, 2, 3, 4, 5, 6}. This implies that there are a total of 6 outcomes. Thus, the probability of obtaining 4 on a dice roll, using probability theory, can be computed as 1 / 6 = 0.167.

NINTH QUIZ

Q9_A1

THE RIGHT ANSWER IS OPTION C.

The **multiplication rule for independent events** (the one event does not affect the existence of the other event) relates the probabilities of two events to the probability that they both occur. In order to use the rule, we need to have the probabilities of each of the independent events. Given these events, the multiplication rule states the probability that both events occur is found by multiplying the probabilities of each event. **Regarding the order of births, we consider each birth as an independent event. The probability of having boy or girl is the same at 0.5. Therefore, any sequence of six children is as likely as any other sequence. The right answer is 72. More specifically, the probability of having a boy is 1/2. The probability of a having a girl is 1/2. The probability of having this sequence G B G B B G is $1/2^6$, and similarly the probability of having this sequence B G B B B B is $1/2^6$.**

Q9_A2

THE RIGHT ANSWER IS OPTION C.

The **multiplication rule for independent events** (the one event does not affect the existence of the other event) relates the probabilities of two events to the probability that they both occur. In order to use the rule, we need to have the probabilities of each of the independent events. Given these events, the multiplication rule states the probability that both events occur is found by multiplying the probabilities of each event. **Regarding the order of births, we consider each birth as an independent event. The probability of having boy or girl is the same at 0.5. Therefore, any sequence of six children is as likely as any other sequence. The right answer is 72. More specifically, the probability of having a boy is 1/2. The probability of a having a girl is 1/2. The probability of having this sequence G B G B B G is $1/2^6$, and similarly the probability of having this sequence B G B B B B is 1/26.**

ON THE CONTRAST, a **t-test** is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another. Clearly, we are not interested in measuring the statistical difference between the first and the second sequence of births.

Q9_A3

THE RIGHT ANSWER IS OPTION C.

The **multiplication rule for independent events** (the one event does not affect the existence of the other event) relates the probabilities of two events to the probability that they both occur. In order to use the rule, we need to have the probabilities of each of the independent events. Given these events, the multiplication rule states the probability that both events occur is found by multiplying the probabilities of each event. **Regarding the order of births, we consider each birth as an independent event. The probability of having boy or girl is the same at 0.5. Therefore, any sequence of six children is as likely as any other sequence. The right answer is 72. More specifically, the probability of having a boy is 1/2. The probability of a having a girl is 1/2. The probability of having this sequence G B G B B G is $1/2^6$, and similarly the probability of having this sequence B G B B B B is 1/26.**

The multiplication rule is a way to find the probability of two events happening at the same time. When we calculate probabilities involving one event AND another event occurring, we multiply their probabilities. In some cases, the first event happening impacts the probability of the second event. We call these dependent events. We cannot apply the rule of **multiplication of dependent events** in our case because, each birth does not affect the other. Thus, we consider them as independent events.

Q9_A4

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 1: **Suppose that we roll a six sided die and then flip a coin.** These two events are independent. The probability of rolling a 1 is 1/6. The probability of a head is 1/2. The probability of rolling a 1 and getting a head is 1/6 x 1/2 = 1/12. We see that there are twelve outcomes, all of which are equally likely to occur. Therefore the probability of 1 and a head is 1/12.

Example 2: **Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. We then ask what the probability that both cards are kings is.** Since we have drawn with replacement, these events are independent and the multiplication rule of independent events applies. The probability of drawing a king for the first card is 1/13. The probability for drawing a king on the second draw is 1/13. The reason for this is that we are replacing the king that we drew from the first time. Since these events are independent, we have the product 1/13 x 1/13 = 1/169.

Example 3: **Imagine you particularly like wearing tan pants with a blue shirt. However, in the morning, you're sleepy and grab your pants and shirt randomly from the closet. The pants are on one side of the closet while the shirts are on the other.** These are independent events because grabbing a pair of pants doesn't affect the probabilities for shirts. You have ten pairs of pants and three are tan. Consequently, the probability of drawing a tan pair is 0.3. You have 16 shirts and four are blue. Hence, the probability of grabbing a blue shirt is 0.25. These are independent events because selecting a pair of pants doesn't affect the likelihood of drawing a blue shirt and vice versa. Using the specific multiplication rule for these independent events: 0.3 X 0.25 = 0.075.

Q9_A5

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 1: **Suppose that we roll a six sided die and then flip a coin.** These two events are independent. The probability of rolling a 1 is 1/6. The probability of a head is 1/2. The probability of rolling a 1 and getting a head is 1/6 x 1/2 = 1/12. We see that there are twelve outcomes, all of which are equally likely to occur. Therefore the probability of 1 and a head is 1/12.

THEORY THAT CANNOT BE APPLIED: T-test of statistical measurement

Example 1: Let's say you have a cold and you try a naturopathic remedy. Your cold lasts a couple of days. The next time you have a cold, you buy an over-the-counter pharmaceutical and the cold lasts a week. You survey your friends and they all tell you that their colds were of a shorter duration (an average of 3 days) when they took the homeopathic remedy. What you really want to know is, are these results repeatable? A t-test can tell you by comparing the means of the two groups and letting you know the probability of those results happening by chance.

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 2: Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. We then ask what the probability that both cards are kings is. Since we have drawn with replacement, these events are independent and the multiplication rule applies. The probability of drawing a king for the first card is 1/13. The probability for drawing a king on the second draw is 1/13. The reason for this is that we are replacing the king that we drew from the first time. Since these events are independent, we use the multiplication rule to see that the probability of drawing two kings is given by the following product 1/13 x 1/13 = 1/169.

THEORY THAT CANNOT BE APPLIED: T-test of statistical measurement

Example 2: A drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there's always a control group (a group who are given a placebo, or "sugar pill"). The control group may show an average life expectancy of +5 years, while the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student's t-test to find out if the results are repeatable for an entire population.

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 3: Imagine you particularly like wearing tan pants with a blue shirt. However, in the morning, you're sleepy and grab your pants and shirt randomly from the closet. The pants are on one side of the closet while the shirts are on the other. These are independent events because grabbing a pair of pants doesn't affect the probabilities for shirts. You have ten pairs of pants and three are tan. Consequently, the probability of drawing a tan pair is 0.3. You have 16 shirts and four are blue. Hence, the probability of grabbing a blue shirt is 0.25. These are independent events because selecting a pair of pants doesn't affect the likelihood of drawing a blue shirt and vice versa. Using the specific multiplication rule for these independent events: 0.3 X 0.25 = 0.075.

THEORY THAT CANNOT BE APPLIED: T-test of statistical measurement

Example 3: In a school, 100 students in class A scored an average of 85% with a standard deviation of 3%. Another 100 students belonging to class B scored an average of 87% with a standard deviation of 4%. While the average of class B is better than that of class A, it may not be correct to jump to the conclusion that the overall performance of students in class B is better than that of students in class A. This is because there is natural variability in the test scores in both classes, so the difference could be due to chance alone. A t-test can help to determine whether one class fared better than the other.

Q9_A6

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 1: **Suppose that we roll a six sided die and then flip a coin.** These two events are independent. The probability of rolling a 1 is 1/6. The probability of a head is 1/2. The probability of rolling a 1 and getting a head is 1/6 x 1/2 = 1/12. We see that there are twelve outcomes, all of which are equally likely to occur. Therefore the probability of 1 and a head is 1/12.

THEORY THAT CANNOT BE APPLIED: Multiplication rule for dependent events

Example 1: **Suppose you are interested in the probability of drawing hearts on two consecutive draws.** Initially, the deck has 13 hearts out of its 52 cards (13/52 = 0.25). If you draw a heart (event H1), that changes the probability of drawing another heart. The dependent probability of drawing that second heart (event H2) is now 12/51 = 0.235.

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 2: **Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. We then ask what the probability that both cards are kings is.** Since we have drawn with replacement, these events are independent and the multiplication rule applies. The probability of drawing a king for the first card is 1/13. The probability for drawing a king on the second draw is 1/13. The reason for this is that we are replacing the king that we drew from the first time. Since these events are independent, we use the multiplication rule to see that the probability of drawing two kings is given by the following product 1/13 x 1/13 = 1/169.

THEORY THAT CANNOT BE APPLIED: Multiplication rule for dependent events

Example 2: **Imagine that we're packing for a short trip and randomly select two pairs of pants and two pairs of shirts to include in our suitcase. We're hoping for two pairs of tan pants and two blue shirts.** We'll start by treating this as two sets of dependent events, one for pants and the other for shirts. We begin with 10 pairs of pants, three of which are tan. Consequently, the probability of the first pair of pants being tan is 0.30. The probability of the second pair being tan is 2/9 = 0.22. Hence: 0.30 * 0.22 = 0.066.

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example 3: Imagine you particularly like wearing tan pants with a blue shirt. However, in the morning, you're sleepy and grab your pants and shirt randomly from the closet. The pants are on one side of the closet while the shirts are on the other. These are independent events because grabbing a pair of pants doesn't affect the probabilities for shirts. You have ten pairs of pants and three are tan. Consequently, the probability of drawing a tan pair is 0.3. You have 16 shirts and four are blue. Hence, the probability of grabbing a blue shirt is 0.25. These are independent events because selecting a pair of pants doesn't affect the likelihood of drawing a blue shirt and vice versa. Using the specific multiplication rule for these independent events: 0.3 X 0.25 = 0.075.

THEORY THAT CANNOT BE APPLIED: Multiplication rule for dependent events

Example 3: Suppose you take out two cards from a standard pack of cards one after another, without replacing the first card. What is probability that the first card is the ace of spades, and the second card is a heart? The two events are dependent events because the first card is not replaced. There is only one ace of spades in a deck of 52 cards. So: 1/52. If the ace of spaces is drawn first, then there are 51 cards left in the deck, of which 13 are hearts: 13/51. So, by the multiplication rule of probability, we have: 1/52 x 13/51 = 1/204.

Q9_A7

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example: **Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. We then ask what the probability that both cards are kings is.** Since we have drawn with replacement, these events are independent and the multiplication rule applies. The probability of drawing a king for the first card is 1/13. The probability for drawing a king on the second draw is 1/13. The reason for this is that we are replacing the king that we drew from the first time. Since these events are independent, we use the multiplication rule to see that the probability of drawing two kings is given by the following product 1/13 x 1/13 = 1/169.

Q9_A8

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example: **Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. We then ask what the probability that both cards are kings is.** Since we have drawn with replacement, these events are independent and the multiplication rule applies. The probability of drawing a king for the first card is 1/13. The probability for drawing a king on the second draw is 1/13. The reason for this is that we are replacing the king that we drew from the first time. Since these events are independent, we use the multiplication rule to see that the probability of drawing two kings is given by the following product 1/13 x 1/13 = 1/169.

THEORY THAT CANNOT BE APPLIED: T-test pf statistical measurement

Example: Let's say you have a cold and you try a naturopathic remedy. Your cold lasts a couple of days. The next time you have a cold, you buy an over-the-counter pharmaceutical and the cold lasts a week. You survey your friends and they all tell you that their colds were of a shorter duration (an average of 3 days) when they took the homeopathic remedy. What you really want to know is, are these results repeatable? A t-test can tell you by comparing the means of the two groups and letting you know the probability of those results happening by chance.

Q9_A9

THE RIGHT ANSWER IS OPTION C. THEORY APPLIED: Multiplication rule for independent events

Example: **Suppose that we draw a card from a standard deck, replace this card, shuffle the deck and then draw again. We then ask what the probability that both cards are kings is.** Since we have drawn with replacement, these events are independent and the multiplication rule applies. The probability of drawing a king for the first card is 1/13. The probability for drawing a king on the second draw is 1/13. The reason for this is that we are replacing the king that we drew from the first time. Since these events are independent, we use the multiplication rule to see that the probability of drawing two kings is given by the following product 1/13 x 1/13 = 1/169.

THEORY THAT CANNOT BE APPLIED: Multiplication rule for dependent events

Example: **Suppose you take out two cards from a standard pack of cards one after another, without replacing the first card. What is probability that the first card is the ace of spades, and the second card is a heart?** The two events are dependent events because the first card is not replaced. There is only one ace of spades in a deck of 52 cards. So: 1/52. If the ace of spaces is drawn first, then there are 51 cards left in the deck, of which 13 are hearts: 13/51. So, by the multiplication rule of probability, we have: 1/52 x 13/51 = 1/204.