Incidental statistical learning and its limitations

A thesis submitted in fulfilment of the requirements

for the degree of Master of Science (By Research)

by

Anish Thankachan

21128402

DEPARTMENT OF COGNITIVE SCIENCE

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

November 2023

POST GRADUATE OFFICE
THESIS SUBMITTED
ON 06/02/2024
I.I.T. KANPUR

Certificate

It is certified that the work contained in this thesis entitled **"Incidental statistical learning and its limitations"** by **Anish Thankachan** has been carried out under my supervision and that it has not been submitted elsewhere for a degree.

Digitally signed by Nisheeth Srivastava
DN: cn=Nisheeth Srivastava, o=IIT Kanpur, ou=CGS, email=nsrivast@iitk.ac.in, c=IN
Date: 2023.11.07 14:31:53 +05'30'

Dr. Nisheeth Srivastava

Associate Professor

Department of Computer Science and Engineering and Cognitive Science

Indian Institute of Technology Kanpur

November 2023

Declaration

This is to certify that the thesis titled **"Incidental statistical learning and its limitations"** has been authored by me. It presents the research conducted by me under the supervision of **Dr. Nisheeth Srivastava**.

To the best of my knowledge, it is an original work, both in terms of research content and narrative, and has not been submitted elsewhere, in part or in full, for a degree. Further, due credit has been attributed to the relevant state-of-the-art and collaborations with appropriate citations and acknowledgments, in line with established norms and practices.

Anish

Anish Thankachan
Roll No. 21128402
CGS Department
Indian Institute of Technology Kanpur

Declaration

(To be submitted at DOAA Office)

I hereby declare that

1. The research work presented in the report titled **"Incidental statistical learning and its limitations"** has been conducted by me under the guidance of my supervisor **Dr. Nisheeth Srivastava**.

2. The thesis has been formatted as per Institute guidelines.

3. The content of the thesis (text, illustration, data, plots, pictures etc.) is original and is the outcome of my research work. Any relevant material taken from the open literature has been referred and cited, as per established ethical norms and practices.

4. All collaborations and critiques that have contributed to giving the thesis its final shape is duly acknowledged and credited.

5. Care has been taken to give due credit to state-of-the-art in the thesis research area.

6. I fully understand that in case the thesis is found to be unoriginal or plagiarized, the Institute reserves the right to withdraw the thesis from its archive and also revoke the associated degree conferred. Additionally, the Institute also reserves the right to apprise all concerned sections of society of the matter, for their information and necessary action (if any).

Anish

Anish Thankachan

Roll No. 21128402

CGS Department

Indian Institute of Technology Kanpur

Abstract

Name of the student: **Anish Thankachan** Roll No: **21128402**

Degree for which submitted: **M.S. (By Research)** Department: **Cognitive Science**

Thesis title: **Incidental statistical learning and its limitations**

Thesis supervisors: **Dr. Nisheeth Srivastava**

Month and year of thesis submission: **November 2023**

This research delves into the phenomenon of incidental statistical learning and its underlying mechanisms in human cognition. The study extends prior research and aims to dissect the mechanisms and constraints associated with incidental statistical learning. The first experiment replicates a previous study but incorporates a modification that removes presumptions about the contexts in which distributions are encountered. This modification renders the conditional priors uninformative, effectively preventing participants from accessing potential generative distributions. The findings indicate that in the absence of access to these generative models, individuals are unable to incidentally learn context-specific distributions.

Moreover, this research introduces various computational models to elucidate the cognitive processes involved in such statistical learning. Notably, the study reveals that compared to a fully Bayesian observer who accurately retains all exemplars, a sequential sampling observer, which compresses the encountered data, better aligns with participants' task performance. This underscores the presence of generative models in human cognition that adeptly abstract the task at hand, primarily by recalling a subset of observations and subsequently along with prior knowledge, invert them to derive context-specific probabilistic assessments.

A follow-up experiment is conducted to delve into the factors influencing category-specific incidental learning of magnitudes. This experiment manipulates the timing, sequence, and content of background information presented to participants. The results illuminate that participants can learn contextual distributions when equipped with generative models before the training phase. However, when background information is presented after training, incidental learning does not occur, emphasizing the necessity of prior access to generative models. Furthermore, the study underscores the pivotal role of the content of background knowledge in facilitating learning. This content can either provide additional insights into generative models or enhance the salience and relevance of the context. Effective background knowledge aids participants in adjusting their generative models based on observed data and in seamlessly integrating new statistical information. Conversely, when background knowledge lacks these crucial elements, learning is impeded or entirely absent. The findings also underscore participants' selective focus on task-relevant dimensions or metrics, while disregarding irrelevant ones.

In contrast to previous research on incidental learning, which primarily centers on replicating observed frequencies in perceptual and language tasks, this study explores a distinct form of statistical learning involving retrospective probability judgments through Bayesian inversion of conditional distributions. The research underscores that mere exposure to distributions is insufficient for this form of learning; participants necessitate a causal model and informative priors regarding the current situation. Without these foundational components, participants struggle to engage in probabilistic reasoning and make accurate probability judgments.

Acknowledgements

I would like to extend my deepest gratitude to my thesis advisor, Dr. Nisheeth Srivastava, for entrusting me with the freedom to conduct my research and for maintaining faith in my capabilities throughout this academic journey. His invaluable guidance, support, commitment, and profound expertise have been instrumental in shaping the success of this research project.

I offer a special appreciation to Mr. Ishan Singhal for his continuous mentorship, encouragement, and the invaluable insights he provided during this period. His guidance has been quite inspiring and has played an indispensable role in enhancing my research skills and fostering my academic growth.

I also wish to convey my gratitude to Professor Narayanan Srinivasan, who serves as the Head of the Department of Cognitive Science. His encouragement and invaluable guidance were readily available during the more challenging phases of this research.

I am also thankful to the esteemed faculty members of our Cognitive Science department for their enlightening contributions and support. Their passion for teaching has been an enduring source of inspiration, and the courses they offered have significantly enriched my intellectual development.

My heartfelt appreciation goes to the dedicated staff and personnel at IIT Kanpur, particularly within our department, for creating a nurturing academic environment and providing the essential resources required for the successful completion of this thesis.

I would like to express my deep gratitude to my family and friends for their understanding and love. Their belief in my abilities and their constant support have served as my driving force throughout this academic journey.

I acknowledge with gratitude the participants who generously dedicated their time and actively took part in this study. Their contributions have been indispensable in collecting the data and insights necessary for this research.

Finally, I wish to extend my sincere appreciation to all those who have contributed in any capacity to the culmination of this thesis, whether directly or indirectly. Your contributions, regardless of their scale, have played a substantial role in shaping this body of work.

Completing this master's thesis has been a challenging yet fulfilling experience, and I am genuinely thankful to each and every individual who has been part of this significant journey.

Contents

List of Figures

List of Tables

Dedicated to family.

Chapter 1

Introduction

1.1 Background

Throughout our daily lives, we constantly encounter ambiguous situations and rely on our perceptual judgments and subsequent actions to navigate and interact with our environment. Human cognition is a multifaceted and complex process that enables us to understand, interpret, and interact with the world around us. It encompasses various mental processes such as perception, attention, memory, language, problem-solving, and decision-making. Numerous studies have been conducted to understand the mechanisms underlying various aspects of human cognition. For example, Gibson proposed what is called the ecological approach to perception (Gibson, 1966). Perception is the process by which we interpret and make sense of sensory information from our environment coming in through various modalities, such as vision, hearing, touch, taste, and smell. This approach emphasizes the role of the environment in guiding perception. According to this ecological approach, perception is an active process where the perceiver directly interacts with the environment and extracts meaningful information. This approach challenges the traditional view of perception as a purely bottom-up process and highlights the importance of context and environmental cues. Another famous example is the Prospect theory, developed by Kahneman and Tversky (1979), which provides a framework for understanding decision-making under uncertainty. This theory suggests that individuals' choices are influenced not only by objective probabilities and outcomes but also by subjective perceptions of gains and losses. It challenges the traditional rational choice model and highlights the role of cognitive biases and heuristics in decision-making.

1.2 Statistical (Bayesian) inference

Similarly, one such influential theory in perception and action, which involves statistical inference and allows the incorporation of subjectivity, is known as the Bayesian inference approach (Pouget et al., 2013). Mathematically, Statistical inference is a general framework encompassing various methods and techniques for drawing conclusions from data. It involves using probability theory and statistical models to make inferences about population parameters based on sample data. Bayesian inference is a specific approach to statistical inference that has its foundations in Bayesian probability theory. In addition to actual sample data, it also incorporates prior knowledge and beliefs into the inference process. In Bayesian inference, probabilities are used to express subjective degrees of beliefs, and such prior beliefs are operationalized through what is known as prior distributions. Bayesian inference treats population parameters as random variables and characterizes uncertainty in terms of probability distributions. Once the current context-specific observations are encountered, these prior distributions are updated, in accordance with the Bayes' theorem, using this sample data to obtain a new distribution (called posterior distribution) that better fits the observed data. This new posterior distribution represents the updated knowledge about the distribution parameters after encountering the data, and serves as the basis for making further inferences and decisions.

This Bayesian inference approach to human cognition, posits that the brain actively combines prior knowledge and incoming sensory data to generate perceptual judgments. According to this framework, our prior beliefs or expectations about the world, derived from past experiences, are combined with the sensory evidence to form a probabilistic representation of the environment (Ernst & Banks, 2002). These representations are used to make perceptual judgments, which play a vital role in guiding our actions by enabling us to estimate the consequences of different action choices under uncertainty and select the most appropriate course of action (Körding & Wolpert, 2004). And this, in turn, guides in updating and revising the initial beliefs based on the current sensory input and the action feedback to form a better representation of the environment. The role of statistical inference in perception can be illustrated through various perceptual phenomena (Rahnev & Denison, 2018). For example, in visual illusions, our brain may use prior knowledge and expectations to interpret ambiguous or conflicting visual cues. The brain tends to perceive what is most likely or probable given the available information, even if it contradicts the actual sensory input. This phenomenon demonstrates how statistical inference plays a crucial role in shaping our perceptual judgments. Similarly, in motor control, for instance, statistical inference allows us to adjust our movements based on sensory feedback

(Trommershäuser et al., 2008). When performing complex tasks, the brain continuously integrates sensory information and predicts the outcomes of different actions. By comparing the predicted outcomes with the actual feedback received during execution, the brain updates its expectations and refines subsequent actions. This iterative process involves statistical inference to optimize action planning and execution. This theory is also tested in the context of computational neuroscience. The studies demonstrate how neural populations can encode and update beliefs based on Bayesian principles. Ma et al. (2006) presents a computational model showing how neural populations can represent and update probability distributions using Bayesian inference. The authors demonstrate how a population of neurons can encode and update beliefs about the orientation of a visual stimulus. Deneve and Pouget (2004) explore how neural populations can perform Bayesian inference for multisensory integration, and they propose a model in which populations of neurons combine visual and auditory information in a Bayesian manner to estimate the location of a sensory stimulus.

1.3 Criticism of Bayesian inference

Although Bayesian models offer a principled framework for reasoning under uncertainty and have been successful in explaining various cognitive phenomena, but many including Jones and Love (2011) point out that they lack theoretical grounding and provide unlimited flexibility to explain the behavior. The freedom to choose from a wide range of prior distributions allows for the incorporation of prior knowledge, but it also introduces subjectivity. With numerous possibilities for prior specification, researchers have the freedom to shape the results based on their beliefs or assumptions. Such analysis is often prone to overfitting, especially when the model is highly flexible, and there is limited data. Overfitting usually occurs when a model becomes overly complex and starts to capture noise or idiosyncrasies in the data, leading to poor generalization of new data. The flexibility of Bayesian analysis, with its ability to accommodate complex models, can inadvertently result in models that fit the data too closely. This flexibility can lead to different conclusions depending on the chosen prior distributions, potentially making it challenging to compare and interpret results across studies. Therefore, the paper by Jones and Love (2011) disparages such results where the experimental data seems to be a good fit, but is essentially so due to the design of that experiment using selective environmental and psychological properties. And so, the authors advocate for the development of theoretical frameworks that go beyond modeling and provide deeper mechanistic explanations for cognitive processes in order to set reasonable boundaries for the degrees of freedom.

1.4 Limitations of such engineered priors

The properties of the environment and the biophysical embodiment of individuals have a significant impact on shaping the nature of prior beliefs or expectations that they possess in perceptual, motor control, and language-related tasks. In perceptual tasks, for example, the properties of the environment, such as regularities, brightness, occlusion, etc., can shape the prior expectations that individuals have about the sensory inputs they receive. Similarly, in motor control tasks, the physical properties of the environment, including its affordances and constraints, as well as our own bodies' physical constraints influence the prior beliefs about the actions and movements required to achieve desired outcomes. Furthermore, in some language-related cognitive tasks, such as speech perception or language comprehension, the acoustic cues or linguistic structures or pauses in between the words provide some constructs about the incoming linguistic information. But we should also acknowledge that producing such strong constraints on prior knowledge becomes more challenging in cognitive tasks where neither embodiment is strongly involved nor are the environmental statistics deterministic. In such cases, the influence of the environment and embodiment on shaping prior beliefs is not as clear or straightforward. Additionally, when environmental statistics are variable and unpredictable, it becomes more challenging to form precise and reliable prior expectations. The lack of consistent environmental cues or patterns makes it harder to constrain the prior knowledge in a way that can consistently guide perception, action, and other cognitive processes. Similarly, in abstract cognitive tasks that do not involve direct physical interactions or specific environmental cues, it becomes harder to establish strong and consistent constraints on prior knowledge. In these cases, individuals may rely more on general cognitive processes, such as reasoning or abstract thinking, rather than specific environmental cues or bodily experiences. Thereby mitigating the prevailing apprehensions surrounding Bayesian inference and rendering it a viable and credible explanation.

1.5 Incidental learning

Numerous instances exist in which individuals employ such general cognitive processes when confronted with markedly diverse environmental conditions. Griffiths and Tenenbaum (2006) proposed that, in the realm of everyday probability judgments, humans tend to exhibit behavior closely aligned with Bayes optimality. In this context, Bayes optimality refers to the capacity to make optimal probabilistic inferences by effectively updating prior beliefs based on observed evidence using Bayes' theorem. To manifest such behavior in the context of routine and commonplace judgments, it implies that individuals must possess an inherent understanding of the underlying distributions that govern a wide array of real-world phenomena. This suggests that people not only demonstrate the ability to perceive and interpret specific instances of events but also possess the capacity to incidentally retain information about the statistical properties and regularities that underpin these events. This acquired knowledge enables them to effectively update their beliefs and make accurate probability judgments in everyday scenarios. On similar lines, Stewart et al. (2006) propose the decision-by-sampling theory as an explanation for risk aversion. According to this theory, individuals make decisions based on sampling experiences, where they mentally simulate and evaluate potential outcomes by sampling from a distribution of possible outcomes. Accepting this theory requires believing that people have the ability to keep track of the money magnitude distributions incidentally. In other words, individuals possess an implicit awareness of the range and variability of monetary values, even in the absence of explicit monetary training or formal learning. By incidentally tracking the distribution of money magnitudes, individuals can make more informed decisions and exhibit risk-averse behavior when faced with uncertain monetary outcomes.

Here, incidental learning refers to the process of acquiring knowledge or information unintentionally or without explicit intention to learn. It occurs when individuals acquire new knowledge or skills as a byproduct of engaging in other activities or tasks that are not specifically designed for learning purposes. In contrast to intentional learning, where individuals actively seek out information or engage in deliberate learning strategies, incidental learning happens spontaneously and without conscious effort. It often occurs in everyday life situations, where individuals are exposed to stimuli or information in their environment and unconsciously absorb or acquire knowledge from these experiences. One common example of incidental learning is language acquisition in children. As children interact with their environment and engage in social interactions, they naturally pick up the sounds, words, and grammar of their native language without explicit instruction (Saffran et al., 1996). They acquire language incidentally through exposure and immersion in a linguistic environment. Another example is incidental learning in educational settings. While students attend classes or engage in activities, they may acquire knowledge or skills related to the subject matter, even if the learning was not the primary goal. For instance, students may learn new vocabulary words while reading a novel or acquire problem-solving strategies while working on a group project. Incidental learning can also occur through exposure to media, such as television, movies, or online content. Individuals may acquire new information, cultural knowledge, or skills through passive exposure to these media without actively intending to learn. Overall, incidental learning highlights the capacity of individuals to acquire knowledge and skills in an unintentional and spontaneous manner through exposure to their environment or engagement in various activities. It demonstrates the inherent learning potential that exists in our everyday experiences, even when learning is not the primary focus.

1.6 Dependence of Incidental Statistical Learning on Perceptual Bias

In the context of incidental statistical learning, one of the primary factors that influence this process is the strength of perceptual biases that individuals possess (Aslin & Newport, 2012). Perceptual biases refer to individuals' inherent cognitive predispositions or preferences, shaping how they perceive and interpret the world around them. These biases play a crucial role in identifying and extracting structural regularities, facilitating generalization across different scenarios. Perceptual biases provide individuals with a cognitive framework that biases their attention towards certain features or patterns in their sensory inputs. These biases can be influenced by various factors, including evolutionary factors, cultural influences, and individual experiences. For example, humans have a natural bias towards detecting faces, known as the face processing bias. This bias leads individuals to prioritize facial features and facilitates the rapid recognition of faces in different contexts. The strength of these perceptual biases influences how individuals attend to and process information in their environment. Strong perceptual biases make individuals more sensitive to specific types of regularities or patterns, allowing them to more readily identify and extract statistical information. For instance, individuals may have a bias towards attending to sequential dependencies or co-occurrences of events, such as the likelihood of one event following another. This bias enables them to detect transitional probabilities and frequency of occurrence, an important statistical cue for learning regularities. By leveraging their perceptual biases, individuals can identify and extract structural regularities from complex and noisy environments. These regularities can manifest in various forms, including temporal, spatial, or relational patterns. Once these regularities are detected, individuals can generalize their knowledge across different scenarios and make predictions about future events or stimuli that share similar structural characteristics. Moreover, perceptual biases can interact with other cognitive processes, such as attention and memory, to facilitate incidental statistical learning. The biases guide individuals' attention toward relevant information and enhance the encoding and retention of statistically relevant patterns. This process supports the formation of memory representations that capture the underlying statistical structure of the environment.

1.7 Prior knowledge influences/manifests as biases

These perceptual biases, which significantly influence human cognition and interpretation of sensory information, do not exist in isolation; rather, they are deeply rooted in individuals' pre-existing expectations, theories, and prior knowledge of the world. In essence, these perceptual biases are intricately entwined with and informed by individuals' existing expectations, theories, and worldly insights. These preconceived expectations and knowledge play a pivotal role in shaping these perceptual biases, a fact well-supported by the literature. The works on illusory correlation (Chapman, 1967) and confirmation bias (Johnson-Laird & Wason, 1977) underscore how individuals have a propensity to exaggerate covariation between events in a manner that aligns with their pre-existing expectations. Individuals carry with them a wealth of knowledge concerning the statistical regularities, causal connections, and distinctive attributes of objects and events within their environment. This prior knowledge significantly influences how they perceive and interpret incoming sensory information. For instance, consider the scenario where someone possesses prior knowledge that associates specific objects with particular sounds; in such cases, they are more inclined to perceive those sounds when confronted with ambiguous auditory stimuli (McGurk & MacDonald, 1976).

These prior expectations are byproducts of past experiences, cultural influences, and individual belief systems, collectively shaping a cognitive framework that individuals employ to interpret and make sense of incoming sensory data. People not only passively assimilate knowledge but actively construct theories and mental models of the world based on their observations and interactions. These theories serve as cognitive frameworks, helping individuals organize information into coherent structures. They guide the interpretation of sensory input by channeling attention toward information that aligns with the theories and frameworks individuals have developed. Perceptual biases can also be forged through the processes of learning and adaptation. When individuals are repeatedly exposed to specific patterns or associations in their environment, their perceptual systems adapt to these regularities over time. These learned associations consequently become integral components of perceptual biases, directing individuals to attend to and interpret stimuli in accordance with the acquired patterns. The integration of these influences from prior knowledge, theories, and learning experiences significantly shapes how individuals perceive and interpret the world around them.

1.8 Need for conceptual coherence

In the realm of incidental learning, another pivotal determinant of effective acquisition and retention of new stimuli or concepts is their compatibility with an individual's existing knowledge base (Murphy, 1985). This entails that these novel concepts should seamlessly mesh with an individual's pre-existing mental structures. When this alignment occurs, it fosters the integration and interrelation of the new concept with other aspects of one's knowledge, resulting in improved retention and comprehension. Murphy's proposition emphasizes that learning transcends the mere accumulation of isolated fragments of information; instead, it revolves around the establishment of connections and associations between new and pre-existing knowledge. When a novel concept harmoniously fits within the framework of an individual's existing knowledge, it enhances comprehension and simplifies the formation of meaningful connections. Conversely, concepts that fail to connect or interact with the pre-existing knowledge may appear isolated and unrelated, rendering them more challenging to remember and apply.

This concept harmonizes with the broader cognitive principles of schema theory and concept formation. Schema theory postulates that individuals organize their knowledge into mental frameworks or schemas, acting as cognitive structures for interpreting and understanding the world. New information is incorporated into these schemas, and the existing schema plays a pivotal role in facilitating the encoding, storage, and retrieval of novel knowledge. Furthermore, when a novel concept lacks alignment with the existing knowledge base, it can introduce cognitive dissonance and inconsistency. In such instances, individuals may encounter difficulties in assimilating the new concept, resulting in reduced stability and increased forgetfulness. The absence of coherence and integration with the existing knowledge base impedes the formation of robust associations and meaningful linkages, rendering the concept more susceptible to decay and overshadowing by more coherent knowledge.

To illustrate, let's consider a student learning about the concept of gravity. If the student possesses a solid foundation in basic physics principles, the concept of gravity seamlessly integrates into their existing knowledge framework. They can draw connections between gravity and concepts like force, motion, and acceleration, facilitating the learning and retention process. Conversely, if the student lacks a fundamental understanding of physics and fails to establish prior connections with the concept of gravity, grasping and retaining this new concept becomes more challenging. Therefore, the coherence of new concepts with an individual's existing knowledge base stands as a pivotal factor in effective learning and retention. Concepts that align and interact with pre-existing knowledge are more likely to be retained and integrated into cognitive frameworks, whereas concepts lacking such coherence may remain unstable and susceptible to fading into oblivion.

1.9 Theories/models in categorization incorporating prior knowledge

Categorization tasks are one of the extensively studied cognitive processes. There are many models proposed to explain how this prior knowledge can affect categorization tasks:

1. Prototype Model: The prototype model assumes that people form a prototype or average representation based on the features of category exemplars. Categorization judgments are made by comparing the similarity between a new stimulus and the prototype. This model predicts that prior knowledge influences categorization by providing a prototype that serves as a reference point for determining category membership.

2. Exemplar Model: The exemplar model suggests that individuals store and compare specific exemplars encountered during category learning. Categorization is based on the similarity between a new stimulus and the exemplars stored in memory. Prior knowledge affects categorization by providing a set of relevant exemplars that influence the determination of category boundaries.

3. Feature Weighting Model: The feature weighting model posits that prior knowledge influences category learning by guiding the allocation of attention or importance to different features. Prior knowledge biases the weighting of features, resulting in selective attention and enhanced learning of relevant features. This model suggests that prior knowledge affects categorization by modulating the contribution of individual features to the categorization process.

4. Rule-Based Model: The rule-based model assumes that people use explicit rules or decision boundaries to categorize stimuli. Prior knowledge influences categorization

by providing knowledge of relevant rules or decision strategies. This model predicts that prior knowledge shapes the selection and application of decision rules, leading to more accurate categorization.

5. Rescorla-Wagner Model: The Rescorla-Wagner model is a well-known model of associative learning. It proposes that learning occurs through the adjustment of associations between stimuli and outcomes. Prior knowledge influences categorization by influencing the strength of associative links between category exemplars and category labels. This model suggests that prior knowledge facilitates learning by strengthening the associations between relevant exemplars and their category membership

The paper by Heit (1994) explores various theories that explain how prior knowledge can affect categorization tasks. The authors propose and compare different computational theories to investigate the influence of prior knowledge on category learning. The models discussed in the paper provide theoretical frameworks to understand how prior knowledge affects categorization. They offer insights into the cognitive processes involved in category learning and highlight the role of prior knowledge in shaping categorization strategies and decision-making. An overview of the models discussed in their paper is as follows:

- Integration Theory: Integration theory, also known as the exemplar-similarity theory, proposes that categorization involves integrating information from multiple exemplars to form a representation of a category. It suggests that people compare new stimuli to the entire set of exemplars they have encountered and make judgments based on their similarity to those exemplars. The theory emphasizes the role of exemplar similarity and the integration of multiple exemplars in categorization.
- Weighing Theory: Weighing theory, or the feature-weighting model, suggests that individuals selectively attend to and weigh different features when making categorization judgments. It assumes that features are not equally important in categorization and that individuals assign different weights or importance to different features based on their relevance to the category. Weighing theory predicts that prior knowledge influences categorization by biasing the weighting of features, resulting in selective attention and enhanced learning of relevant features.
- Distortion Theory: Distortion theory, also known as the distortion hypothesis, proposes that people categorize stimuli based on a prototype or central tendency but allow for some flexibility or distortion in categorization judgments. It suggests that

individuals are sensitive to the distribution of exemplars within a category and allow for variations or distortions around the prototype. Distortion theory posits that prior knowledge influences categorization by shaping the range or extent of distortions allowed within a category.

1.10 The boundary of research

Substantial evidence underscores the remarkable capacity of humans to engage in probabilistic reasoning when equipped with an understanding of the pertinent generative model of the world. This model acts as a blueprint, encapsulating the fundamental processes responsible for generating the observed data or events. In numerous scenarios, individuals adeptly employ probabilistic reasoning to make precise predictions and judgments, harnessing their knowledge of these generative models to navigate uncertainty. A common example involves the domain of coin flips, where individuals, armed with an appreciation for the probabilistic nature of such events, can effectively forecast the likelihood of outcomes like "heads" or "tails" based on their profound comprehension of the underlying generative process.

However, the pivotal question arises when we consider how individuals are able to, incidentally acquire, update, and effectively utilize these generative models in novel contexts. As mentioned earlier, in this context, "incidentally" refers to the spontaneous and implicit acquisition and application of these models without explicit, deliberate instruction. While ample evidence supports humans' ability to engage in probabilistic reasoning when provided explicit instruction and equipped with knowledge of generative models, the mechanisms through which individuals incidentally acquire this knowledge remain less clear. This intricacy prompts inquiries into the means by which individuals naturally develop an understanding of appropriate generative models across diverse domains and situations. The process of incidental generative model acquisition likely encompasses a medley of factors, including exposure to pertinent information, the ability to glean statistical insights from the environment, the cumulative impact of prior experiences, and inherent cognitive mechanisms for pattern recognition and generalization. Despite the strides made in understanding of probabilistic reasoning and generative models, the precise mechanisms and processes underpinning the incidental acquisition and ongoing refinement of these models continue to be subjects of ongoing research and investigation.

Chapter 2

Literature Review

2.1 The Role of Abstract Knowledge and Bayesian Models in Understanding Human Cognition

In the field of cognitive science, researchers have been facing the daunting task of unraveling how our minds can extract meaningful information from the limited and often noisy input provided by our senses. Despite this challenge, our cognitive abilities allow us to construct complex mental models, make generalizations, and form abstract concepts. One remarkable aspect of human cognition is our ability to learn the meanings of words with only a few examples and then apply that knowledge to new situations. For example, let's say you are shown a set of dinosaurs as seen in Figure 2.1, and told that the ones that are highlighted in green belong to a specific category X. And your task is to judge which among the two marked in red belong to the same category X?

We see that even with such little data, there is a lot of processing that is happening in the back of your mind to bring out this response. Out of all the attributes, you might have observed the similarity of the exoskeleton and long tail, which is what we call the "singly necessary and jointly sufficient attribute" for category X, while being a quadruped is not sufficient. We see how our mind goes over and beyond the data available, makes strong generalizations, and constructs powerful abstractions over this limited data to come up with solutions. One of such abstraction or mental model might be something like a cladogram, as shown in Figure 2.2.

To understand the mechanisms behind this ability to generalize from sparse data Tenenbaum et al. (2011) have proposed Bayesian models. These models incorporate prior beliefs

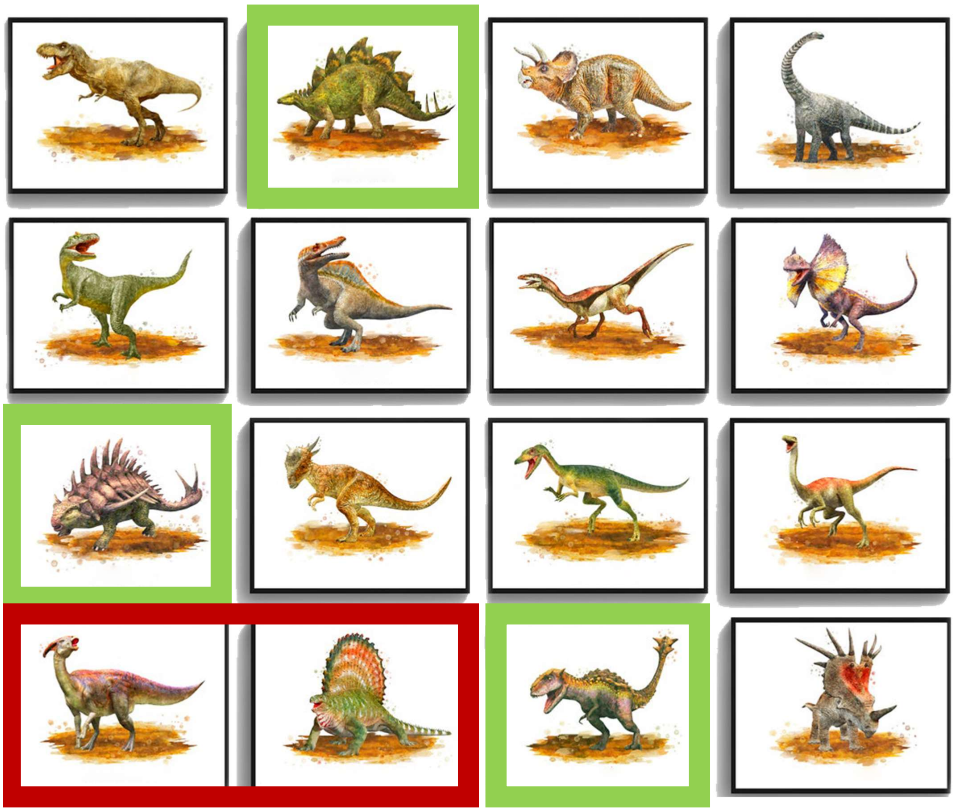


FIGURE 2.1: Samples from several categories of dinosaurs

or background knowledge with observed data to make inductive inferences. By combining existing knowledge with new information, our minds can navigate the vast complexity of the world and make sense of it. Abstract knowledge plays a crucial role in learning and cognitive development. It takes various forms, such as hierarchical tree structures, graphs, grammars, and schemas. These structures provide frameworks for organizing information and identifying patterns and relationships. Through the acquisition of abstract knowledge, we discover the appropriate organizational structures and inductive constraints that govern different domains. The acquisition of abstract knowledge is an integral part of human intelligence. It allows us to go beyond the specific instances we encounter and derive general principles that apply to a wide range of situations. Bayesian models offer a powerful framework for studying human intelligence by combining sophisticated knowledge representation with statistical inference. By simulating the learning and inference processes observed in human cognition, these models provide valuable insights into the underlying mechanisms of how we learn, reason, and form abstract concepts.

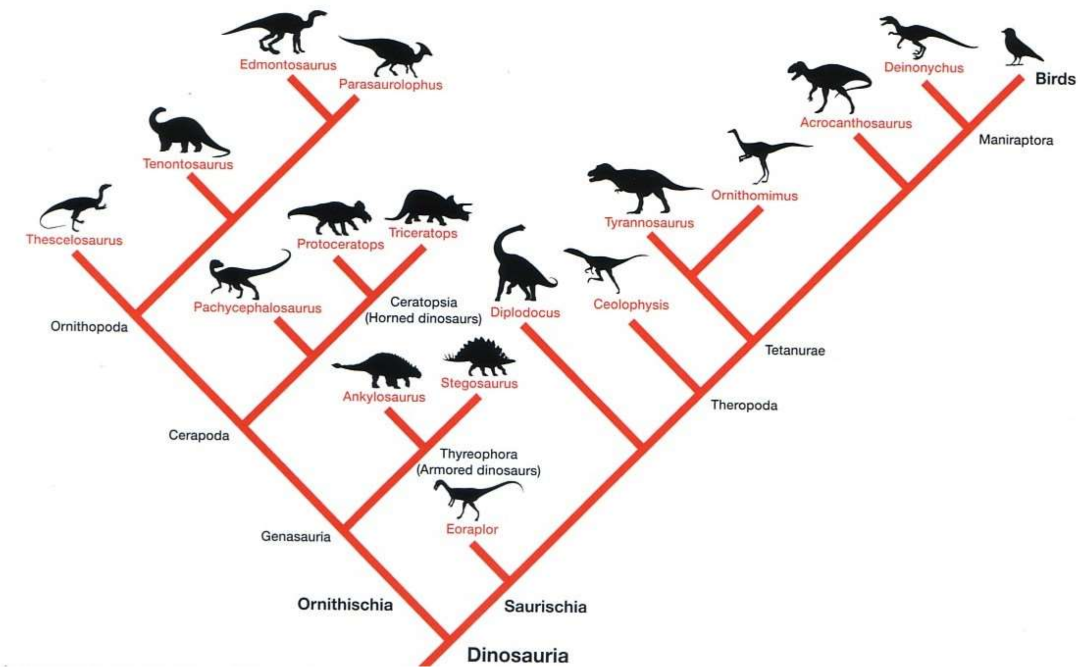


FIGURE 2.2: A possible mental model for categories of dinosaurs

2.2 Human cognition follows optimal statistical principles in real-world situations

When it comes to perceiving the world and remembering information, humans are often thought to operate in a way that is consistent with the principles of sound statistical reasoning, making use of available information and prior knowledge. But on the other hand, human cognitive judgments are believed to be fallible and subject to biases. The study conducted by Griffiths and Tenenbaum (2006) sheds light on the relationship between human cognition and optimal statistical inference, challenging the conventional notion that human judgments are inherently erroneous. The researchers designed experiments where participants were required to make predictions regarding various real-life scenarios, such as estimating the total box office gross for a movie or predicting the life span of a teenager, given their present values. These scenarios encompassed a broad range of phenomena whose actual empirical distributions could be Gaussian, exponential, Erlang, etc. These participant responses were then compared to the predictions generated by a Bayesian model. The Bayesian model employed in this research applied Bayes's rule to compute probability distributions, for example, concerning gross earnings for a given movie, given specific observed data points, such as a movie's box office earnings at several points in time. The likelihood function is the probability of encountering specific box office earnings given the general expectations regarding the underlying phenomenon. The comparison between the predictions of the Bayesian model and that of participants is shown in Figure 2.3. Each vertical column represents the specific scenarios presented to the participants. The first figure plots the empirical distribution for the total extent of each phenomenon. The types of empirical distributions include Gaussian, exponential, and Erlang distributions. The next row presents the various predictions. The straight dotted lines show predictions of and uniform prior, while the solid line is the prediction by the aforementioned Bayesian model. On top of this, we see participants' median predictions as black dots.

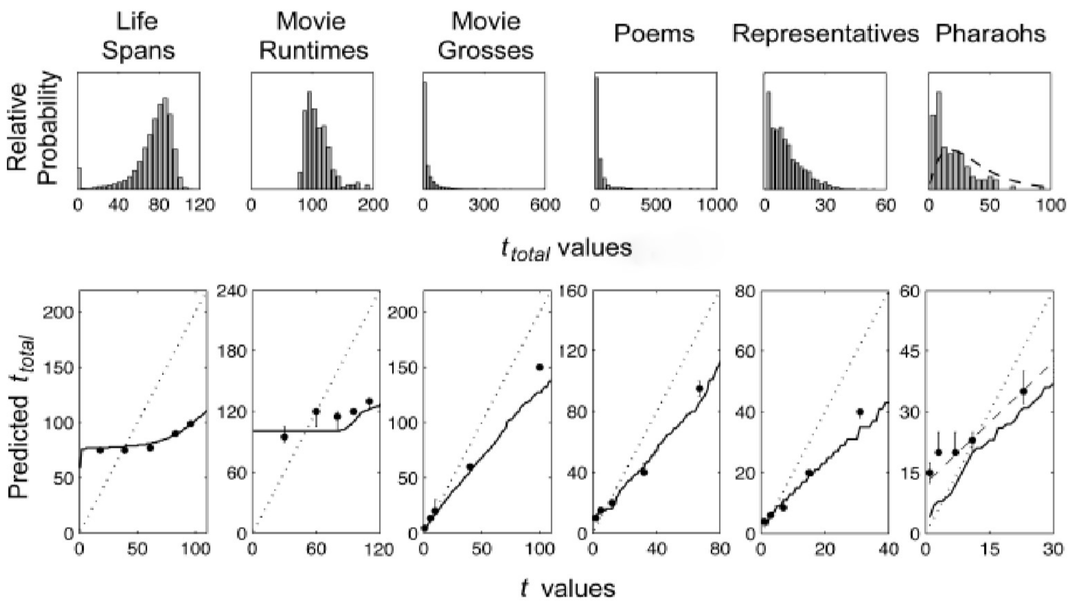


FIGURE 2.3: (Griffiths & Tenenbaum, 2006) Participants' judgments and the predictions of the Bayesian model across various scenarios

The study's findings revealed a striking alignment between the predictions made by participants and those generated by the Bayesian model, demonstrating high accuracy in participants' judgments across a range of scenarios. This suggests that people possess fairly accurate mental models for understanding real-life situations. Notably, these models are not generic or context-neutral but are tailored to specific situations. This is in contrast with the common belief that cognitive judgments are predominantly influenced by biased mental shortcuts, implying that people's intuitive judgments actually reflect the statistical properties of the world. In exploring cases where participants' predictions deviated from the optimal Bayesian predictions, such as when estimating the box office returns, the researchers proposed that participants might use a strategy involving analogical reasoning. This approach draws on their knowledge of more familiar situations to make predictions about unfamiliar events. In this process, participants may correctly infer the generic underlying distribution but make inaccurate assumptions while tweaking these distributions to fit the current scenario. The research underscores the vital role of optimal statistical inference in the realm of everyday cognitive judgments, mirroring its importance in processes like perception and memory. The capacity to seamlessly amalgamate well-defined preexisting beliefs is not a novel concept, having found applications across various domains of human reasoning. Nevertheless, the distinctive value of this study lies in its ability to showcase the pragmatic optimality of these models within the complexities of real-world contexts. By doing so, it accentuates the potential of Bayesian statistics as a robust framework for comprehending the intricacies of human inductive reasoning.

2.3 Incidental learning of the shape of probability distributions

To test the human ability to acquire such statistical knowledge incidentally, Tran et al. (2017) did a study where they found results both in favor and against such a form of learning. The study's objective was to explore whether exposing participants to a restricted number of samples from a specific distribution allows them to acquire accurately the fundamental structure of the distribution and subsequently replicate it. They ran a series of studies in which participants had to click on things on the screen. A bimodal distribution (purple distribution in Figure 2.4) was used to produce the horizontal placements of these items on the screen. Participants were shown 180 such samples; this training phase was common to all experiments.

Then, for the test phase in experiment 1, participants were asked to generate another set of target positions from the same distribution. The results (orange distribution in Figure 2.4) showed that participants' responses did not exhibit a bimodal distribution or accurately reproduce the learned distribution. Experiment 2 involved pre-announcing the testing

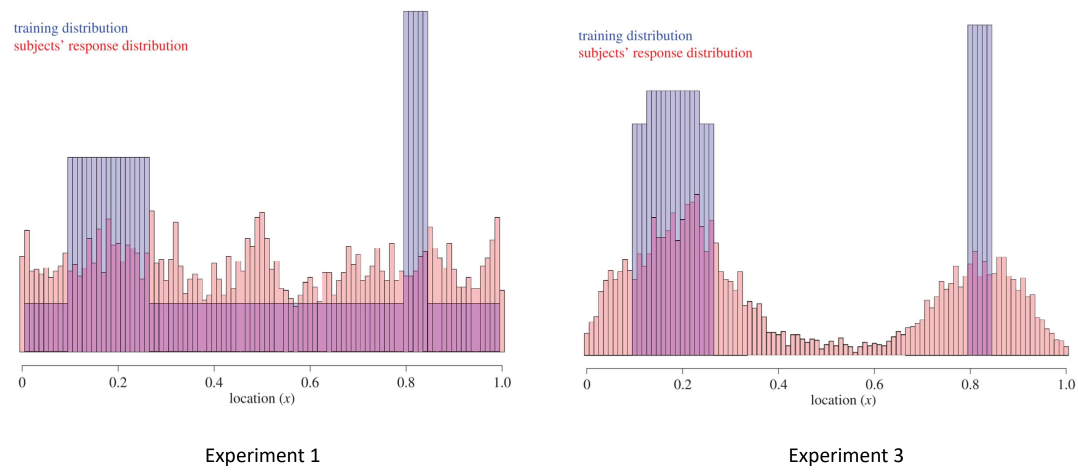


FIGURE 2.4: (Tran et al., 2017) Training distributions and participant responses for experiments 1 and 3

phase to the participants, informing them that they would be asked to generate target positions. Despite this explicit instruction to learn the distribution, participants still failed to reproduce it accurately. Experiment 3 introduced a modification to the distribution by adding a significant gap between the two distributions. Participants were able to reproduce a bimodal distribution in this case (as seen in orange distribution over experiment 3 in Figure 2.4), indicating that they could acquire knowledge about the modified bimodal distribution. The paper discusses several possible interpretations of the findings. It implies that the discreteness of the distribution may affect learning when learning happens non-parametrically for continuous probability distributions. Furthermore, learning continuous probability distributions in such manners may have some constraints, and individuals often learn by fine-tuning parameters rather than non-parametrically. The authors acknowledge that incomplete memory or prior assumptions favoring unimodal distributions could also play a role in participants' failure to reproduce the learned distribution accurately.

We tried to explain the results seen in experiments one and three in the above study using a form of observer model. This model draws a few samples from the memory corresponding to the training phase observations. And tries to fit a function to these samples that would approximate the generative model of the training distribution. We propose that participants may perceive the need to fit a bounded and positive continuous distribution to the training phase observation, and the beta function fits this requirement. However, due to the limited number of samples drawn from memory, the parameter estimation of the posterior beta distribution may be significantly skewed. In experiment 1, where participants observed non-zero probability density throughout the horizontal space, they may randomly choose samples from throughout the range of the learned distribution. This can lead to a distorted parameter estimate of the beta function. This distortion arises because the samples may not be representative of the underlying distribution, causing the fitted beta function to deviate from the true shape. However, in experiment 3, the participants are exposed to a perfectly discriminable bimodal distribution, where the two modes are easily distinguishable, and there is a zero-density gap between the two modes. In such cases, participants are more likely to draw samples from these modes, and therefore the predicted beta function would also be bimodal. This observation suggests that participants' ability to accurately reproduce the learned distribution depends on the discriminability of the modes.

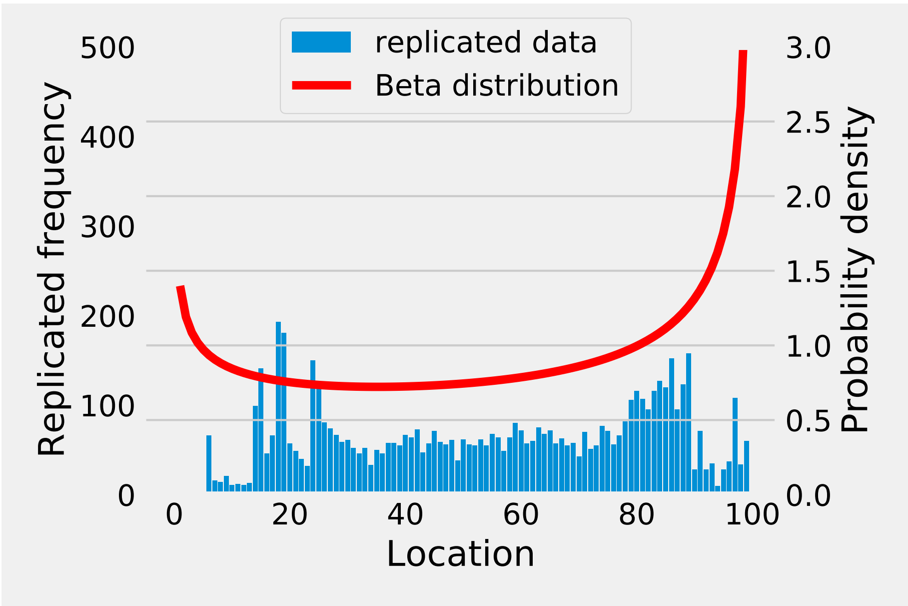


FIGURE 2.5: Simulated reproduction of training distribution using data from Tran et al. (2017) paper, across 35 runs. In each run, 180 values are drawn from a beta distribution, which is fit to 5 random samples from training distribution for Experiment 1 from Tran et al. (2017) paper

One possible simulation illustrating the potential outcome of fitting a beta function to the randomly sampled observations from experiments 1 and 3 is shown in Figure 2.5 and 2.6, respectively. This explanation provides a possible account for why participants in the study can sometimes accurately replicate the learned distribution while at other times they struggle to do so. It suggests that the choice of sampling strategy, the number of samples, and the assumption of a specific functional form of generative function (here, a

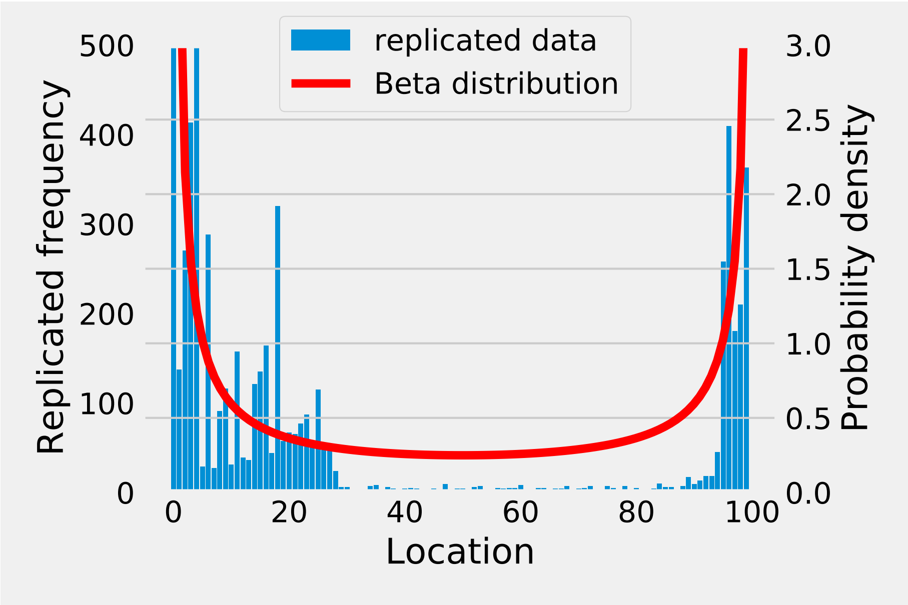


FIGURE 2.6: Simulated reproduction of training distribution using data from Tran et al. (2017) paper, across 35 runs. In each run, 180 values are drawn from a beta distribution, which is fit to 5 random samples from training distribution for Experiment 3 from Tran et al. (2017) paper

beta function) may all contribute to the challenges encountered in distribution learning tasks.

2.4 Incidental Learning of Context-Dependent Probability Distributions

While previous research has provided evidence for humans' probabilistic reasoning abilities when they have an understanding of the generative model, the mechanisms driving such learning are still unclear. In a recent paper by Singhal et al. (2021), they investigate how people are able to learn statistical distributions that are context-specific, and give probability judgments based on them retrospectively. They designed an experiment to test whether participants could learn and utilize probabilistic information about context-specific distributions without explicit instruction or motivation, and invert it later to reason about those distributions. They recruited 76 participants to play a game in which their capacity to learn context-specific probability distributions was assessed by having them pretend to be auditors and inspect some bills for accuracy. To ensure that participants paid attention to the numerical values on the bills while remaining unaware of the true objective of the experiment, they gave them a straightforward arithmetic activity. The experiment had two stages: a training stage and a testing stage. In the course of the training, 20 restaurant bills were displayed to the participants, and each of them had three items along with their monetary values(see Figure 2.7). These invoices clearly stated that they came from either "cheap" or "expensive" restaurants. The values in the bill were taken from two normal distributions with mean values - 1600 and 2900. The lower and higher value distribution was for cheap and expensive restaurants, respectively. Variances for these distributions were defined as a factor of mean, these factors being 0.25 for low, 0.5 for medium, and 0.7 for high. One of three variance groups was assigned at random to each participant. The aim of the training phase was for the participants to get acquainted with the value distributions p(magnitude|context) for each restaurant and know the frequency of context occurrence p(context) throughout the training phase. Later, participants were asked to express p(context|magnitude) in each of the 40 testing phase trials.

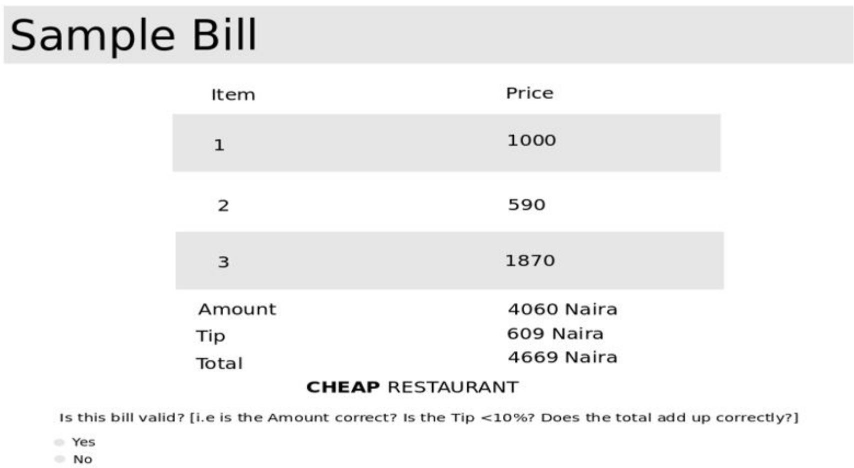


FIGURE 2.7: (Singhal et al., 2021) An image of stimuli that was displayed during one of the iterations in the experiment's training phase (Bill from a restaurant).

The results from each participant during the testing phase were fitted using a psychometric function (see Figure 2.8). The psychometric analysis supports the hypothesis that the variance of numerical observations during the training phase has an impact on participants' posterior probability judgments during the surprise testing phase. In other words, the degree of variability within the category distributions influences participants' ability to accurately categorize new instances. However, what makes the current study unique is the focus on incidental category learning. Participants in the experiment were not explicitly instructed to learn or memorize values for different category labels. They were engaged in an auditing task where they had to identify totaling mistakes in bills during the training phase. Crucially, participants were not informed that a testing phase would follow the testing phase or given any feedback on their replies during the testing phase. Participants were not explicitly informed that they needed to learn or recall the values associated with various categories because of the retroactive nature of the design.

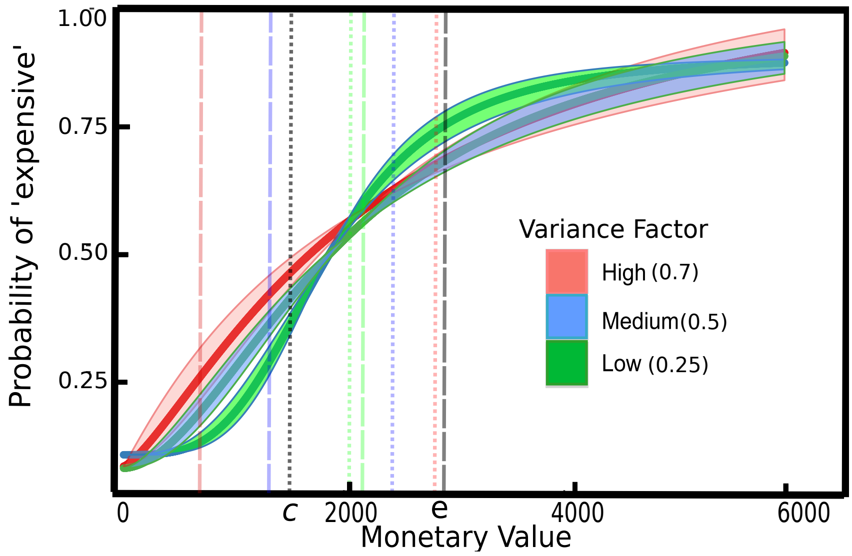


FIGURE 2.8: (Singhal et al., 2021) The mean of the underlying distribution is shown as a dashed black line for expensive restaurant, and a dotted black line for cheap restaurant. The three variances are also represented as a dashed line for expensive restaurant and a dotted line for cheap restaurant (the color for each variance value is as shown in legend). Using the averages of parameter estimates for each participant within each of the three variance groups, a psychometric function was fit for these three groups. The shaded region denotes the 95% confidence interval for the slope parameters.

In addition to examining participants' performance in the experiments, the paper employs observer models to analyze their behavior. These models are fitted to the observations that participants saw during incidental learning (see Figure 2.9), allowing the researchers to gain insights into the mental representations used in the retrospective probabilistic reasoning task. The authors analyze the human behavior using three variations of observer models:

1. Ideal Observer model: It updates the generative model by fetching up each of the 60 training instances encountered during training.

2. Prototype Observer Model: It considers that users will be able to recall the average of the category distributions shown during the training phase. And use this pair of means to update an abstract generative model of the process in order to provide probability judgments.

3. Memory Sampling Extension of Prototype Observer Model: It is a modification of the prototype observer model, which considers that people construct prototypes using a

subset of samples from stored or recalled observations corresponding to the training phase. Instead of using all the training samples, participants select these prototypes that are then used to update the generative model, similar to the prototype observer model.

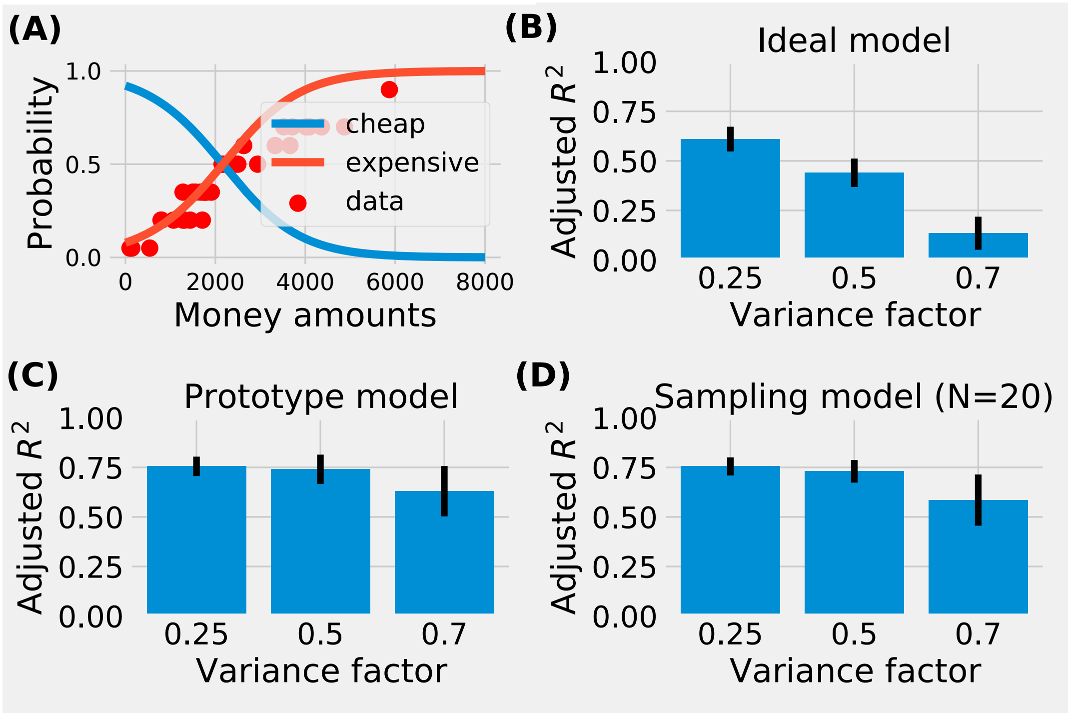


FIGURE 2.9: (Singhal et al., 2021) (A), The test phase responses of a random participant fit to an ideal observer model's predictions. (B-D) The median goodness of fit for each variance condition corresponding to each observer model.

The paper provides empirical evidence for incidental supervised category learning from a small number of observations. The results support the assumption in Bayesian cognitive science that humans intuitively store statistical information about events in the world, even when not explicitly motivated to do so. Additionally, computational modeling is used to demonstrate two important properties of this incidental learning. First, participants in tests of incidental category learning behave in a way that is consistent with a Bayesian observer who selects a subset of observations from memory in order to form probability judgments. This finding suggests that humans rely on memory-based sampling to reason probabilistically about the learned distributions. Second, the study shows that the number of samples drawn from memory varies depending on the discriminability of the categories. When categories are easily discriminable, fewer samples are drawn, whereas in cases of less discriminable categories, more samples are drawn. These results are consistent with the concept of resource-rationality in memory sampling, which posits that individuals adjust their sampling strategies based on the available cognitive resources and the complexity of the task.

2.5 Effect of prior knowledge concept learning

In a study conducted by Wattenmaker (1999), they investigated how prior knowledge affects concept learning in intentional and incidental learning contexts. They explore the idea that prior knowledge plays a significant role in shaping the process of concept learning. The author suggests that when individuals encounter new concepts, their existing knowledge base, theories, and expectations about the world around them come into play. Prior knowledge serves as a foundation upon which new information is integrated and understood. To examine the influence of prior knowledge, the study compares intentional and incidental concept learning. Intentional learning refers to situations where individuals are aware of the goal to learn and actively engage in learning the concepts. Incidental learning, on the other hand, occurs when individuals acquire knowledge without an explicit intention to learn. The author compares the encoding of conceptually related co-occurrences (congruent with prior knowledge) with conceptually unrelated co-occurrences (incongruent with prior knowledge). The focus is on how prior knowledge influences the encoding and retention of information in both intentional and incidental learning contexts.

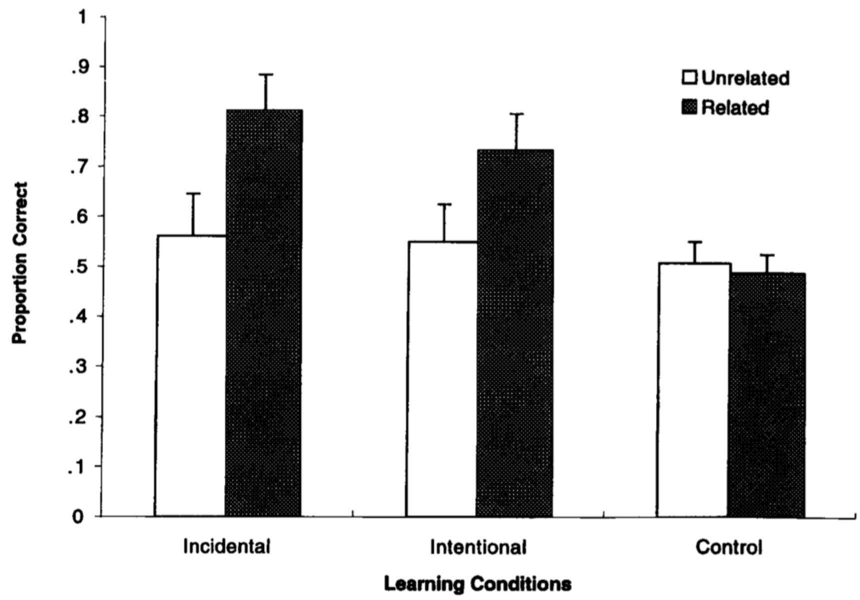


FIGURE 2.10: (Wattenmaker, 1999) Proportions of responses in which transfer descriptions with co-occurring features were selected, as a function of encoding condition and type of co-occurrence.

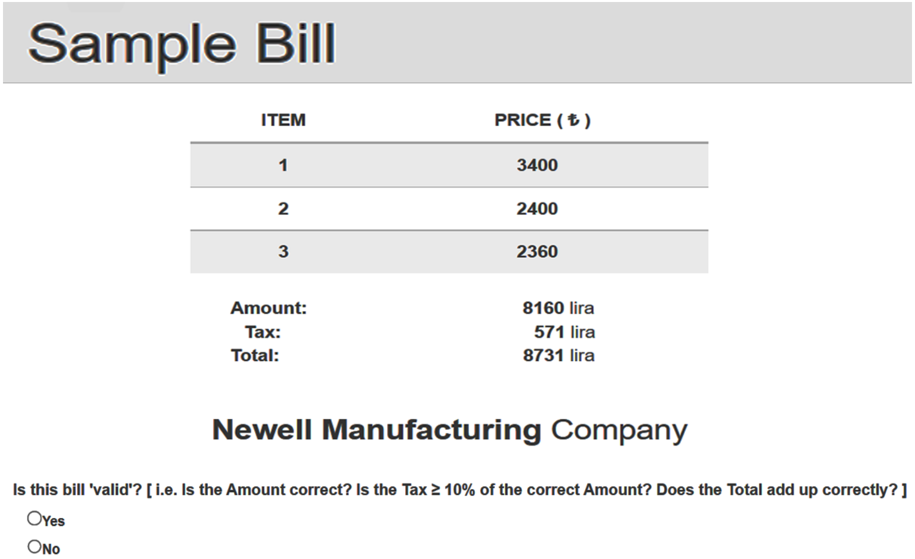
The results of the study (shown in Figure 2.10) indicate that regardless of the encoding method (intentional or incidental), past knowledge structures are readily available and employed on their own during concept learning. The findings also reveal a clear influence of prior knowledge on incidental concept learning, where conceptually related co-occurrences are better encoded and remembered compared to conceptually unrelated co-occurrences. These findings suggest that prior knowledge, even when acquired incidentally, significantly impacts the encoding and retention of new information. The study highlights the robust influence of prior knowledge in the incidental learning process and emphasizes its role in facilitating the integration and understanding of new concepts.

Chapter 3

Experiment 1

Positive results from the experiment conducted by Singhal et al. (2021), indicating people's ability to learn the context-specific statistical distribution and give probability judgments retrospectively, led us to test the limits of such incidental learning. We wished to see if the prior knowledge about the relative positions of these distributions (acquired through the nomenclature - 'cheap' and 'expensive') played a role in such learning. To test this, we designed a variation of the experiment conducted by Singhal et al. (2021). Again, participants were instructed to play the part of auditors and assess the tax conformity of a company's supplier procurement invoices. But now, these bills were mentioned as being from either "Preston Industrial" or "Newell Manufacturing" companies (to ensure that these company names did not provide any informative cues about their distribution's relative means, we intentionally selected common nouns as the company names). Similar to the original experiment, participants were explicitly asked to check for the correct total (sum of all items' cost) and tax values (Tax values have to be at least 10% of the total cost of the products.).

The experiment consisted of two distinct phases: training and testing. In the training phase, participants were presented with a series of supply procurement bills (see sample bill in Figure 3.1). These bills were designed to resemble real invoices and contained three items listed with their respective money values. Importantly, each bill was explicitly labeled as originating from either the "Preston Industrial" or "Newell Manufacturing" company. This labeling ensured that participants associated specific bills with the corresponding company names. The training phase of the experiment aimed to establish two important aspects: the value distributions of money magnitude given a specific context (company), and the probability of context occurrence in participants' minds. By exposing



Sample Bill

ITEM	PRICE (₺)
1	3400
2	2400
3	2360

Amount: 8160 lira
Tax: 571 lira
Total: 8731 lira

Newell Manufacturing Company

Is this bill 'valid'? [i.e. Is the Amount correct? Is the Tax ≥ 10% of the correct Amount? Does the Total add up correctly?]

Yes

No

FIGURE 3.1: An image of stimuli that was displayed during one of the iterations in experiment-1's training phase (procurement bill of a company)

participants to various supply procurement bills labeled with specific company names, we sought to instantiate these value distributions and contextual probabilities within the participants' cognitive frameworks. Following the training phase, the experiment transitioned to the testing phase. In this phase, individual items were presented to participants along with their corresponding prices. For each item, participants were required to assess the likelihood that it belonged to the "Newell Manufacturing" company's procurement bill. In other words, they were asked to express the probability of a specific context given the observed price.

3.1 Participants

An Institutional Review Board (IRB) examined and approved the study protocol for this investigation. The experiment was carried out via an online platform. This allowed for a convenient and efficient data collection process, potentially reaching a broader participant pool. The sample size for the study was determined based on the original study that served as a reference (Singhal et al., 2021). Initially, a total of 88 participants took part in the experiment. These participants had a mean age of 25.1 years, with 21 of them being female. However, certain exclusion criteria were applied to refine the sample and focus on participants who met specific performance standards. Eight participants were excluded from the study due to poor performance in the training phase. Specifically, these participants demonstrated an accuracy level below 80%. Additionally, 10 participants were excluded from the analysis because they provided identical responses for all the problems in the test phase. This suggests that these participants might not have fully engaged with the task or may have provided unreliable data. After applying these exclusion criteria, the final sample for analysis consisted of 70 participants. These individuals were retained for further data analysis and interpretation.

3.2 Stimuli

During the training phase of the experiment, participants were presented with values that were drawn from two separate normal distributions. These values were represented as bills from companies labeled as "Preston Industrial" and "Newell Manufacturing" (See Figure 3.1). For each company label, "Preston Industrial" and "Newell Manufacturing," a specific normal distribution was assigned. The mean value for "Preston Industrial" was fixed at 1600, while the mean value for "Newell Manufacturing" was fixed at 2900. Additionally, scaling variance factors were applied to each distribution to manipulate the spread of the values. The three variance factors used were low (0.25), medium (0.5), and high (0.7). These factors determined the variability of the values within each distribution. The manipulation of the experiment involved varying the variance of these distributions among different groups of participants. There were three participant groups, and each group was assigned a specific variance level. The number of participants in each variance group was 27, 24, and 19, respectively, totaling 70 participants in the final analysis.

3.3 Procedure

3.3.1 Training Phase

To ensure randomization and control in the study, three different variance groups were randomly chosen for the participants. This assignment determined which distribution (low, medium, or high variance) the participants would encounter during the training phase. The means of the value distributions for both "Preston Industrial" and "Newell Manufacturing" companies remained constant across all participants, ensuring that any observed differences in participant responses were due to the manipulation of variance rather than the means. As part of the experiment, each participant was given the responsibility of auditing 20 bills. Three different monetary values, drawn as samples from the particular distribution allocated to the participant's variance group, made up each bill. To maintain consistency and balance in the experiment, there were an equal number of bills for each company label. Specifically, there were 10 bills each for "Preston Industrial" and "Newell Manufacturing" procurement bills, resulting in a total of 30 sample money magnitudes (3 x 10) for each context. Participants were explicitly instructed to carefully examine the supply procurement bills and check for two specific aspects: the correct total of all items' costs and the accuracy of the tax values associated with the bills. Participants were required to assess whether the total cost of the items listed on the bill was calculated correctly. This involved summing up the individual costs and comparing it to the provided total. Additionally, participants were instructed to verify the accuracy of the tax values. Specifically, they were asked to ensure that the tax amount indicated on the bill was equal to or greater than 10% of the total cost of the items. During the training phase, participants received feedback on their responses to each bill. This feedback indicated whether their audit was correct or incorrect, serving as an attention check to ensure participants were actively engaged in the task. With an accuracy rate below 80%, participants who struggled with the maths assignment were disqualified from the analysis that followed. This exclusion was required to weed out individuals who might not have understood the task or completed it correctly, ensuring the validity and dependability of the results.

3.3.2 Testing Phase

During the testing phase of the experiment, participants were shown a set of 40 individual items, each assigned a specific value. These values were generated from a distribution with a fixed variance factor of 0.5, which was the same for all participants. The use of a common distribution with a consistent variance factor allowed for a direct comparison of participants' responses across the three groups that were assigned different variance conditions during the training phase. In this testing phase, participants were asked to rate the likelihood of each item being drawn from the procurement bill of the "Newell Manufacturing" company. They provided their ratings on a seven-point Likert scale, which allowed for a range of responses from "highly unlikely" to "highly likely." The purpose of this task was to assess participants' subjective assessments of the association between the item values and the specific company, specifically the probability of a specific context given the observed price. Notably, participants were not informed throughout the training phase that they would subsequently be evaluated on the values presented in the bills.

3.4 Analysis

We used a psychometric function to fit the testing phase responses of each participant.

$$\gamma + \frac{\lambda - \gamma}{1 + (\frac{x}{c})^{\beta}}. \tag{3.1}$$

Note, the psychometric curve's inflexion point was c, its upper asymptote was gamma (γ), its lower asymptote was lambda(λ), and its steepness or slope was beta (β).

In the context of psychometric curves, the parameters γ, λ, c, and β are used to describe the shape and characteristics of the curve. Here's an explanation of each parameter:

- γ (Upper asymptote): This parameter represents the upper limit or maximum value that the psychometric curve can reach. It indicates the level of performance or response rate when the stimulus value is very high. In other words, γ defines the ceiling of the curve.
- λ (Lower asymptote): This parameter represents the lower limit or minimum value that the psychometric curve can reach. It indicates the level of performance or response rate when the stimulus value is very low. λ defines the floor of the curve.
- c (Inflexion point): This parameter represents the point on the psychometric curve where it transitions from a lower response rate to a higher response rate. It is the midpoint of the curve's transition from the lower asymptote (λ) to the upper asymptote (γ). In our case, the inflexion point indicates the stimulus value at which participants' responses are more likely to be ambiguous.
- β (Steepness or slope): This parameter determines the steepness or slope of the psychometric curve. It reflects how quickly the response rate changes as the stimulus value increases or decreases. A higher value of β indicates a steeper curve, meaning that a small change in stimulus intensity leads to a larger change in the response rate. In our case, this would determine the degree or ease of discriminability between the two distributions.

In order to analyze and interpret the Likert ratings provided by the participants, we used the mapping technique that converts these ratings into probability judgments (See Table 3.1). This mapping is the same as the one used in the original study. This mapping was based on a previously validated approach described in a study by Hancock and Volante

(2020). The purpose of this mapping technique is to assign numerical probabilities to the verbal labels used in the Likert scale.

TABLE 3.1: The table lists our conversion from Likert ratings to assigned probabilities.

Likert Ratings	Assigned Probability
1 = Very Unlikely	0.05
2 = Unlikely	0.2
3 = Somewhat Unlikely	0.35
4 = Undecided	0.5
5 = Somewhat Likely	0.6
6 = Likely	0.7
7 = Very Likely	0.9

3.5 Results

In JASP (0.16.3), we performed a between-subject Bayesian ANOVA with three conditions of variance (0.7, 0.5, 0.25) and assumed a standard uniform prior for each of the four parameters. Moderate evidence was found for the null hypothesis that variance conditions would not alter the slopes (β) of the fitted curves (BF_{01} = 7.833, error% = 0.026). For information on the slope variations across the groups, see Figure 3.2. Similarly, for the inflexion point (BF_{01} = 3.452), the lower asymptote (BF_{01} = 3.921), and the upper asymptote (BF_{01} = 4.180), there was weak to moderate evidence in favor of the null hypothesis. We also performed Bayesian one-sample T-test on the slopes (β) of the fitted curves, which provided only weak evidence in support of the null hypothesis, that the population mean slope parameter did not differ significantly from zero (BF_{01} = 5.244, error% = 0.078).

3.6 Discussion

The psychometric analysis of the data from Experiment 1 reveals important findings about participants' category judgments and their ability to learn the context-specific distributions presented during the training phase. Firstly, the Bayesian T-test conducted on the slopes of the fitted curves provides evidence that the posterior probability during the testing phase is not dependent on the monetary values provided as inputs. This is indicated by the lack of evidence for a non-zero value for the slope parameter. In other words, participants' judgments of category membership were not influenced by the values they encountered, suggesting that they were unable to learn the specific distributions presented

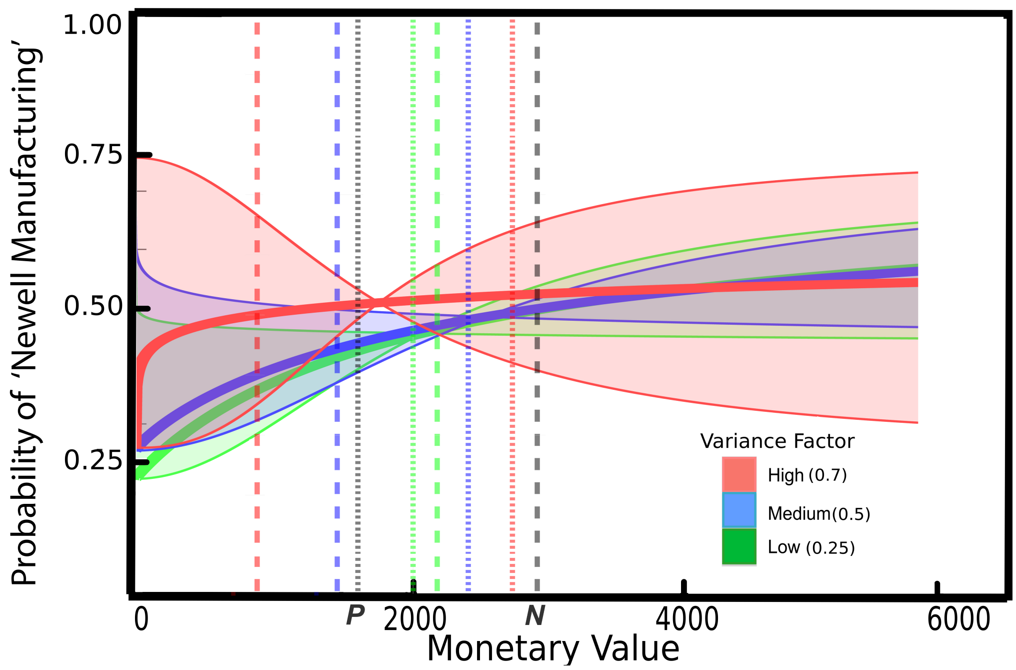


FIGURE 3.2: The mean of the underlying distribution is shown as a dashed black line for 'Newell' company, and a dotted black line for 'Preston' company. The three variances are also represented as a dashed line for 'Newell' company, and a dotted line for 'Preston' company (the color for each variance value is as shown in legend). Using the averages of parameter estimates for each participant within each of the three variance groups, a psychometric function was fit for these three groups of Experiment 1. The shaded region denotes the 95% confidence interval for the slope parameters.

during training. This lack of learning is further supported by the results of the ANOVA, which examined the effects of variance conditions on the slopes. The analysis showed that regardless of the amount of spread in the distribution during the training phase (low, medium, or high variance), participants still exhibited a lack of context-specific learning. The absence of significant differences in the slopes across variance conditions not only suggests that participants were unable to infer the underlying generative distributions even when exposed to different levels of variance, but also discards any effect of the spread of the distribution on learning.

Moreover, when considering the task as a categorization task with two labels, the data from the testing phase indicate that participants' judgments of category membership were no better than chance (See Figure 3.2). This finding is reflected in the random chance level of selecting the correct category label in the data. The lack of discrimination in category judgments further supports the notion that participants were unable to acquire meaningful context-specific anchors during the training phase. Overall, these results suggest that the absence of context-specific anchors during the training phase hindered participants' ability to engage in incidental learning, as observed in the original study by Singhal et al. (2021). Participants were unable to logically infer the underlying generative distributions based on the information provided, regardless of the spread of the distribution or the categorization task at hand. This highlights the importance of context-specific anchors in facilitating the inference processes in incidental statistical learning tasks.

Chapter 4

Computational Modelling

4.1 Ideal observer model

To perform model-based analysis on the empirical results seen in experiment 1, we borrow the Bayesian observer models used in Singhal et al. (2021). The most basic or the ideal observer they consider is the one that uses the complete set of training examples to modify the generative model that is assumed to be underlying these contexts.

Let us assume that K is a discrete random variable representing an index for a specific company in experiment 1. Therefore, it can take two discrete values corresponding to each company. Let Y be a continuous random variable that represents the natural numbers that correspond to the amounts of money depicted on the bills of our experiment. So p_K denotes the probability of encountering one particular context (company) in a given trial. And $p_{Y|K}$ denotes the probability of getting the bill value as Y given that the bill is from company K. For a given context, this probability is given by the pdf function $f_{Y|K}$. There are two separate such pdf functions for each context.

During training, participants are shown bills which are explicitly labeled as being from one of the companies. These bill values are generated from the conditional pdfs of their respective companies. Thereby incidentally exposing the participants to values from these underlying conditional distributions $p_{Y|K}$. And across trials, as they are presented with bills from one of the two companies, they observe the frequency of occurrence for each company name or context i.e. p_K.

Then, during the testing phase, participants are shown a specific magnitude Y and asked how probable it is that if the bill item has this magnitude Y, then it belongs to a specific company K. So, given any specific numerical magnitude Y, they are required to give the conditional probability of context K, as $p_{K|Y}$,

The ideal observer would calculate this quantity as follows, using the Bayes rule:

$$p_{K|Y}(k|y) = \frac{p_K(k) f_{Y|K}(y|k)}{f_Y(y)}, \tag{4.1}$$

Here f_Y is a probability density function and is made up of two components corresponding to each context. These individual components for a given context is represented here as $f_{Y|K}$ We consider a single parameter θ that models the p_K as a Bernoulli trial $p_K(\theta)$. And both $p_{Y|K}$ for each context is assumed to be Gaussian distributions $f_{Y|K}(y|\mu^k, (\tau^k)^{-1})$, with parameters μ^k as mean and τ^k as the precision for K^{th} distribution. And p_Y is a combination of these Gaussian distributions weighted by the Bernoulli distribution. The distribution $p_K|Y$ is considered in a tabular form and remains without a precise parametric definition.

Similarly, observers will track the category-specific magnitude distribution,

$$f_{Y|K}(y|k) = \frac{f_Y(y) p_{K|Y}(k|y)}{p_K(k)}, \tag{4.2}$$

and the relevant marginals can be computed over these conditionals,

$$f_Y(y) = \sum_{k'} p_K(k') f_{Y|K}(y|k') \tag{4.3}$$

$$p_K(k) = \int_{y'} f_Y(y') p_{K|Y}(k|y') dy' \tag{4.4}$$

We assume that participants who join the experiment possess some form of Gaussian distribution for $f_Y|K$ and f_Y. And that these distributions are updated as the trial progresses. We also suppose that they assume some Bernoulli for p_K and consider an improper uniform prior on the positive axis for $p_K|Y$ with a value of 0.5. These initial conditions are described using hyper-parameters, in accordance with a generative model that is very similar to the Bayesian Gaussian mixture model present in Bishop (2006), as illustrated in Figure 4.1

When a new value $\{y_{obs}, k_{obs}\}$ is encountered, two sequential updates happen. The y_{obs}, which corresponds to context $K = k_{obs}$, updates the $p_{Y|K}$ distribution for this context.

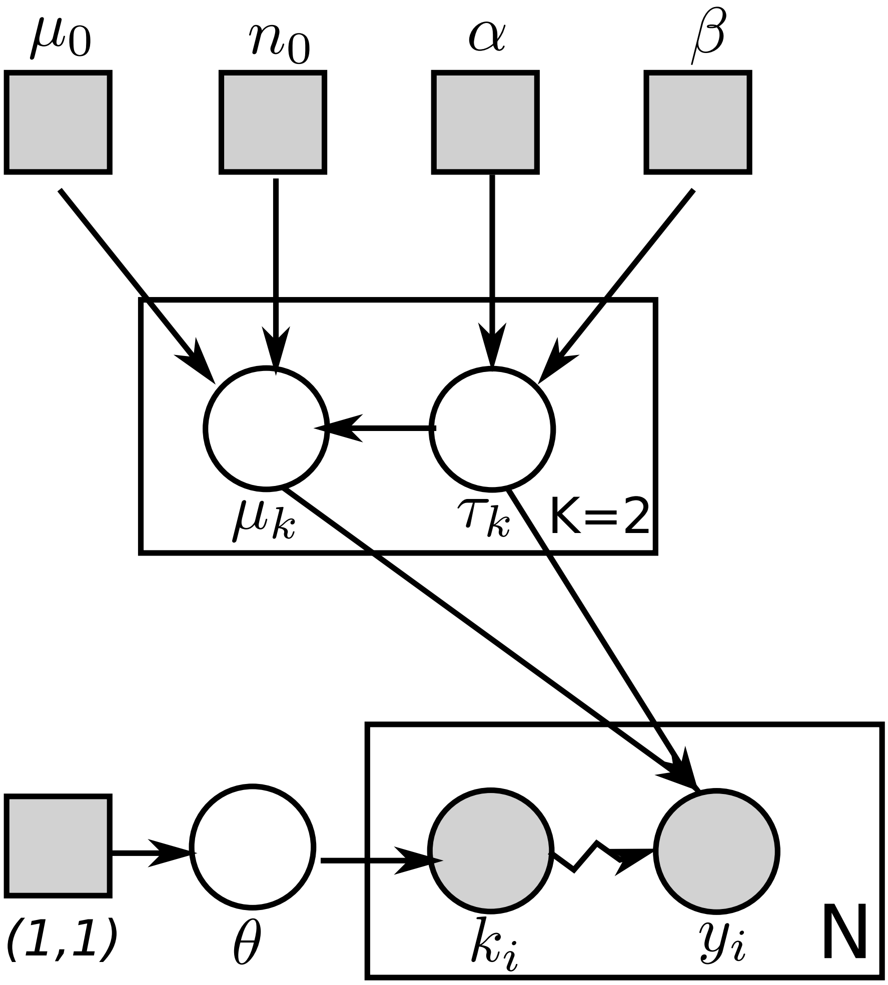


FIGURE 4.1: Generative model for the experimental task in plate notation.

And k_{obs}, in turn, updates the p_k distribution, where the k update is modeled as a Bernoulli update. The prior used for the Bernoulli update is an uninformative prior $Beta(1, 1)$.

This yields the update:

$$\theta \leftarrow \frac{r+1}{n+2}, \tag{4.5}$$

where r is the number of occurrences of category 'expensive' out of n total observations.

We model the $y|k$ as a Gaussian with unknown mean and precision. And it gets updated as a sequential inference using a normal-inverse-Gamma conjugate prior K. P. Murphy (2007).

For each Gaussian component, this inference follows the generative model

$$
\begin{aligned}
y_i|\mu,\tau &\sim N(\mu,\tau),\\
\mu|\tau &\sim N(\mu_0, n_0\tau),\\
\tau &\sim Ga(\alpha,\beta),
\end{aligned}
$$

yielding analytic parameter updates on observing y_i.

The α and β parameters for precision τ are updated as follows:

$$
\begin{aligned}
&\alpha \leftarrow \alpha + \frac{n}{2},\\
&\beta \leftarrow \beta + \frac{1}{2}\sum(y_i-\bar{y})^2 + \frac{nn_0}{2(n+n_0)}(\bar{y}-\mu_0)^2,
\end{aligned}
\tag{4.6}
$$

And the μ_0 and n_0 parameters for mean μ are updated as follows:

$$
\begin{aligned}
&\mu_0 \leftarrow \frac{n\tau}{n\tau+n_0\tau}\bar{y} + \frac{n_0\tau}{n\tau+n_0\tau}\mu_0,\\
&n_0 \leftarrow n + n_0.
\end{aligned}
\tag{4.7}
$$

The ideal observer first updates the hyperparameters $\{\alpha,\beta,\mu_0,n_0\}$ for $f_{Y|K}$ and the hyperparameter θ for p_K. The learning for an ideal observer happens by updating the parameter for p_K using Equation 4.5 and the parameters for $f_{Y|K}$ for the correct mixture component using Equations 4.6 and 4.7 given above. Following this, it updates the marginal f_Y using Equation 4.3, and post that uses Equation 4.1 to estimate $p_{K|Y}$.

4.2 Model-based Analysis

For our experiment, an ideal observer would update the generative model by fetching up each of the 60 instances encountered during training. And it inverts this distribution to provide judgment about a new probe magnitude's category membership. Along with this model, Singhal et al. (2021) also considers two frugal alternatives to this baseline observer model:

1. They consider a *prototype* observer model which considers that users will be able to recall the average of the category distributions shown during the training phase. And use this pair of means to update an abstract generative model of the process in order to provide probability judgments.

2. Next, they consider a modification of the prototype observer model, which considers that people construct prototypes using a subset of samples from stored or recalled observations corresponding to the training phase. Instead of using all the training samples, participants select these prototypes that are then used to update the generative model, similar to the prototype observer model. This is the memory *sampling* extension of the *prototype* observer model.

Apart from these three observer models, in this study, we also consider two additional observer models, which are a form of exemplar models:

1. A memory *sampling* observer model, which considers that people either store or recall just a subset of observations encountered during the training phase. But instead of constructing prototypes with these, the samples are used to update the generative model sequentially, in order to provide probability judgments.

2. We also consider an extension of memory *sampling* observer model, which assumes primacy and recency (PR) effects when people encode or retrieve a subset of training phase observations. The PR effects are modeled with beta priors (with shape parameters $\alpha, \beta \leq 0.5$). These samples are used to sequentially update the generative model, which is then used to construct probability judgments during the testing phase.

In order to provide a detailed understanding of the observer models used in the study, a comprehensive visualization is presented in Appendix (A). This visualization offers insights into the structure and functioning of the observer models, aiding in the interpretation and analysis of the experimental results.

We initially assessed all five models' capacity to account for empirical data from the experiment conducted by Singhal et al. (2021), using a single set of pooled hyperparameters computed across all participants. Then we perform the same analysis for the data from our experiment 1. To analyze the data, we employed the observer models mentioned above and followed a sequential access approach. The observer models were fitted to the data by sequentially presenting them with the 60 original observations that each participant encountered during the training phase of each of the three conditions. This sequential access allowed the models to learn from the observed data and capture the patterns present in the training phase. Once the models were trained using the condition-specific data, we utilized these trained models to construct posterior probabilities, and then predict the outcome for the 40 observations that participants encountered during their test phase. By predicting the posterior probabilities, we aimed to assess how well the models could capture and explain the participants' responses in the test phase. To evaluate the performance of the models, we obtained maximum likelihood estimates for all the model parameters. These estimates represent the values that maximize the likelihood of the observed data given the models, therefore representing the parameter values that best align with the participants' actual responses.

Further, to compare the models' predictions with the observed data, we calculated Bayesian Information Criterion (BIC) scores. The BIC is a measure that takes into account both the goodness-of-fit of the model and its complexity. Lower BIC scores indicate a better trade-off between model fit and complexity. It is important to note that the sampling models, specifically the sampling prototype, sampling, and sampling-pr, have stochastic likelihoods, resulting in varying posterior probabilities. To account for this stochasticity, we conducted 100 simulations for each of these three models. By performing multiple simulations, we aimed to capture the variability in the posteriors emitted by these models.

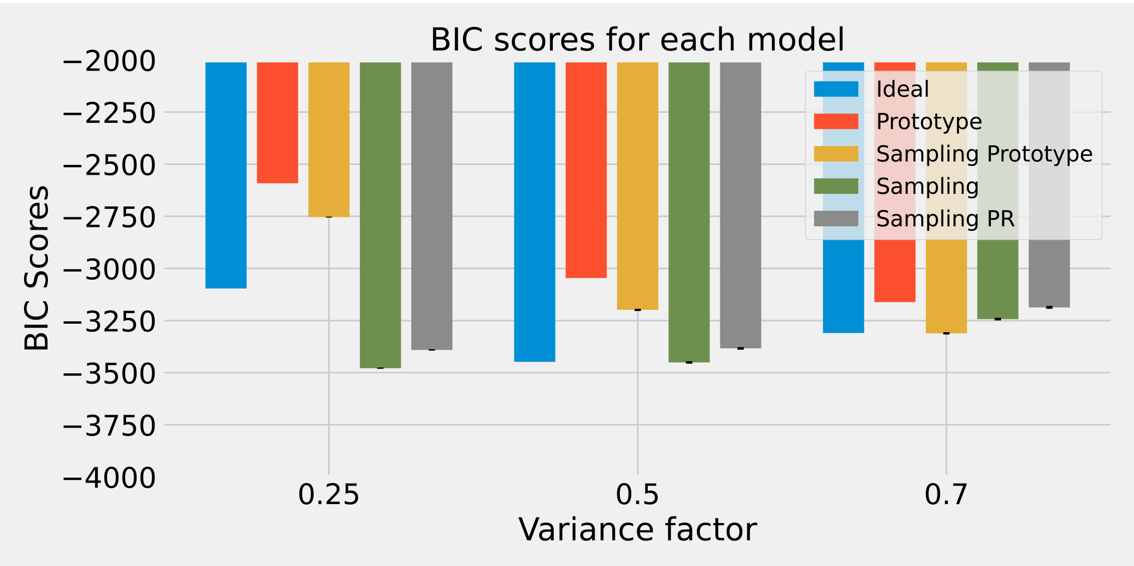


FIGURE 4.2: BIC scores of each observer model for the original experiment from Singhal et al. (2021) paper. SEM across multiple runs is represented with black vertical markers.

Across low-variance training conditions, we find that both alternative sequential sampling models outperform the ideal observer. Thus, at least empirically, it suggests that models that condense the incoming magnitude information into some kind of category prototype representations are better predictors for participants' judgments in our task. These models are better than a fully Bayesian observer that uses from memory all the examples it encountered. We see a high correspondence between the sampling models and participant responses (absolute goodness-of-fit $RMSE = 0.15$ across all three conditions), indicating that humans possess generative models that are broadly appropriate. This generative model is adjusted using a subset of past observations from memory, which then can be inverted to obtain scenario-specific probabilistic judgments.

Next, we examined model fits using individually generated hyperparameters. In Figure 4.3, panel A displays the probability judgment of a participant alongside the ideal observer model's projected posterior distribution. Panel B summarises the median and mean adjusted $R2$ values for participants. This is obtained by training the model using the same values observed by each participant during the training phase. Then, we find the hyperparameters for each participant that maximizes the MLE estimate. The findings indicate that the ideal observer does reasonably well in capturing the variability in people's responses when the distribution has less variability, or in other words, the distributions are more discrete. However, this model becomes less efficient in modeling variability in human responses when the distributions overlap (high variance conditions), making it tougher to discriminate.

Figure 4.3C-D provides the goodness-of-fit for two prototype observer models (prototype and sampling prototype). While figure 4.3E-F shows the goodness-of-fit for two sequential sampling observer models (sampling and sampling-pr). Even by considering only a subset of training data, overall, they afford at-par descriptions of individuals' data when compared to the ideal observer model. This could be because of the introduction of one (sampling model - sample count) and two (sampling-pr model - sample count and primacy/recency prior) additional parameters, which provide the model with extra flexibility.

On the other hand, none of the observer models could explain the variations in participant responses from the testing phase of our experiment 1, as shown in figure 4.4 where hyperparameters are estimated at each participant level, to get median and mean goodness-of-fit for all five model fits across the three spread conditions. This is because the participants were unable to infer the category distributions from the training examples provided to them. And their testing phase responses for this experiment were essentially random in nature, therefore, these could not be predicted reliably by any of our models.

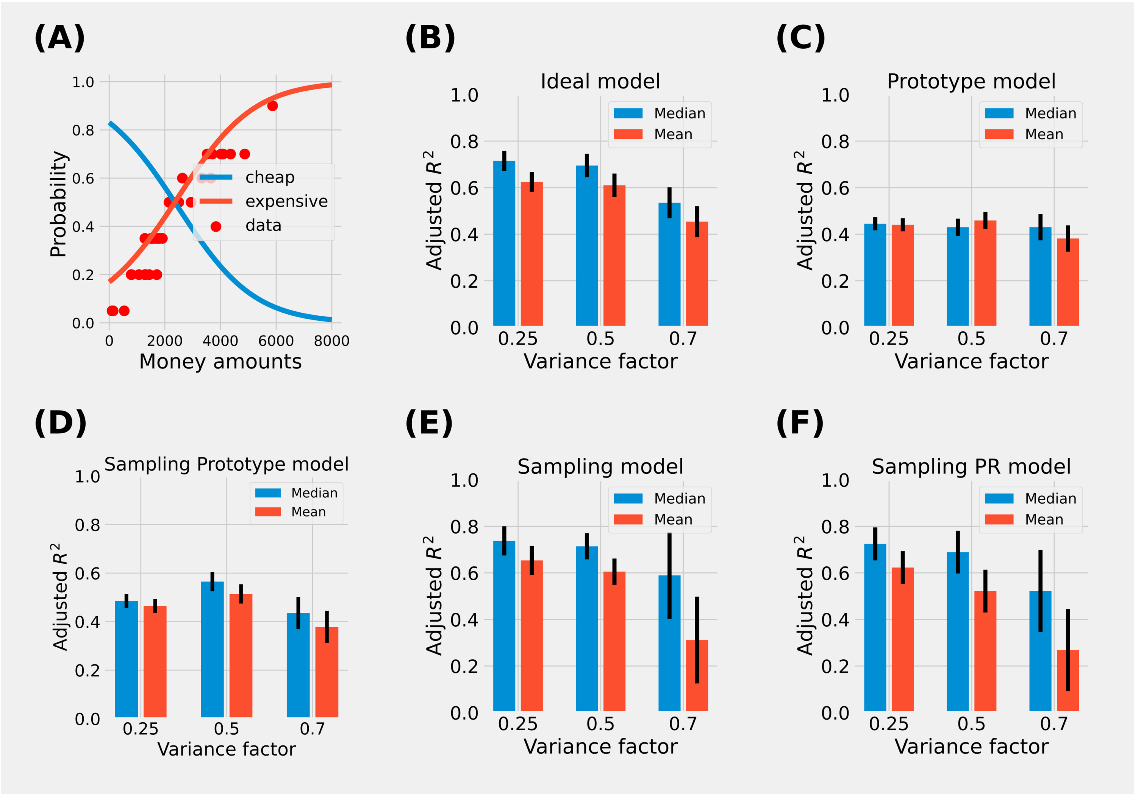


FIGURE 4.3: In panel A, the red dots indicate the test phase responses of a random participant from the original experiment by Singhal et al. (2021). The solid lines are the ideal observer model's fit to this participant's data. Panels B, C, D, E, and F show the median goodness of fit for each variance condition corresponding to the Ideal observer model, Prototype observer model, Sampling prototype observer model, Sampling observer model, and Sampling observer model considering primacy and recency effects (PR) respectively for the same experiment. Error bars show ± 1 s.e.m.

Further analysis of the sampling models (sampling prototype, sampling, and sampling-pr) affords additional insight into people's behavior. Figure 4.5 displays adjusted $R2$ values considering all participants for sampling models when they are provided a certain limited number of observations per category, which are drawn from memory. For each sample count level, the number shown in the graph represents the median of 100 runs of that specific model, which takes into account the stochasticity of memory sampling. As more additional samples are permitted, the model gets better at explaining response variation in participants subjected to high-variance training distributions. Nonetheless, it is crucial to note that for individuals who are assigned to the low variance training condition, the sequential sampling observer models are able to explain over 60% of the variation in people's behavior, even with just one or two samples per category.

But we see no such explanatory powers with any of our models for data from our experiment

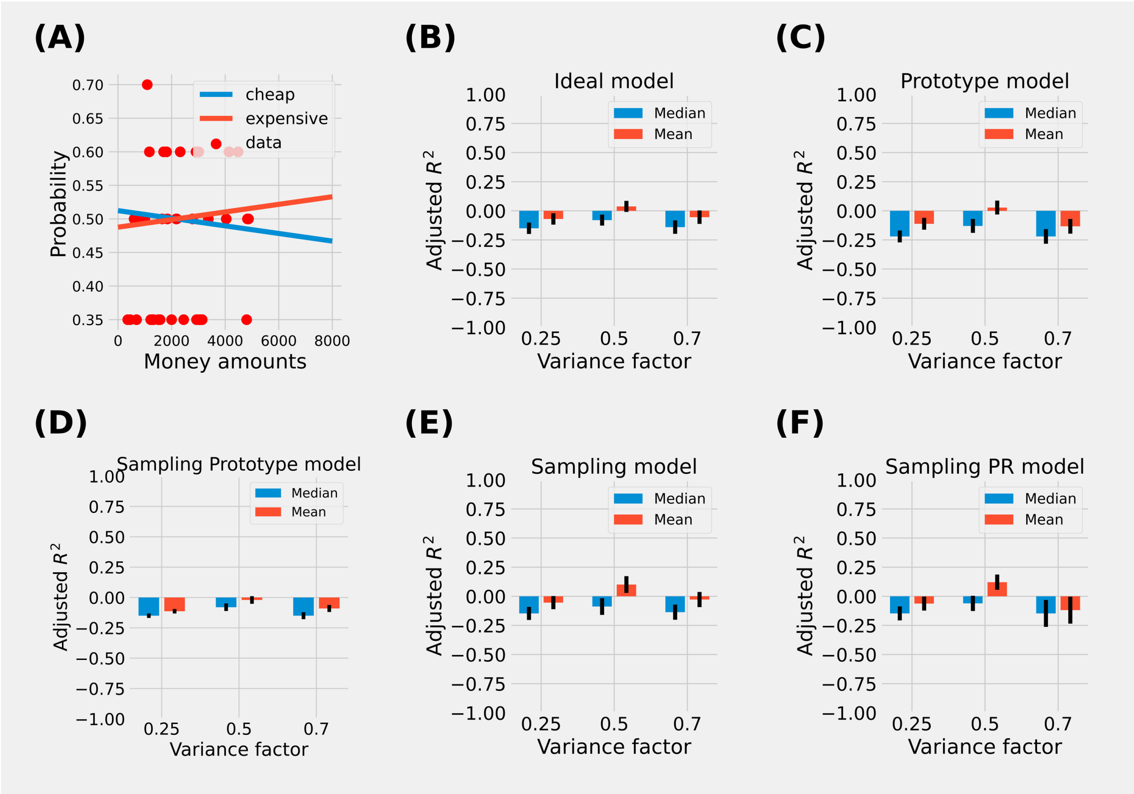


FIGURE 4.4: In panel A, the red dots indicate the test phase responses of a random participant from our experiment 1. The solid lines are the ideal observer model's fit to this participant's data. Panels B, C, D, E, and F show the median goodness of fit for each variance condition corresponding to the Ideal observer model, Prototype observer model, Sampling prototype observer model, Sampling observer model, and Sampling observer model considering primacy and recency effects (PR) respectively for the same experiment. Error bars show ±1 s.e.m.

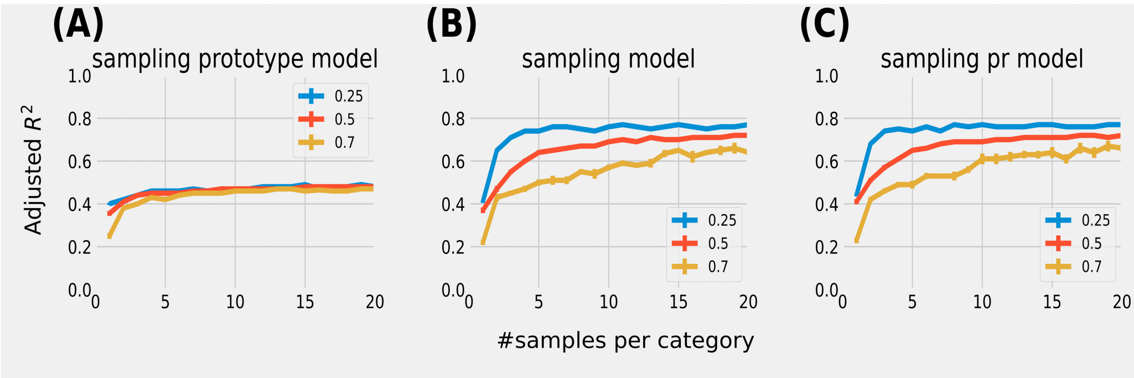


FIGURE 4.5: For original experiment from Singhal et al. (2021) paper - (A) sampling prototype observer model, (B) sampling observer model, (C) sampling observer model considering primacy and recency effects (PR). For each count of observations sampled, median model R^2 values measured for each participant's data across 100 iterations are reported. Error bars show 1 s.e.m. across participants and iterations.

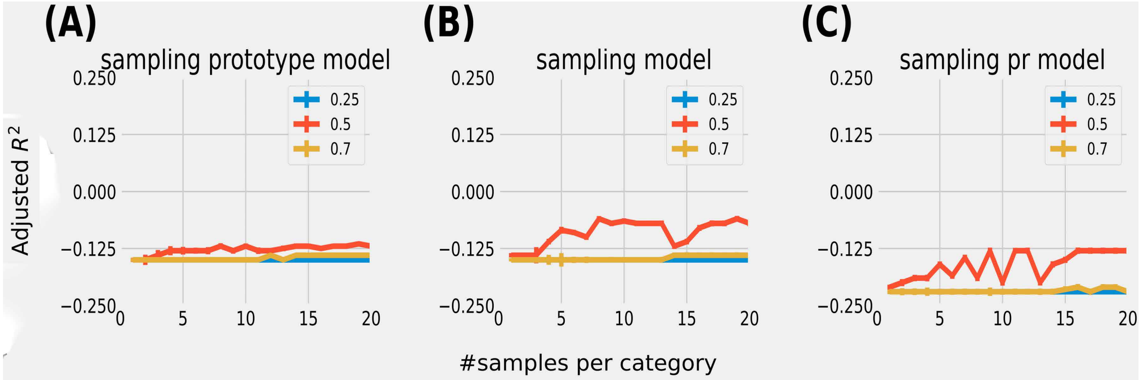


FIGURE 4.6: For the replication study done in our experiment 1 - (A) sampling prototype observer model, (B) sampling observer model, (C) sampling observer model considering primacy and recency effects (PR). For each count of observations sampled, median model R^2 values measured for each participants' data across 100 iterations are reported. Error bars show 1 s.e.m. across participants and iterations.

1. (See Figure 4.6). Even when the observer models are provided with almost all the data points that participants saw during their training phase, it is unable to predict people's behavior in any variance condition. This inability to explain the variance in participants' data, as stated earlier, can be attributed to the random nature of their responses during the testing phase.

Chapter 5

Experiment 2

Although in the earlier study conducted by Singhal et al. (2021) we see how incidental learning of distributions is possible given some category-specific context, in our first experiment, we show how it gets impeded by just removing these context tags, which alluded to ordinal positions of the distribution means. Such mixed results are also seen in previous works in this area Hasher and Zacks (1984); Tran et al. (2017) as discussed in the literature review chapter. Thereupon raising questions about the limits of such statistical learning and, more generally, the question of what makes such kind of category-specific incidental learning of magnitudes possible.

In our investigation of incidental learning of distributions, we propose a hypothesis regarding the mechanism underlying this learning process. We suggest that statistical learning may involve minor adjustments to pre-existing generative models that individuals possess in their minds. These adjustments serve to align the models with the current observed state. However, we speculate that without prior knowledge in the form of causal models, this incidental learning may not occur as effectively as observed in the experiment done by Singhal et al. (2021). To facilitate incidental learning through this hypothesized mechanism, context-specific background information needs to be available to individuals before they are exposed to statistical information. Our hypothesis is rooted in the idea that individuals already possess mental models or representations of the world, which include implicit assumptions about the underlying generative processes. These models help individuals make predictions and judgments based on prior knowledge and experiences. In the context of statistical learning, these models can be considered as priors that individuals use to guide their learning and inference processes. In the original experiment by Singhal et al. (2021), they demonstrated that when participants were provided with the category-specific context in the form of labeled examples, they were able to incidentally learn the underlying distributions of magnitudes. The explicit labels alluded to the ordinal positions of the distribution means, providing participants with a reference point or anchor. This context-specific information helped participants adjust their mental models to align with the observed magnitudes, leading to successful incidental learning. However, in our experiment 1, we saw that the absence of such informative labels curtails such learning. Therefore, we propose that in the absence of prior knowledge or appropriate mental models, the incidental learning process may be impeded. Without pre-existing generative models that incorporate background information, individuals may struggle to calibrate their mental models based solely on statistical information. This lack of calibration could hinder their ability to accurately infer the underlying distributions and make appropriate category judgments.

To test our hypothesis, we need to investigate whether learning occurs when we provide the background information post-exposure to labeled examples. If participants can still learn the distributions even after being given the context-specific background information later, it would suggest the involvement of a different mechanism. One possibility is that participants encode and store the background information and statistical information independently in separate cognitive modules. Later, they combine these pieces of information from memory to generate their probability judgments. Such behavior would deviate from the ideal Bayesian observer, who has generative models of the world based on prior knowledge and sequentially adjusts the parameters of those distributions to account for the observed context values, which is what we hypothesize in this paper.

To test our hypothesis, we designed an experiment with three phases - *training*, *testing*, and *background impartation*. Similar to previous experiments, during the training phase, participants are explicitly asked to perform a cover task in order to expose them to sample values from context-specific probability distributions. But contrary to previous experiments, we also provide participants with context-relevant background information. Finally as usual, in the testing phase, participants are asked to give a probability judgment for a probe magnitude's membership to a specific context. The experiment had four participant groups based on the variations in background knowledge provided to them. Each group was provided with different kinds of background information regarding the contexts, at different stages of the experiment to study its effect on the probabilistic reasoning elicited in the testing phase. For two groups, the context-relevant background information was provided before the training phase, while for the remaining two groups, it was provided after training but before the testing phase(as shown in Table 5.1). Each participant group is also provided with a different type/nature of background information. Two of the groups are provided with just informative labels for the contexts, while the other two groups are additionally provided ordinal positions of the context-specific distributions along with the informative labels as background knowledge (as shown in Table 5.1).

Table 5.1: Number of participants in different experimental groups, along with the kind of background information provided to them. The kind of background information includes the stage/phase at which this is provided to the participants and the type of information (informative labels/Names and ordinal position)

		Background information		
		Stage of impartation	**Type of information**	
Group label	**Number of participants**		**Names as label**	**Ordinal position**
ATNO	25	Ante-Training	Yes	Yes
ATNX	26	Ante-Training	Yes	No
PTNO	25	Post-Training	Yes	Yes
PTNX	25	Post-Training	Yes	No

5.1 Participants

We used a sample size of around 25 participants per experimental group, in accordance with the previous studies. A total of 102 participants took part in the study, as part of four experimental groups. We identified one participant who demonstrated poor performance, with an accuracy rate below 80% in the training phase. Due to this criterion, this participant was excluded from the final analysis. The remaining sample consisted of 101 participants, of which 82 identified as male, 18 as female, and 1 as non-binary. The average age of the participants was 24.4 years, representing a diverse range of individuals. To ensure the validity of our findings, participants were randomly assigned to one of the four experimental groups. The assignment process was designed to distribute participants evenly across the groups and minimize potential confounding variables. The specific assignment of participants to each group can be found in Table 5.1, which outlines the experimental conditions. Each participant's involvement in the study lasted approximately 20 minutes. They completed the designated tasks and responded to the presented stimuli during this time. For their time and participation, participants were remunerated with ₹50.

5.2 Stimuli

The context in our experiments corresponds to the simulated Google Scholar profiles of two scientists. In each stimuli presentation, along with the scientist label, three metrics (i.e., citations, h-index, and i10-index) related to that scientist's performance corresponding to a random country are also shown. The type of scientist label shown to each participant depended upon the experiment group that they were assigned to. For groups that were provided with informative semantic context labels as background knowledge before the training phase (groups 'ATNO' & 'ATNX'), the two profiles were labeled as 'Albert Einstein' and 'Isaac Newton' during the training phase (see sample in Figure 5.1); while for groups which aren't provided with these informative tags before the training phase (groups 'PTNO' & 'PTNX'), the profiles were labeled as 'Scientist 1' and 'Scientist 2' during training (see sample in Figure 5.2).

The values for these performance metrics for each scientist were generated from a context (scientist) specific uni-variate probability distribution. In our case, this is modeled as a normal distribution with mean and standard deviations as shown in Table 5.2 (Also see Figures 5.3 to 5.5). The standard deviations for each metric were the same for both scientists. To maintain minimal overlap, the distribution means were 4 standard deviations apart for task-relevant metrics ('Citations' and 'h-index'), and 10 standard deviations apart for task-irrelevant metric ('i10-index'). The separation for task-irrelevant metric ('i10-index') was increased to 10 standard deviations because no learning was elicited during the pilot run for this experiment where the separation was kept at 4 standard deviations, and people are more effective in learning continuous distribution when they are easily discretized (Srivastava et al., 2014; Tran et al., 2017). The ordinal position of distribution means was also swapped between the task-relevant metrics ('Citations' and 'h-index') to account for any order preferences.

TABLE 5.2: Mean and standard deviation (in brackets) of underlying normal distribution for various performance indicators of each scientist.

Performance Metric	**Scientist 1**	**Scientist 2**
Citation	190,000(20,000)	110,000(20,000)
h-index	75(12.5)	125(12.5)
i-10 index	150(10)	250(10)

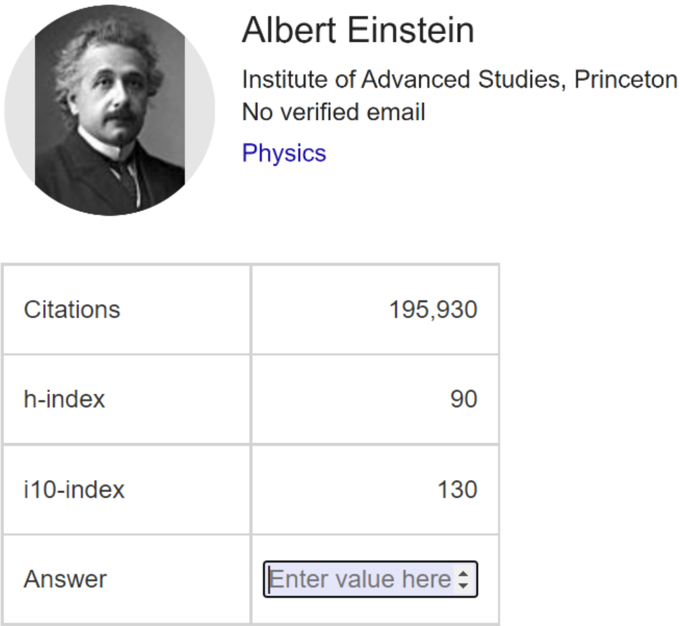


FIGURE 5.1: Sample profile presented during the training phase for ATNO & ATNX experiment groups

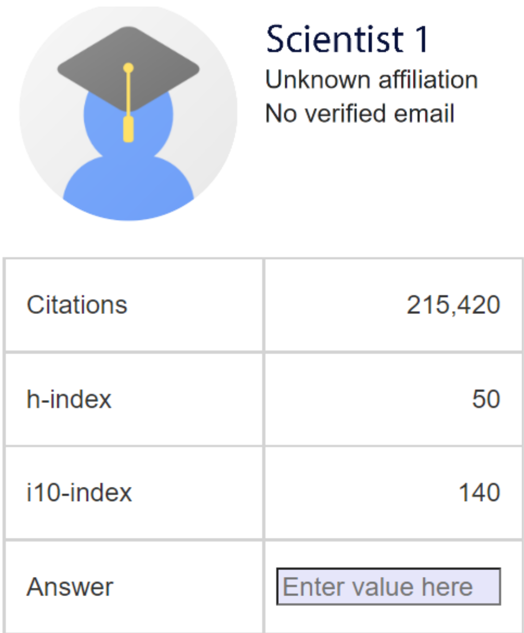


FIGURE 5.2: Sample profile presented during the training phase for PTNO & PTNX experiment groups

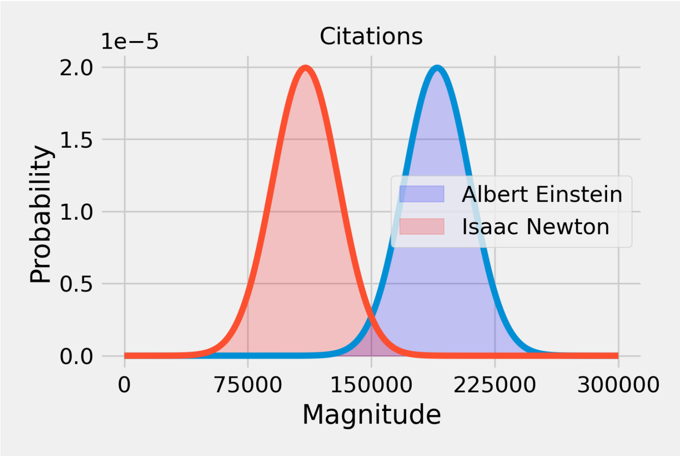


FIGURE 5.3: Visualisation of the underlying generative model corresponding to ’citations’ metric for both scientists

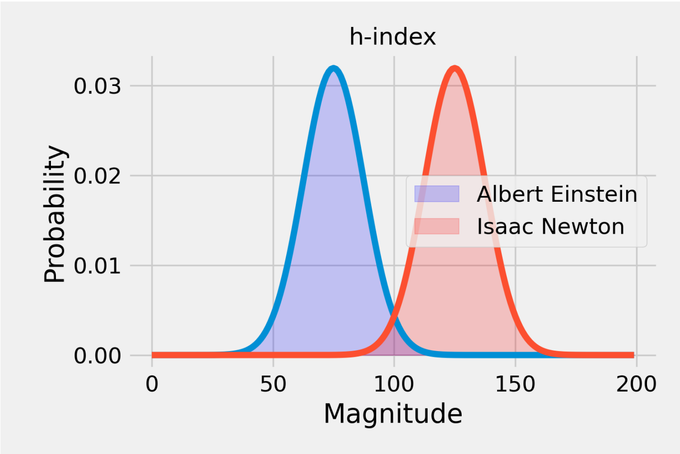


FIGURE 5.4: Visualisation of the underlying generative model corresponding to ’h-index’ metric for both scientists

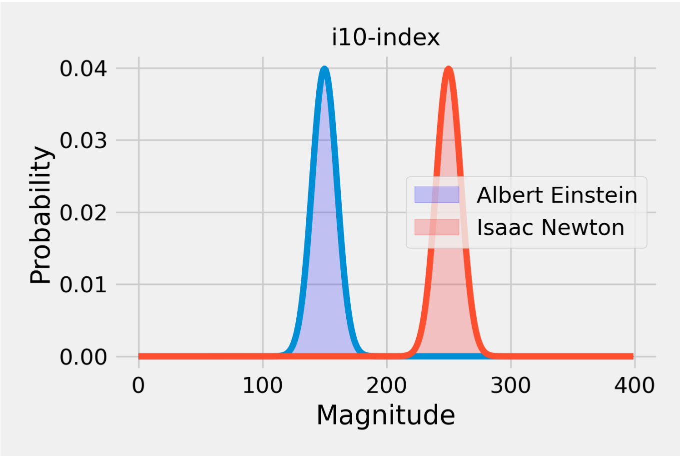


FIGURE 5.5: Visualisation of the underlying generative model corresponding to ’i10-index’ metric for both scientists

5.3 Procedure

5.3.1 Background impartation

The participants are provided with two kinds of background information about the statistical distribution of performance metrics for each context (shown in Table 5.1). First is the informative tags having semantic significance, and instantiated as scientist names - 'Albert Einstein' and 'Isaac Newton', which is provided to participant groups 'ATNO' & 'ATNX' before the training, and for groups 'PTNO' & 'PTNX' it is provided immediately after the training. Due to this reason, during the training phase the profiles are tagged with the names 'Albert Einstein' and 'Isaac Newton' for the former groups, and for the latter ones the tags are just 'Scientist 1' and 'Scientist 2' (see Figures 5.1 and 5.2). To ensure that the participants in groups 'ATNO' & 'ATNX' (which were given meaningful tags) had a conceptual knowledge of the provided contexts, we asked them to rate on a 5-point scale that how familiar they are with these two scientists' names, before the training phase. The other kind of background information is the relative positions of two context-specific distribution means for each performance metric. This ordinal information is provided to participants as *"It was seen that even though Isaac Newton has fewer total citations than Albert Einstein, he has a higher h-index probably owing to the fact that he was a multi-disciplinarian"*. And this is provided only for groups 'ATNO' & 'PTNO' (and not for groups 'ATNX' & 'PTNX') along with the informative labels, that is, before the training for group 'ATNO' and after training for 'PTNO'.

5.3.2 Training Phase

During training, participants were shown 20 Google Scholar profiles, 10 for each scientist, and each stimulus corresponded to their performance with respect to a particular country. These profiles were explicitly tagged as belonging to one of the scientists - 'Albert Einstein' and 'Isaac Newton' for groups 'ATNO' & 'ATNX' while 'Scientist 1' and 'Scientist 2' for groups 'PTNO' & 'PTNX', as seen in figures 5.1 and 5.2. Although this context information/profile labels had no bearing on the task assigned to participants. The profile consisted of 3 performance metrics: two task-relevant metrics ('Citations' and 'h-index') and one task-irrelevant metric ('i10-index'). The explicit (or cover task) given to them was a simple arithmetic task of calculating the number of citations that are not considered for the h-index. (They were instructed to calculate this as the number of 'citations' minus the square of 'h-index'). Participants were given the option to use a calculator if they chose to for this task. As an attention check, we gave them feedback (correct/incorrect) on their replies to each stimulus. Participants who did not do well on this arithmetic exercise (accuracy less than 80%) were removed from the analysis.

5.3.3 Testing Phase

In the subsequent testing phase, participants are retrospectively tested for the incidental learning of the underlying generative model corresponding to performance metrics for each context. On each test trial, participants are shown a single value from one of the three performance metrics, and they are asked to indicate on a 7-point Likert scale how likely it is that the shown metric value corresponds to 'Albert Einstein'. We did not ask to indicate the likelihood for 'Scientist 1' because all groups are provided with the scientist names (informative labels) as background knowledge before the test phase starts. Groups 'ATNO' & 'ATNX' are provided this information before the training phase, while groups 'PTNO' & 'PTNX' are provided immediately after the training and before testing. Each participant faced 60 such trials, 20 for each metric. The values for each metric in the testing phase were drawn from a uniform distribution encompassing at least 95% of the training distributions, as shown in Table 5.3. No feedback was provided during this phase.

Table 5.3: Lower and upper bound of underlying uniform distribution for various performance indicators of each scientist in the test phase.

Performance Metric	**Test values**
Citation	$\mathcal{U}[70000,230000]$
h-index	$\mathcal{U}[50,150]$
i-10 index	$\mathcal{U}[100,300]$

5.4 Results

The psychometric curves were fit separately for each of the three performance metrics - two task-relevant metrics ('Citations' and 'h-index') and one task-irrelevant metric ('i10-index'), by using the 20 Likert scale responses for each metric that were converted to probability judgments as done in the previous experiment. The psychometric function was fit using the average of parameters estimated for participants in experiment 3 for each participant group within each performance metric, i.e., Citations, h-index, and i10-index.

The shading indicates 95% CI for slope parameter estimates. Here, dashed and dotted lines indicate underlying distribution parameters for 'Albert Einstein' and 'Isaac Newton', respectively. The black color denotes the mean of the distribution, while gray denotes 1 standard deviation in either direction. (see Figures 5.6 to 5.8).

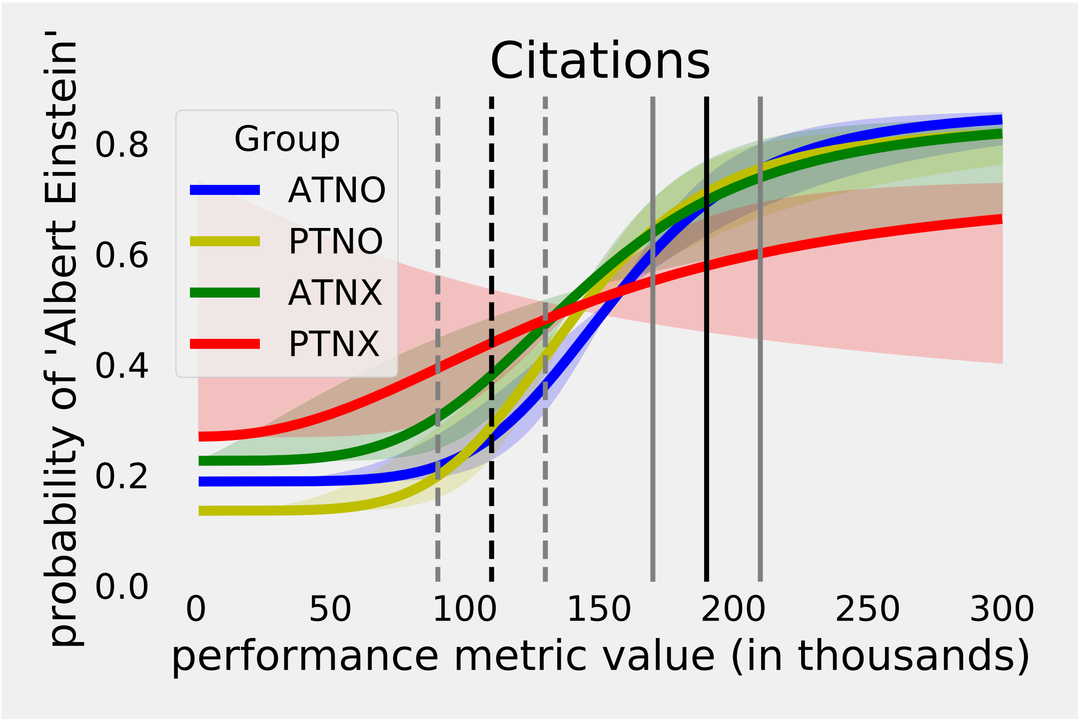


Figure 5.6: Psychometric function fit to participant responses for *citations* metric

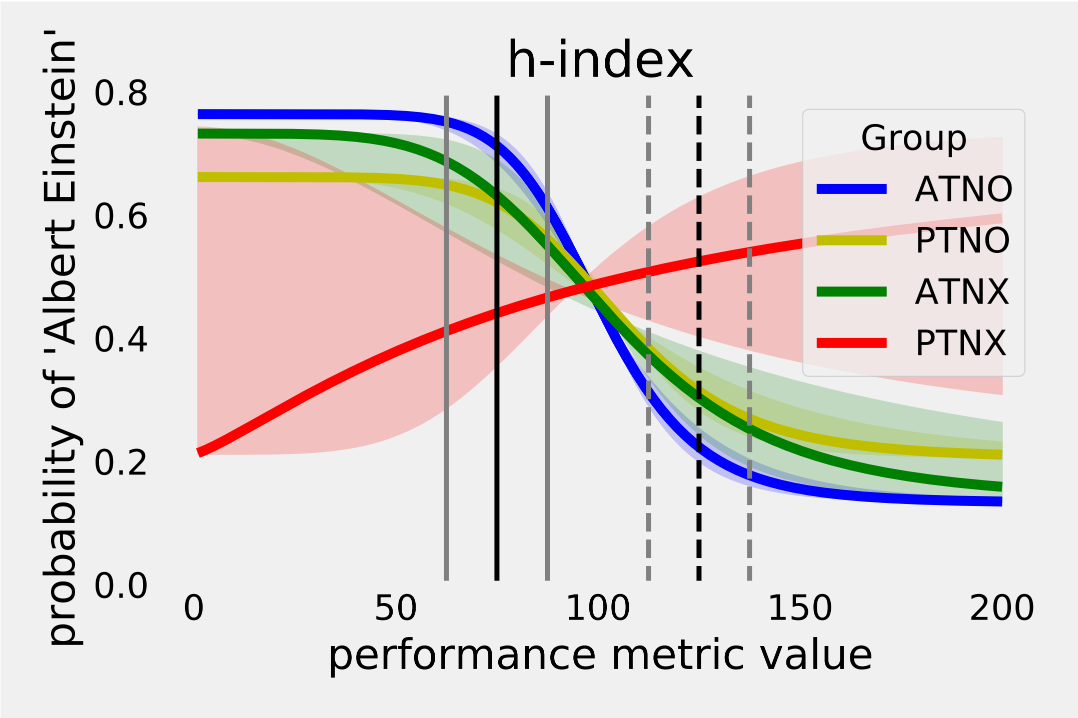


FIGURE 5.7: Psychometric function fit to participant responses for *h-index* metric

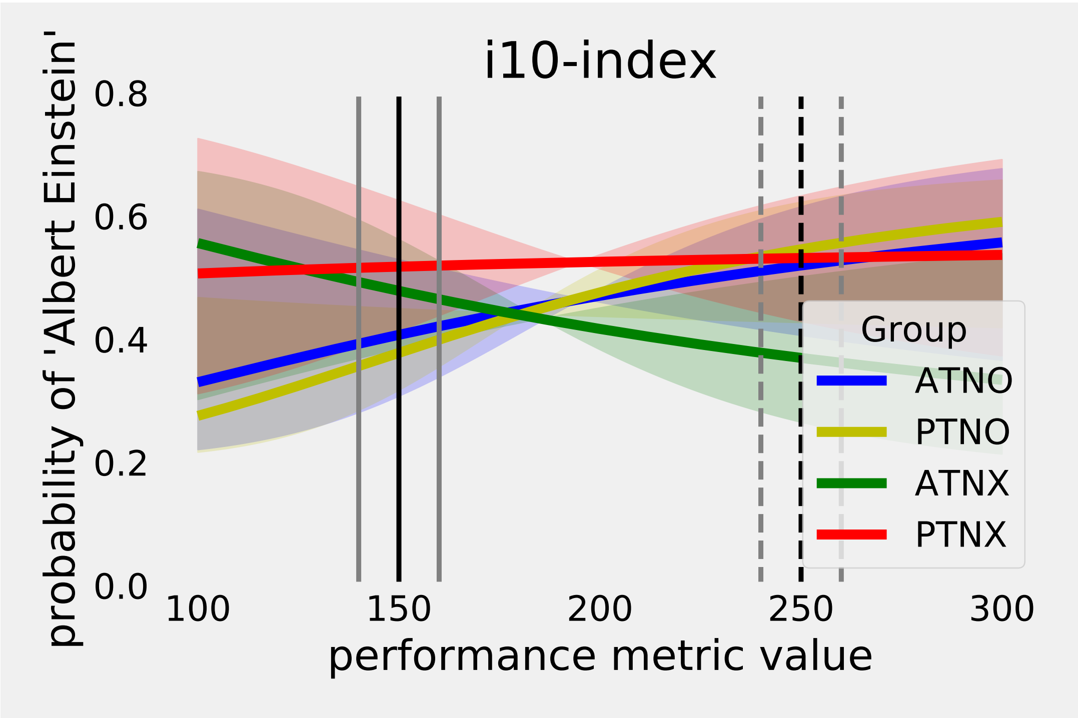


FIGURE 5.8: Psychometric function fit to participant responses for *i10-index* metric

We ran a one-sample T-test on the slopes of fitted curves for each participant group and each metric to check if the slope parameter significantly differs from zero. For the i10-index, which is task-irrelevant, none of the participant groups show slopes that differ from zero (see supplementary material for test statistics). For task-relevant metrics (Citations and h-index), the results show significant slopes with large effect sizes for the group that was given both informative labels and ordinal positions before the training (ATNO(Citations): $t(24) = 5.15, p =< 0.001, d = 1.03$; ATNO(h-index): $t(24) = -10.07, p =< 0.001, d = -2.01$). Similarly, for the group that was just given informative labels as background information before the training, we see significant slopes but with medium effect sizes (ATNX(Citations): $t(25) = 2.83, p = 0.009, d = 0.56$; ATNX(h-index): $t(25) = -3.27, p = 0.003, d = -0.64$). Out of the two remaining groups that were provided background information after the training, the group that was given ordinal positions additionally showed significant slopes with large effect size (PTNO(Citations): $t(24) = 4.47, p =< 0.001, d = 0.89$; PTNO(h-index): $t(24) = -5.45, p =< 0.001, d = -1.09$). While for the last group (PTNX), which wasn't provided any ordinal information and just labels post-training, the slope did not differ significantly from null for any metric (For details of all the test results for each participant group and each metric, see Appendix-C).

The average familiarity on a 5-point scale towards scientist names for group 'ATNO' was 4.52 (SD=0.59), and for group 'ATNX', it was 4.50(SD=0.65). Post-hoc analysis of task-relevant metrics shows that the slope parameter for fitted curves becomes significantly different from zero at the beginning of the test phase itself for groups ATNO and ATNX that were given meaningful tags before the training. But interestingly, for group PTNO, which also had significant slopes but was only given these informative labels post-training, the slope parameter becomes significant only after encountering 5-6 test samples. (see Figures 5.9 and 5.10).

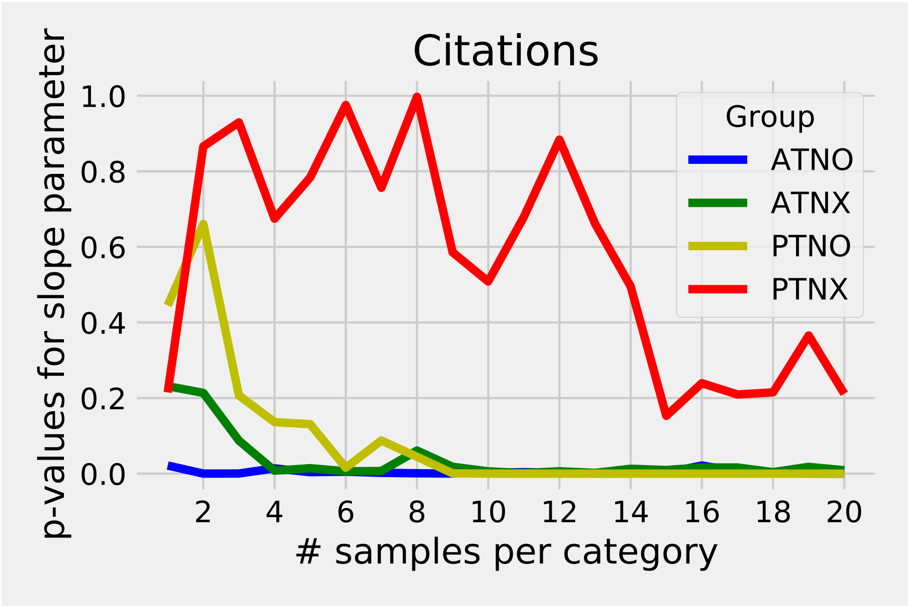


FIGURE 5.9: p-values corresponding to one-sample T-test conducted on slope (beta) parameter of the psychometric function fit considering test samples observed hitherto at each iteration during the testing phase for *citations* metric

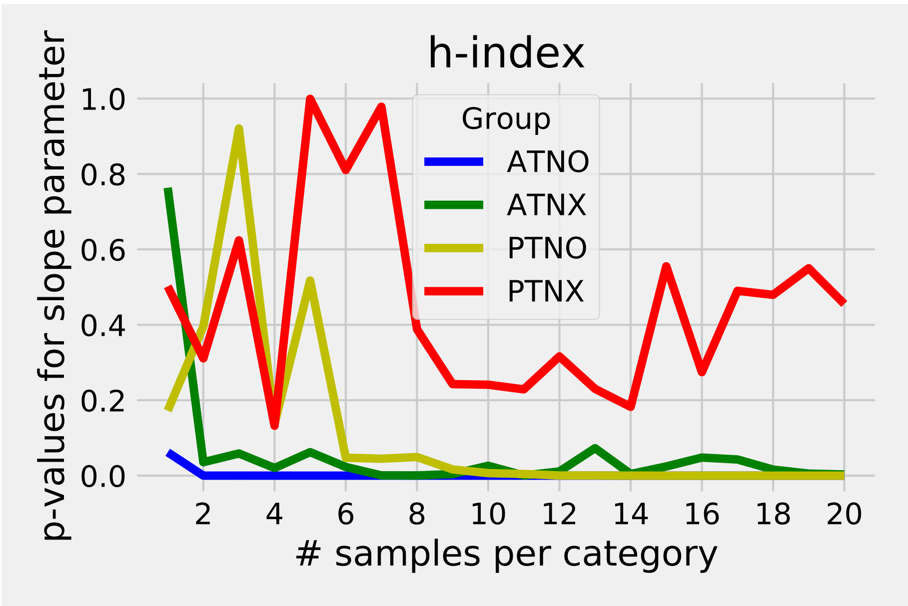


FIGURE 5.10: p-values corresponding to one-sample T-test conducted on slope (beta) parameter of the psychometric function fit considering test samples observed hitherto at each iteration during the testing phase for *h-index* metric

5.5 Discussion

Our findings support the hypothesis that relevant background knowledge is necessary for individuals to learn contextual distributions effectively. We observed significant slope values in participant groups that were provided with context-specific labels before the training phase (ATNO and ATNX). This suggests that when participants had prior knowledge about the context associated with each distribution, they were able to learn the underlying distributions. This finding supports the notion that incorporating relevant background information facilitates the learning process. In contrast, the group that received background information post-training, but only in the form of context labels (PTNX), did not demonstrate any significant learning. This aligns with our hypothesis that the absence of prior knowledge impedes incidental learning. Without the necessary background information, participants were unable to learn the distributions based on the provided statistical information. Interestingly, the group that received ordinal information along with context labels as part of the background information after training (PTNO) showed significant learning. However, post-hoc analysis revealed that this learning only became apparent after encountering approximately 5-6 test samples for each metric during the testing phase. This suggests that participants may have utilized the ordinal information acquired just before the testing phase to divide the observed samples into different ranges corresponding to each context. This behavior aligns with the concept of the ”one-and-known” heuristic (Singhal et al., 2021), where individuals rely on limited information to make judgments or categorizations.

Furthermore, our findings highlight that incidental learning is primarily elicited for task-relevant information and is absent for task-irrelevant information. Even when the context for the task-irrelevant metric was more distinguishable due to the larger separation of distribution means, none of the participant groups were able to track or learn this metric. This underscores the importance of task relevance and the influence it has on incidental learning processes.

Chapter 6

General Discussions

The research conducted by Singhal et al. (2021) offers compelling empirical evidence of incidental supervised category learning, highlighting the human capacity to acquire categories with minimal exposure. This discovery reinforces the concept that individuals are adept at engaging in Bayesian inversion, wherein they can engage in probabilistic reasoning and adjust their beliefs based on statistical data, even when this information is incidentally presented. It aligns with the longstanding notion in Bayesian cognitive science that humans possess an inherent inclination to store and leverage critical statistical information about real-world events, even in scenarios where explicit motivation for such actions may be lacking. The primary objective of our study is to delve further into the intricate mechanisms underpinning such incidental statistical learning in the human cognitive framework. Our goal is to find how individuals can incidentally assimilate statistical information, and to identify the factors that influence the efficacy of such a learning process. In doing so, we aim to shed light on the cognitive processes and underlying assumptions that enable individuals to glean insights from statistical data. A central focus of our study is to delineate the boundaries and constraints of incidental statistical learning. Through an examination of the circumstances conducive to successful incidental learning and those leading to its failure, we aim to comprehensively understand the conditions that either facilitate or impede incidental learning in humans.

In our first experiment, we sought to replicate a prior study's observation of incidental statistical distribution learning, as outlined by Singhal et al. (2021). However, we introduced a deliberate modification aimed at eliminating any presumptive knowledge regarding the contexts in which the distributions were presented. This manipulation was designed to render conditional priors uninformative, thereby hindering participants' access to the observation's likely generative distributions. Despite exposing participants to several labeled examples, our results revealed a notable absence of learning under these modified conditions. This outcome suggests that in the absence of access to generative models governing conditional probabilities, individuals struggle to incidentally acquire context-specific distributions. This also hints at the crucial role of the availability of relevant background knowledge to the individuals, particularly the generative model that they believe is responsible for shaping the observed conditional probabilities. The nature of this information significantly influences the capacity for incidental learning. In situations where the relevant generative models are absent, individuals face significant challenges in grasping and internalizing context-specific distributions.

In order to further investigate the impact of background knowledge on incidental statistical learning, a follow-up experiment (Experiment 2) was devised. The primary objective of this experiment was to systematically modulate the availability of background information to the participants, allowing a comprehensive exploration of its effects on the learning process. This was accomplished through a two-fold approach. Firstly, by controlling the timing and sequence of background information delivery. This manipulation aimed to scrutinize the temporal dimension's role in shaping incidental learning. In doing so, we assessed whether the order of presentation had an influence on participants' capacity to acquire and assimilate statistical knowledge. Secondly, the content of the background information itself was intentionally varied, thus regulating the type of information accessible to participants regarding the experimental context. This variation enabled us to investigate the specific elements within the background information that either facilitated or impeded the incidental learning process. The comprehensive control over the nature and presentation order of background information allowed for a systematic exploration of the driving factors behind incidental statistical learning.

The results obtained from our experiment 2 underscore the pivotal role played by the availability of generative models, which was provided in this study as background information, in shaping participants' ability to acquire knowledge of contextual distributions. Notably, participants demonstrated the capability to learn contextual distributions when they had access to these generative models before being exposed to labeled examples during the training phase. In contrast, for participants who did not have access to these generative models because the background information was provided after the training phase, they were unsuccessful at learning the underlying statistics, despite encountering numerous contextual examples. This challenges the alternative hypothesis, which posits that incidental statistical learning operates independently of background knowledge acquisition and that both modules can function autonomously, accessible at a later point for retrospective reasoning. Instead, findings strongly support the notion that generative models, in the form of background knowledge here, function as a foundation and are a necessary element in the learning process. Access to this generative model during the observation of labeled examples allows participants to build upon it and adapt the model based on the contextual data they encounter. According to our observer models, participants likely possess the ability to modify this generative model based on the observed data to minimize prediction errors, ultimately enabling the learning of contextual distributions. Furthermore, an insightful aspect emerged from the responses of participants who received background knowledge (in the form of ordinal information about the distributions) after the training phase but before the testing phase. In such cases, participants demonstrated learning; however, notably, this learning commenced only after encountering a few samples from the testing phase. This intriguing response suggests that participants utilized their initial probe magnitudes encountered during the test phase as surrogates for training examples to adjust the recently acquired generative model (post-training phase) to fit the specific context. While these magnitudes are not explicitly labeled examples, one possibility is that participants used them to estimate a range for these distributions. Combining this information with their knowledge of the order of these distributions, participants could devise an effective strategy for the remaining test phase. In conclusion, these results decisively affirm that access to generative models, represented in our experiments as background knowledge, is imperative for the learning of contextual distributions. Furthermore, the timing of this access emerges as a critical factor, emphasizing that participants must have access to these generative models before incidental exposure to labeled examples to effectively engage in the learning process.

While prior access to generative models is undoubtedly a necessary condition for incidental learning, it is essential to emphasize that it alone is not sufficient. Incidental statistical learning is also significantly influenced by the content of the background knowledge, which can facilitate learning in two distinct ways. Firstly, the content of the background knowledge can enhance learning by providing additional insights into the nature or properties of the generative models. For instance, background knowledge may explicitly mention the shape of the distribution or offer insight into the relative ordering of contextual distributions. This knowledge serves to constrain the degrees of freedom governing the underlying statistics, providing valuable information in limiting the possibility space and effectively reducing uncertainty. This phenomenon is observed in specific groups within Experiment

2, where participants were provided with ordinal information regarding the underlying distributions. This information made participants aware of the specific dimension along which distributions were ordered and informed them of the relative positions of these distributions along this dimension. Armed with this information, participants were able to adjust their generic models to align with the actual statistical distributions encountered during the training phase. It's noteworthy that providing this additional information explicitly is not the only pathway to its acquisition. Background knowledge about the underlying distributions could also be derived from pre-existing knowledge, such as semantics or certain beliefs. For instance, the experiment conducted by Singhal et al. (2021) demonstrated that knowledge about the relative positions of distributions was likely acquired through the nomenclature assigned to each context, such as 'cheap' and 'expensive.' This nomenclature inherently carried ordinal information about the underlying distributions and was possibly utilized by participants to inform their generative models.

Additionally, the content of background knowledge can also enhance learning by rendering the provided contexts more salient to the observers. This increased saliency facilitates the harvesting of relevant generative models from individuals' existing theories and expectations regarding the world. However, to extract these meaningful models effectively, there must be conceptual coherence between the distribution-specific contexts presented to individuals and their pre-existing knowledge. In essence, there should be a meaningful connection or familiarity between the information presented in the learning environment and individuals' existing understanding of the world (G. L. Murphy & Medin, 1985). These generative models or informative priors supply pertinent and valuable insights into the current contexts. These priors serve to highlight significant aspects or features within the context that can influence the underlying statistical distributions. They accentuate the principal dimensions along which these distributions can be most effectively distinguished. This form of learning is demonstrated in Experiment 2 by the group that did not receive ordinal information but was provided with informative tags for each context before the training phase. Within this group, participants relied on the informativeness of the context tags to guide the formulation of potential models for the underlying data. They developed informative priors based on their pre-existing knowledge about these scientists. The salience of scientist names in this experimental group likely assisted them in identifying the primary metrics by which they could be evaluated. Then, during the subsequent training phase, they modified their priors based on the samples of these metrics they encountered for each scientist.

Furthermore, our results shed light on a critical aspect that certain metrics or dimensions of stimuli, which do not contribute to the primary task at hand, remain unlearned, regardless of the specific experimental conditions. In Experiment 2, we specifically examined the learning of metrics or dimensions that held no relevance to the intentional task. Our results demonstrated that participants exhibited no significant learning of these irrelevant dimensions. This held true regardless of the experimental conditions, background knowledge availability, or even the distinguishability of contextual distributions. It appears that participants exercised selective focus on dimensions pertinent to the task, while effectively disregarding the irrelevant ones. This observation finds resonance with the channel capacity model as described within the realm of information theory (Rabbitt, 1968). The channel capacity model sets a finite limit to the quantity of information that a channel can process and transmit. In the context of cognitive processing, this implies that individuals have a constrained capacity for processing and acquiring information. Consequently, only those dimensions or metrics of stimuli that are relevant to the task are effectively learned, while irrelevant dimensions are either disregarded or not learned. An important caveat to consider here is the substantial impact of cognitive load in such scenarios. This alignment with the channel capacity model reinforces the idea that there exists a limit to information processing capability. Participants adopt a strategy wherein they prioritize and allocate their cognitive resources to process and acquire dimensions directly tied to their task objectives, while omitting or dismissing those dimensions deemed extraneous. Within the framework of information theory, this phenomenon can be understood as a deliberate optimization of cognitive resource allocation. By eschewing irrelevant dimensions, participants enhance the efficiency and effectiveness of their learning process, concentrating exclusively on the information indispensable to the task at hand.

In contrast to prior literature, which predominantly exemplifies incidental learning in perceptual tasks, such as those explored by Turk-Browne et al. (2005) and some language tasks illustrated by Saffran et al. (1997) that primarily entail the replication of observed samples, our study delves into a broader form of statistical learning. The type of learning elucidated in this investigation is akin to the one probed in a previous study by Tran et al. (2017). It emphasizes that mere exposure to samples from distributions is insufficient for effective learning, and underscores the necessity of a foundational causal model upon which such learning can take place. In these tasks, participants are not merely tasked with recalling cumulative frequencies; instead, they are actively engaged in probabilistic reasoning and making judgments based out of their updated beliefs derived from the inversion of conditional distributions. Importantly, it's crucial to note that our findings do not serve to invalidate existing literature on incidental learning. Rather, they contribute an added layer of complexity and nuance to our understanding of cognitive judgment tasks, offering a deeper and more comprehensive understanding of these statistical learning mechanisms.

Our experiments show that when participants are provided with specific properties of the underlying generative models, or when the given context is pertinent, or even better, when both conditions are met, we observe the occurrence of context-specific incidental learning. In such cases, participants already possess the appropriate generative model for the current scenario. And during the training period, they acquire the parameters of the context-specific distributions that underlie the training data. The process of learning in this context-specific incidental learning scenario involves calibrating the parameters of existing prior beliefs to best fit the observed data sample. In other words, participants modify their existing mental representations or models to align with the statistical patterns present in the training data. By adjusting the parameters of these generative models, participants are able to optimize the fit between the model and the observed data. This calibration process allows them to capture the key characteristics and distributional properties of the training data accurately. As participants progress through the training period, they continually update the prior distribution based on the observed examples. The resulting posterior distribution, which represents their updated beliefs or probabilities about the context-specific distributions, is then used to make probability judgments corresponding to the current situation. These probability judgments are based on the learned generative models, which have been adjusted to match the observed data. By utilizing the posterior distribution, participants can estimate the probability of certain events or outcomes within the context, allowing them to make informed judgments about the test scenarios.

Conversely, when the background information available to participants lacks the two crucial forms of content discussed above, namely the properties of the underlying generative models and context-specific saliency, we observe a complete absence of learning. This is evident in both Experiment 1 and Experiment 2. In Experiment 1, the lack of any informative company tags or distribution-specific information hindered the learning of context-specific parameters. Without these key pieces of information, participants were unable to establish a clear connection between the contextual cues and the underlying statistical distributions. As a result, their ability to learn and adjust their generative models based on the observed data was significantly impaired. Similarly, in Experiment 2, the group of participants who were neither provided with any context-specific labels before the training phase nor any ordinal information after training, they exhibited no learning. Without the prior knowledge or background information about the context-specific distributions, participants lacked the necessary foundation to form accurate generative models. Consequently, even after being exposed to multiple labeled exemplars during the training phase, their behavior was essentially random in nature. The reason for this lack of learning in such cases can be attributed to the absence of correct prior distributions. Prior distributions play a crucial role in Bayesian inference as they provide the initial beliefs or expectations about the parameters of the generative models. Without the correct prior distributions, participants are unable to accurately update their beliefs and establish the correct posterior distribution. As a result, the posterior distribution, which represents participants' updated beliefs about the context-specific distributions, becomes distorted or inaccurate. The lack of accurate posterior distributions leads to flawed probability judgments and decision-making processes. Consequently, participants' behavior becomes random or arbitrary, as they lack the necessary information to make informed and contextually appropriate judgments.

In conclusion, our study has yielded crucial insights into incidental statistical learning. We have shown that the mere observation of categorical examples is insufficient to produce a generative model, indicating that such a learning process is not purely unsupervised, and that causal models do not emerge solely from statistical information. Our study has substantiated that the incidental acquisition of observed frequencies is contingent upon the pre-existence of context-specific causal relationships within participants' cognitive schemata. This antecedent knowledge must be firmly in place before any exposure to labeled data, thereby serving as the foundational substrate upon which the assimilation of observed statistical information can effectively occur. The thematic congruence between the generative model and context-specific examples facilitates them to mesh seamlessly together, permitting the Bayesian update mechanism to refine participants' beliefs and enable rational reasoning within the current context. While our study underscores the pivotal role of a precise generative model in elucidating statistical learning, it simultaneously beckons a fundamental question: How do individuals initially acquire these generative models? What delineates the array of scenarios in which individuals possess such generative models? Moreover, What are the conceivable forms of these causal models and their defining characteristics? This study signifies that substantial research endeavors are still warranted to comprehensively understand the acquisition, properties, and utilization of these generative models that individuals possess. An in-depth inquiry in this domain is indispensable to further unravel the intricacies of human cognition and furnish a more profound understanding of the cognitive framework within which these generative models operate.

Appendix/Supporting Information

Additional supporting information and appendices may be found at this project's OSF link.

Bibliography

1. Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current directions in psychological science*, *21*(3), 170–176.

2. Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

3. Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 151–155.

4. Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris*, *98*(1-3), 249–258.

5. Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

6. Gibson, J. J. (1966). The senses considered as perceptual systems.

7. Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological science*, *17*(9), 767–773.

8. Hancock, P., & Volante, W. G. (2020). Quantifying the qualities of language. *PloS one*, *15*(5), e0232198.

9. Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American psychologist*, *39*(12), 1372.

10. Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1264.

11. Johnson-Laird, P. N., & Wason, P. C. (1977). *Thinking: Readings in cognitive science*. CUP Archive.

12. Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and brain sciences*, *34*(4), 169.

13. Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to jonathan cohen. *7*(4), 409–411.

14. Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

15. Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, *9*(11), 1432–1438.

16. McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

17. Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, *92*(3), 289.

18. Murphy, K. P. (2007). Conjugate bayesian analysis of the gaussian distribution. *def*, *1*($2\sigma 2$), 16.

19. Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature neuroscience*, *16*(9), 1170–1178.

20. Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly journal of experimental psychology*, *20*(3), 241–248.

21. Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*.

22. Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

23. Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science*, *8*(2), 101–105.

24. Singhal, I., Srinivasan, N., & Srivastava, N. (2021). One and known: Incidental probability judgments from very few samples. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

25. Srivastava, N., Vul, E., & Schrater, P. R. (2014). Magnitude-sensitive preference formation. *Advances in neural information processing systems*, 1080–1088.

26. Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, *53*(1), 1–26.

27. Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, *331*(6022), 1279–1285.

28. Tran, R., Vul, E., & Pashler, H. (2017). How effective is incidental learning of the shape of probability distributions? *Royal Society Open Science*, *4*(8), 170270.

29. Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in cognitive sciences*, *12*(8), 291–297.

30. Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552.

31. Wattenmaker, W. D. (1999). The influence of prior knowledge in intentional versus incidental concept learning. *Memory & cognition*, *27*(4), 685–698.