COG250Y1: Introduction to Cognitive Science

Jaehyeon Park

Fall 2023 - Winter 2024

Course Description

An introduction to the problems, theories and research strategies central to the interdisciplinary field focusing on the nature and organization of the human mind and other cognitive systems. Interrelations among the philosophical, psychological, linguistic and computer science aspects of the field are emphasized.

Disclaimer

Please note that this document may not cover all the material from the lecture, and some information may contain errors. If this is the case, I may have been falling asleep or not paying attention during that lecture. Unless necessary, I will not update the document, meaning the content is provided as-is, and I am not responsible for any inaccuracies or consequences resulting from its use. It is strongly recommended to cross-check the material with the current lecture materials.

Special thanks to:

- **Dominic Le** for substantial suggestions
- **Dominik Kucz** for massive explanations
- **Tiana Chan** for bringing her friends

Contents

1 Introduction to Cognitive Science **3**

2 Resemblance and Classical Theory of Categorization **5**

3 Prototype Theory and Its Critiques **8**

4 Micro Theory and the Problems with Classical and Prototype Theories **10**

5 Memory: Accuracy and Relevance **12**

6 Memory and Problem Solving **14**

7 Rationality, Intelligence, and Problem Solving **17**

8 Insight and Memory: The Search Inference Framework **19**

9 Attention and Insight: Beyond the Spotlight Metaphor **21**

10 Procedural Knowledge and Insight **24**

11 Introduction to the History of AI **26**

12 The Formal System and Computation **28**

13 Algorithmic Systems and Cognitive Science **29**

14 Cognition Beyond Language: Challenges to GOFAI **31**

15 Challenges in Cognitive Science: Connectionism vs. GOFAI **33**

16 Criteria for Cognition and the Mind-Body Problem **34**

17 The Fall of Dualism and the Rise of Materialism **37**

18 Relevance Realization and the Limitations of Computational Functionalism **39**

1 Introduction to Cognitive Science

1.1 Levels and Disciplines

Cognitive science explores the mind by integrating insights from several disciplines. Each discipline interprets the mind from its own perspective:

- **Neuroscience**: focuses on the brain as the physical basis of cognition.
- **Artificial Intelligence (AI)**: views cognition through the lens of machine learning and information processing.
- **Psychology**: investigates behavior through experimental methods.
- **Linguistics**: examines language and its role in cognitive processes.
- **Anthropology**: studies cognition within cultural contexts.

However, cognitive science faces several key challenges, such as equivocation, causal ignorance, and fragmentation, which demand a cohesive framework that integrates diverse disciplines.

1.2 The Role of Philosophy in Cognitive Science

Philosophy plays a vital role in cognitive science by addressing foundational questions and building bridges between disciplines. It does so through:

- **Ontology**: the study of the structure of reality, which helps unify different vocabularies used by each discipline.
- **Epistemology**: the study of knowledge, which helps connect diverse methodologies.

Philosophical insights provide the tools to address the issues of cognitive science by fostering integration.

1.3 Three Visions of Cognitive Science

There are three primary approaches to cognitive science, each with its own strengths and limitations:

1. **General Nominalism**: Treats cognitive science as a set of loosely connected disciplines without much integration. This approach does not address the fundamental problems in cognitive science, such as equivocation or causal ignorance.
2. **Interdisciplinary Eclecticism**: Disciplines interact to learn from each other, enhancing mutual understanding. However, it lacks the deep integration required to truly advance cognitive science, and there is no strong bridging between them.

3. **Synoptic Integration**: Aims for a deep connection between disciplines where they challenge and change each other. This approach seeks strong relationships between disciplines and represents the ideal form of cognitive science.

The goal of cognitive science is to achieve synoptic integration, which forms a unified theory of the mind that addresses its complexity.

1.4 Problems in Cognitive Science

Cognitive science must address several key problems to develop a unified theory:

- **Equivocation**: The same term, such as "mind," is used differently across disciplines, leading to confusion. This problem arises when a word has different meanings depending on the context in which it is used.
- **Causal Ignorance**: Although disciplines are interconnected, their causal relationships are often ignored. It is plausible that these relationships are crucial to understanding the mind, yet they are often left out.
- **Fragmentation**: The study of cognitive science is fragmented, leading to a scattered and incomplete understanding of the mind. This fragmentation prevents the development of a cohesive theory.

1.5 Bridging Between Disciplines

To integrate cognitive science, we rely on conceptual tools like metaphors. These help bridge gaps between disciplines by providing apt conceptual vocabularies. For effective integration, we need:

- **Multi-Aptness**: A metaphor must apply across multiple domains, allowing transfer of meaning while respecting differences. This concept is also referred to as elegance, where a theory describes multiple aspects of the mind with simplicity.
- **Convergence**: A theory must be supported by different arguments but should not be too trivial. If too much convergence occurs, the theory becomes overly simplistic and loses its explanatory power.

The balance between elegance and convergence is crucial for a theory to be plausible. A plausible theory must not only be integrative but also avoid being too far-fetched or trivial.

1.6 Naturalistic Imperatives

To firmly ground cognitive science within the scientific worldview, we follow the naturalistic imperatives:

1. **Analyze**: Break down complex phenomena into simpler terms. This is the first step towards creating a comprehensive understanding of cognitive processes. For instance, Thales' claim that all things are made of water represents an early form of analysis.

2. **Formalize**: Transform mental terms into non-mental, formal terms. Descartes, for example, invented the Cartesian plane, formalizing spatial concepts into a system of coordinates. This step also helps avoid the Homuncular Fallacy, where complex concepts are explained by simply reusing the original concept without proper definition.

3. **Mechanize**: Go beyond formalization to mechanize the process of explanation. This step, exemplified by Turing's work on computational theory, allows full access to the inner workings of cognitive processes. Mechanization formalizes and explains complex processes in a way that can be systematically understood.

1.7 Conclusion

In Conclusion, cognitive science seeks a plausible theory of mind that integrates insights from various disciplines through synoptic integration. By balancing multi-aptness and convergence, we develop metaphors and frameworks that bridge gaps between fields. The naturalistic imperatives—analyze, formalize, and mechanize—anchor cognitive science in the scientific worldview, providing a path toward a unified understanding of the mind.

2 Resemblance and Classical Theory of Categorization

2.1 Introduction to Categorization

Categorization is a fundamental process in cognitive science that allows us to simplify complex information by grouping entities based on shared characteristics. According to the *naturalistic imperatives*, we aim to refine broad concepts, such as intelligence, into smaller, more manageable components through analysis. Categorization serves as the underlying process that enables coding and decoding of information.

Psychologically, a category is a class of entities that we mentally sense belong together. While some entities are treated as individuals, many must be categorized to simplify our interactions with them. Categorization allows us to control, manipulate, and replace category members without expending excessive time and attention. This process enables generalization and inferential connections—once an entity is categorized, we can anticipate its behavior rather than encountering it as a completely new instance. This anticipation highlights the importance of categorization in our cognitive processes.

2.2 Inferences in Categorization

Categorization is closely tied to three types of inference:

- **Deduction**: If the premises are true, the conclusion must be true.
- **Induction**: If the premises are true, there is a high likelihood that the conclusion is also true.
- **Abduction**: The process of inferring the best possible explanation for a given set of facts.

These inference methods allow us to make sense of invisible entities by drawing conclusions based on categorization patterns.

2.3 Theories in Categorization

2.3.1 Scott Atran's View

Scott Atran explored how people across different cultures categorize objects and concepts. He argued that common sense leads us to categorize by taking what is familiar and using it to explain the unfamiliar. In science, however, this process is often reversed. In Atran's view, we mentally group similar objects together to form categories, and categorization occurs because these objects share certain similarities.

2.3.2 Nelson Goodman's Critique

Nelson Goodman challenged the notion of categorization based purely on similarity, arguing that this approach leads to *equivocation*. He explained that two objects may share some properties, making them appear similar, but similarity itself can have two different meanings:

- **Logical Similarity**: Objects are similar if they share many, but not all, properties. However, this approach faces the problem that any two objects can be said to share an infinite number of properties, meaning that everything could be considered similar to everything else.
- **Psychological Similarity**: Here, categorization relies on judgments of relevance—deciding which properties are important to compare. This type of similarity, however, often falls into the *Homuncular Fallacy*, where the concept being explained is assumed without proper definition.

2.3.3 Resemblance Theory

Resemblance Theory was introduced by Smith, who argued that similarity generates categorization. According to this theory, if two objects are perceived as similar, they are categorized together. Tversky attempted to formalize this idea with a *similarity equation*:

$$sim(I, J) = af(I \cap J) - bf(I - J) - cf(J - I)$$

where a, b, and c represent the importance attached to each comparison, and f stands for salience. However, this formalization has been critiqued for falling into the same problems of equivocation and relevance.

Lance Rips argued against Resemblance Theory, suggesting that categorization drives similarity judgments rather than the other way around. He presented the concept of *double dissociation*, where categorization and similarity judgments can vary independently. This challenges the idea that similarity necessarily leads to categorization.

Smith responded to Rips' critique by distinguishing between two types of categorization:

- **Similarity-Based Categorization**: This is a bottom-up process where categorization is based on visible features that entities share.
- **Reasoning-Based Categorization**: This is a top-down process that relies on knowledge and belief. In this approach, we categorize entities based on reasoning, even if some features are not directly observable.

Despite these distinctions, Resemblance Theory still faces challenges, particularly in its formalization and reliance on inductive generalizations.

2.4 Classical Theory of Concepts

In contrast to Resemblance Theory, the *Classical Theory of Concepts* views categorization as the process of forming mental definitions. These definitions are based on finding necessary and sufficient features of a concept, known as the concept's *essence*. Classical Theory claims that categorization occurs by defining entities according to a conjunctive list of attributes that must be both necessary and sufficient.

Classical Theory solves many of the problems associated with Resemblance Theory, as it focuses on essential properties rather than superficial similarities. However, this theory has also been challenged. For example, Wittgenstein questioned whether necessary and sufficient conditions are always enough to define a category.

2.5 Six Characteristics of Classical Theory

Classical Theory posits six key characteristics:

1. The meaning of a concept can be captured by a conjunctive list of attributes.
2. The features of a concept are atomic and cannot be broken down further.
3. Each feature is individually necessary and jointly sufficient for categorization.
4. The boundaries between what is and what is not a member of a category are clearly defined.

5. All members of a category are equally good representatives of that category.

6. Categories can be explained through hierarchical taxonomy, using transitive inference.

Despite these principles, Classical Theory is not without its own challenges, as later discussed by philosophers like Wittgenstein.

2.6 Conclusion

Categorization is a fundamental cognitive process that allows us to manage and simplify the vast amount of information we encounter. Various theories have been proposed to explain how we categorize, from Resemblance Theory, which emphasizes similarity, to Classical Theory, which focuses on defining essential features. While each theory has its merits, both face challenges in explaining how we group entities together in the mind. As cognitive science continues to evolve, so too will our understanding of the processes behind categorization.

3 Prototype Theory and Its Critiques

3.1 Problems of Classical Theory

Wittgenstein raised several issues with classical theory, noting that it struggles with transitivity and fails to account for certain categories. Classical theory suggests that categories are defined by necessary and sufficient features, but Wittgenstein argued that not all categories fit this model. For example, the category of "games" has no clear essence or defining features, yet we intuitively recognize it as a valid category. This critique laid the groundwork for the development of alternative theories, including prototype theory.

3.2 Introduction to Prototype Theory

Eleanor Rosch proposed prototype theory as a response to the limitations of classical theory. Instead of defining categories based on strict features, prototype theory suggests that categories are organized around typical features. According to this theory, some members of a category are more representative or "typical" than others. For example, a robin might be considered a more typical bird than a penguin, even though both are classified as birds.

Prototype theory also addresses several problems that classical theory could not explain:

- Why some concepts do not have a clear essence.
- Why not all members of a category are equally representative.
- Why category boundaries are often fuzzy or unclear.

3.3 The Five Claims of Prototype Theory

Prototype theory rests on five key claims:

1. Concepts have prototype structures. In the *exemplar camp*, the prototype is the best existing example of a concept, while in the *prototype camp*, it is a set of typical features.

2. There are no necessary or sufficient features to determine category membership. This explains why some concepts do not have a clearly defined essence.

3. Concept boundaries are unclear and fuzzy. Family resemblances hold concepts together, so some members exist in a grey zone.

4. Members are rated for typicality based on a *typicality gradient*. The higher an entity is on this gradient, the more typical features it has.

5. Membership is determined by similarity to a prototype with a finite number of features. The fewer features an entity shares with the prototype, the less typical it is.

3.4 Prototype Theory and the Naturalistic Imperatives

One of the main questions surrounding prototype theory is whether it fits within the naturalistic imperatives of cognitive science. The creation of prototypes is tied to frequency, which introduces elements of resemblance theory. In addition, typicality depends on frequency, requiring an explanation of similarity before prototypes can be established. This process implies psychological similarity, as proposed by Nelson Goodman. However, determining the causation of similarity remains difficult to explain within the framework of prototype theory.

3.5 Problems of Prototype Theory

While prototype theory offers many advantages over classical theory, it has its own issues. Some of the key problems are:

1. It cannot explain how to generate a prototype, as noted by Medin and Shoben.

2. It struggles to explain how smaller concepts are integrated into larger ones.

3. Armstrong, Gleitman, and Gleitman criticized the experimental protocols used to test prototype theory.

4. It faces challenges with mechanization, particularly when dealing with interactions between features.

3.6 Contextual Variability and Typicality

Roth and Shoben demonstrated that typicality ratings are not stable across contexts. For example, is milk more typical of a beverage than tea? The answer often depends on the context and the individual making the judgment. This variability suggests that prototype theory struggles to maintain the consistency required for categorization across different situations.

Medin and Shoben further showed that prototype theory relies on circular explanations. Their research on spoons found that small metal spoons were rated as more typical than large wooden spoons, but people were familiar with both. The theory could not explain why certain features were more salient in different contexts, relying instead on psychological similarity.

3.7 Mechanization Challenges

A significant criticism of prototype theory is its difficulty with mechanization. Unlike classical theory, which defines categories based on necessary and sufficient features, prototype theory depends on typical features. This introduces a threshold for categorization, which does not account for interactions between features and often leads to issues with boundary conditions. The theory also struggles to evaluate the truth of quantified statements, especially when combined with fuzzy set theory.

3.8 Conclusion

Prototype theory offers a compelling alternative to classical theory by focusing on typical features rather than strict definitions. However, it faces several challenges, particularly in explaining how prototypes are generated, how concepts are integrated, and how variability in typicality is handled across contexts. Despite these issues, prototype theory remains an influential approach to understanding how we categorize the world around us.

4 Micro Theory and the Problems with Classical and Prototype Theories

4.1 Problems of Classical and Prototype Theories

Both classical and prototype theories face significant challenges. Classical theory struggles with transitivity and fails to account for certain categories that do not possess necessary and sufficient conditions. For example, subcategories and supercategories may not always be related—such as the relationship between car seats and chairs or chairs and furniture. Wittgenstein pointed out these issues, highlighting how classical theory could not always explain the hierarchical structure of concepts.

Similarly, prototype theory has its limitations. One of the main issues is the reliance on similarity, which becomes problematic when trying to generate

a prototype. The typicality gradient, which measures how "typical" an entity is within a category, is unstable across contexts. For example, people may rate certain spoons as more typical based on subjective preferences, meaning that typicality is not always tied to concrete concepts.

Additionally, prototype theory predicts linearly separable judgments, where relationships between features can be explained straightforwardly. However, people often make nonlinear judgments, which prototype theory cannot easily account for.

4.2 Productivity and Compositionality

The idea of productivity—how we can generate an infinite number of ideas from a finite number of elements—plays a central role in understanding conceptual theory. In classical theory, productivity is explained through compositionality: combining a finite number of basic elements and rules can create an indefinite number of new concepts. This is why classical theory excels in explaining productivity.

Prototype theory, however, does not provide a clear explanation for compositionality. Prototypes, as wholes, cannot be broken down into the typical features of their parts. For example, the prototype of a car cannot be constructed by simply combining the prototypes of its components like wheels and seats. This inability to explain compositionality limits the scope of prototype theory.

4.3 Gestalt and the Problem of Feature Lists

Both classical and prototype theories share two presuppositions:

1. Concepts are built out of feature lists. However, this is problematic because it assumes features are well-defined and discrete.

2. Concepts serve a descriptive function, primarily labeling the world rather than explaining how entities behave.

Gestalt theory challenges these presuppositions by suggesting that "the whole is greater than the sum of its parts." In other words, concepts do not just consist of isolated features; they have a structural-functional organization that allows the features to fit together. This organization helps predict and explain how entities behave in the world.

4.4 Conceptual Combination and Micro Theory

The idea of conceptual combination suggests a functional relationship between abstract and concrete features. Concepts are not merely lists of unrelated features; rather, they are small theories—called micro theory. Micro theory explains aspects that classical and prototype theories cannot, such as how entities behave and how concepts interact with one another.

Micro theory provides a more dynamic approach to categorization by addressing the structural organization of concepts. It suggests that concepts are smaller theories that work together to explain both the abstract and concrete levels of features.

4.5 Problems with Micro Theory

Despite its advantages, micro theory also faces challenges. One major issue is the problem of underdetermination: every theory will eventually be contradicted by a more plausible one. Additionally, selecting features for a concept is not always straightforward, and categorization may be influenced by memory and context.

While micro theory addresses many of the shortcomings of classical and prototype theories, it must contend with these problems, particularly when it comes to mechanization and explaining how features interact.

4.6 Memory and Categorization

Micro theory posits that categorization may be based on memory. Intuitively, memory is often thought of using a spatial metaphor. According to Eysenck and Keane, memory functions like a stable object placed in a location, with related memories stored nearby. When recalling information, we search these memories and retrieve what is relevant.

However, experiments dating back to 1932 suggest that memory is not as stable as previously thought. Memory is not simply reproductive but reconstructive: when we try to remember something, we do not accurately preserve the past but instead reconstruct it in a way that makes sense based on what is meaningful.

4.7 Conclusion

Micro theory provides a more flexible and explanatory approach to categorization, offering solutions to many problems in classical and prototype theories. Micro theory offers a new perspective on how we understand and categorize the world by considering the structural-functional organization of concepts and the role of memory in categorization. However, like any theory, it is not without its challenges, particularly in terms of mechanization and feature selection.

5 Memory: Accuracy and Relevance

5.1 The Manipulation of Memory

Memory is not as reliable as it may seem and can be easily manipulated based on context. Loftus and Zanni (1975) demonstrated this through an experiment in which participants were shown videos of car crashes. The wording of the questions significantly impacted the participants' recollections. For instance,

when asked, "Did you see a broken headlight?" versus "Did you see the broken headlight?" the latter question implied the presence of a broken headlight, leading participants to falsely remember seeing one. This experiment revealed how easily human memory can be influenced by suggestion.

Further studies questioned the reliability of eyewitness testimony, particularly in cases where memory is recovered under hypnosis. Putnam (1979) modified the Loftus experiment and found that hypnotized participants showed increased incorrect confidence in their false memories, highlighting that memory under hypnosis is not necessarily accurate.

These experiments underscore the idea that memory is not a perfect record of events, but rather a reconstructive process, influenced by context, suggestion, and other external factors.

5.2 Working Memory and Its Limitations

Working memory is bound to consciousness and perceptual experience, but it has limitations, particularly in its capacity. Miller (1956) proposed that working memory can hold around four chunks of meaningful information at a time, a phenomenon known as chunking. However, chunking only describes the phenomenon; it does not explain how working memory operates.

Two primary theories attempt to explain working memory:

1. Miller's bottleneck theory: Working memory is limited in capacity, creating a bottleneck between sensory input and long-term memory.

2. Hasher's relevance filter theory: Working memory is not a bottleneck but a filter that helps prioritize relevant information. This theory suggests that working memory is key to fluid intelligence, as it allows us to focus attention in complex or novel situations.

Despite these theories, working memory remains a topic of debate, and it is generally agreed that working memory cannot be improved through training.

5.3 Long-Term Memory and Meaning

Unlike working memory, long-term memory (LTM) focuses on encoding meaning rather than the surface form of information. Sachs (1967) demonstrated this in an experiment where participants were asked to recall sentences they had read. While they often struggled to remember the exact wording, they easily recalled the meaning of the sentences. This suggests that LTM stores meaning and discards surface details.

Bransford and Frank (1971) extended this research by investigating how meaning is tied to propositions, or truth claims. They found that participants were more confident in remembering sentences with more propositions, indicating that meaning plays a central role in memory retention.

5.4 Memory as an Integration Process

Craik and Lockhart's levels of processing theory proposes that the more integrated information is with existing knowledge, the more likely it is to be remembered. For example, shallow processing tasks like identifying whether a word is in capital letters are less likely to result in long-term retention than deeper tasks, such as determining whether a word fits meaningfully into a sentence.

However, Anderson and Pichert (1978) showed that perspective also plays a crucial role in memory. In their experiment, participants read a story from different perspectives, and when they were asked to recall the story from an alternate perspective, they were able to retrieve additional information. This suggests that memory is not solely about integrating meaning but also about relevance to a particular situation.

5.5 Relevance in Memory

Memory, particularly long-term memory, appears to be organized around relevance. Tulving and Thomson's theory of transfer-appropriate processing suggests that information is best remembered when the encoding and retrieval processes align with the problem-solving situation. In other words, the relevance of the information to the task at hand determines how well it is remembered.

This idea is further supported by the concept of encoding specificity, which explains that memory retrieval is influenced by the context in which the information was learned. The more relevant the context, the easier it is to recall the information.

5.6 Conclusion

Memory is not just about recalling facts; it is a reconstructive and relevance-driven process. While working memory is limited and serves as a filter for relevant information, long-term memory focuses on encoding meaning and integrating information. Both memory systems are essential for problem-solving and future prediction, highlighting the adaptive nature of memory in human cognition.

6 Memory and Problem Solving

6.1 Challenges to Levels of Processing

The levels of processing (LOP) theory, while influential, faces both theoretical and empirical challenges. Eysenck pointed out a circular problem: there is no independent measure of meaning or depth, and the theory relies heavily on vague intuitions. The word "meaning" itself is ambiguous in cognitive science, encompassing everything from truth conditions and relevance to personal significance.

Empirically, Morris, Bransford, and Franks (1977) demonstrated that the LOP framework can be problematic. In their modified experiment, participants were more likely to remember words that had procedural relevance, such as rhyming with a target word, rather than words with deeper meaning. This result suggests that relevance plays a key role in memory retention, not just meaning.

6.2 Relevance and Memory

The concept of relevance has become increasingly important in understanding memory. Transfer appropriate processing suggests that how we learn information (its relevance) must align with how we later retrieve it. This process is essential for problem-solving. Anticipation and relevance realization are crucial for predicting future needs and guiding memory retention.

Encoding specificity further explains how memory functions by suggesting that external context and internal states are stored together. For instance, if a bad event triggers sadness, the brain is more likely to remember other sad events, while failing to recall happy moments. Encoding specificity and transfer appropriate processing converge to explain how memory anticipates relevant information for future use.

6.3 Memory and Problem Solving

Potter proposed a three-stage model for memory retrieval:

1. Large amounts of information are retrieved, much of which is irrelevant.
2. A selection process identifies relevant material.
3. The selected relevant material is passed to consciousness.

This model highlights the close relationship between memory and problem solving. Memory helps filter through irrelevant information to identify the most pertinent data for solving a problem.

6.4 The Interconnection Problem

Cognitive science faces the challenge of integrating theories of categorization, memory, and problem solving into a cohesive framework. These theories often point to one another, creating what is known as the interconnection problem. Solving this issue requires an integrated account that addresses all three processes in a parallel fashion.

6.5 Analyzing Problem Solving

Intelligence is closely tied to problem solving. However, intelligence is not just about solving one type of problem—it's about being able to solve many different

kinds of problems. A general problem solver is someone who can navigate a wide variety of challenges, and this ability is the hallmark of intelligence.

Newell and Simon's General Problem Solver (GPS) project aimed to develop a system that could solve any well-defined problem. They analyzed problem solving through four key elements:

1. The representation of the initial state (IS).
2. The representation of the goal state (GS).
3. Operators or operations that change the IS and move toward the GS.
4. Path constraints that guide the search space without derailing the solution process.

A problem solution, in this framework, is any sequence of operations that transforms the IS into the GS while obeying the path constraints.

6.6 Types of Problem Solving

There are two primary approaches to problem solving:

1. **Algorithm**: A systematic technique that guarantees a solution or proves that a solution is impossible. Algorithms work well under conditions of certainty but often require exhaustive searches of the entire search space, which can lead to combinatorially explosive searches.
2. **Heuristics**: These are rules of thumb that improve the chances of solving a problem by limiting the search space. Heuristics pre-judge where to look, introducing biases but reducing the computational load. Examples include availability heuristics (estimating probability based on what comes to mind easily) and representativeness heuristics (estimating likelihood based on prototypical examples).

6.7 Flawed Assumptions of the General Problem Solver

The General Problem Solver project made two flawed assumptions:

1. All problems are fundamentally the same, with the same underlying structure.
2. Problem formulation is a trivial matter.

In reality, most real-world problems are ill-defined, lacking clear IS, GS, operators, or path constraints. Formulating the problem is often the most difficult step, as seen in examples like the chessboard and 9-dot problems, which require insight to transform ill-defined problems into solvable ones.

6.8 Intelligence and Problem Formulation

Intelligence is not just about applying algorithms or heuristics—it involves the ability to reformulate problems and realize what is relevant. Moments of insight allow us to escape ill-defined problems and avoid combinatorially explosive searches. The relevance realization process is what makes human problem solving adaptable and effective in a wide range of contexts.

6.9 Conclusion

Memory and problem solving are deeply interconnected processes that rely on relevance, anticipation, and insight. The ability to filter irrelevant information, recognize patterns, and reformulate problems is key to both intelligence and effective problem solving. While algorithms and heuristics provide structured approaches, it is the human capacity for relevance realization and insight that drives true problem-solving success.

7 Rationality, Intelligence, and Problem Solving

7.1 Rationality versus Intelligence

Rationality is distinct from intelligence because intelligence is often considered a fixed trait, whereas rationality can be developed or improved. The General Problem Solver (GPS) project, which aims to model problem solving, ties into psychometric approaches to intelligence, especially Spearman's theory of the general factor of intelligence, or g.

Spearman divided intelligence into:

- **Fluid Intelligence**: The raw ability to solve novel problems, requiring explanations like pattern detection (eduction) and mental energy (processing power).
- **Crystallized Intelligence**: The ability to use accumulated knowledge to solve problems. However, relying solely on crystallized intelligence to explain intelligence is problematic.

Two interlocking meta-problems for problem solvers are anticipation and relevance realization. The more an individual can anticipate future problems and realize what is relevant, the more adaptive and intelligent their behavior becomes. These processes are intertwined; without anticipation, relevance realization would be unnecessary, and vice versa.

7.2 Newell and Simon's GPS and Heuristics

Newell and Simon, in their work on the General Problem Solver (GPS), explored various heuristics to determine what makes individuals fluidly intelligent. They examined two main heuristics:

- **Hill Climbing**: This heuristic presupposes an evolutionary mechanism where feedback indicates whether one is getting closer to the goal. However, it is limited by the possibility of getting stuck at a local maximum.
- **Means-Ends Analysis**: This heuristic involves working backward from the goal state (GS) to the initial state (IS). The process involves:
  1. Noticing salient differences between IS and GS.
  2. Choosing an operation to reduce the difference.
  3. Applying the operation.
  4. If sub-goals are not achieved, returning to step 1.

  While effective, means-ends analysis is limited because it cannot be formalized, leading to the homuncular problem—who decides what is salient?

7.3 Gestalt Psychology and Problem Solving

Gestalt psychology provides an alternative view of problem solving, focusing on aspect shifts and insight. Kohler's experiments demonstrated that problem solving involves a learning curve, with moments of insight that allow individuals to reformulate poorly defined problems into solvable ones. This contrasts with Thorndike's law of effect, which suggests that learning happens through trial and error.

Weisberg and Alba (1981) tested the idea of "thinking outside the box" by giving participants the 9-dot problem. They argued that insight, as proposed by Gestalt theory, is about breaking fixation. However, when participants failed to solve the problem, Weisberg and Alba concluded that insight does not exist. They assumed there was only a single difficulty in the problem, but in reality, there are multiple factors at play.

7.4 Different Kinds of Knowing

Weisberg and Alba's work highlighted the importance of different types of knowing:

- **Propositional Knowing**: Knowing that something is true (semantic memory).
- **Procedural Knowing**: Knowing how to do something (skills, procedural memory).
- **Perspective Knowing**: Knowing what something is like from a certain point of view (episodic memory).
- **Participatory Knowing**: Knowing by being or embodying a role (affordance, self).

These different types of knowing play a crucial role in problem solving and insight.

7.5 Memory Organization and Compartmentalization

Weisberg and Alba's counterproposal to Gestalt insight involved memory organization. They suggested that memory is compartmentalized, meaning that when solving a problem, only certain compartments of memory are searched. This process requires insight, leading to another version of the interconnection problem—how different types of memory, problem solving, and categorization work together.

Casselman and Maier (1970) attempted to explore the limits of "thinking outside the box" by using a larger visual frame in their experiments. Interestingly, this intervention helped men solve the problem but not women, highlighting that problem solving is also influenced by historically constructed knowings.

7.6 Conclusion

Rationality and intelligence intersect with problem solving in complex ways. The GPS project, while providing valuable insights into heuristics and algorithms, has limitations when it comes to solving ill-defined problems. Gestalt psychology offers an alternative, emphasizing insight and the reformulation of problems. Different types of knowing—propositional, procedural, perspective, and participatory—add layers to how we understand problem solving. Memory, especially its compartmentalized nature, plays a central role in how we approach and solve problems.

8 Insight and Memory: The Search Inference Framework

8.1 Weisberg and Alba's Memory Search Theory

Weisberg and Alba argued that insight is simply a form of memory search. They claimed that when people experience moments of insight, they are merely retrieving information from memory, similar to how they would recall facts or details.

8.2 Metcalfe's Counterargument: The Feeling of Knowledge and Warmth

Metcalfe proposed a counterargument to Weisberg and Alba's theory. She introduced the concept of a cognitive probe known as the Feeling of Knowledge (FOK), which is a reliable predictor of accurate memory recall. FOK works well in predicting when memory retrieval will be successful, but Metcalfe found that it did not predict moments of insight.

Metcalfe replaced the Feeling of Knowledge with the Feeling of Warmth, borrowed from Newell and Simon's work. Warmth ratings reflect a person's

perceived closeness to solving a problem. Metcalfe and Wiebe (1987) tested this with both non-insight and insight problems. They found that:

- For non-insight problems, warmth ratings increased incrementally as the problem approached a solution.
- For insight problems, warmth ratings showed an S-shaped learning curve, reflecting a sudden leap in understanding.

This experiment demonstrated that memory search is effective for non-insight problems but fails to predict insights.

8.3 Kaplan and Simon's Search Inference Framework

Kaplan and Simon (1990) recognized that problem formulation is critical in solving problems. They introduced the concept of two types of search spaces:

- **Meta Search Space**: Involves searching for possible problem formulations.
- **Primary Search Space**: Involves searching for solutions to a formulated problem.

Kaplan and Simon claimed that insight does not involve any new cognitive processes but is simply the result of shifting between these two search spaces. They proposed that shifting from the left hemisphere (which processes problems step by step) to the right hemisphere (which processes problems holistically and rapidly) explains the leap from problem formulation to insight.

However, this approach also faces challenges, particularly with combinatorially explosive searches in the meta search space. Kaplan and Simon emphasized the importance of noticing invariants—features that remain unchanged across different problem formulations. By paying attention to failed problem formulations, individuals can refine their problem-solving approaches.

8.4 The Gestalt View and Insight

The Gestalt view of insight contrasts with the "business as usual" perspective of the Search Inference Framework. Gestalt psychologists argue that insight involves a sudden, ineffable shift in perspective that cannot be easily explained by gradual problem solving. Baker-Sennett and Ceci (1996) explored this ineffability, arguing that some types of knowing, particularly non-propositional knowing, cannot be put into words.

They described this leap from "I don't know" to "I know" as inductive or abductive leaping. Their research found that individuals who are better at making these cognitive leaps are more successful at solving insight problems. Pattern detection and completion are key abilities that contribute to effective cognitive leaping.

8.5 Dunker's Radiation Problem

Dunker's radiation problem is a classic example of an insight problem. The problem involves using X-rays to kill a tumor without harming surrounding tissue. Dunker's solution was to propose abstract strategies that could then be applied to concrete situations. This illustrates how solving an insight problem abstractly can lead to real-world solutions.

However, Dunker's problem also highlights the difficulty of transferring solutions between abstract and concrete problems, a failure often associated with transfer appropriate processing.

8.6 Gick and Holyoak's Fortress Problem

Gick and Holyoak (1980) explored a similar problem with their fortress analogy, attempting to transfer abstract problem-solving strategies to concrete ones. Despite the structural similarity between the problems, participants often failed to make the connection. This failure challenges the idea that insight is purely about recognizing structural similarities between problems.

8.7 Conclusion

The Search Inference Framework provides a structured approach to understanding insight, but it does not fully explain the sudden, ineffable nature of insight that Gestalt psychologists emphasize. The role of memory, problem formulation, and cognitive leaping are crucial components in solving both non-insight and insight problems. While Kaplan and Simon's framework offers valuable insights into the cognitive processes underlying problem solving, Gestalt theory reminds us that some aspects of human cognition, particularly insight, may be beyond simple explanation.

9 Attention and Insight: Beyond the Spotlight Metaphor

9.1 Polanyi's Focal and Subsidiary Awareness

Polanyi introduced the idea of focal and subsidiary awareness, which forms the basis of his theory of attention.

- **Focal awareness**: Being aware of an object directly, such as focusing on a physical object.
- **Subsidiary awareness**: Being aware of something through another medium, such as feeling something with your fingers.

Polanyi argued that attention involves both types of awareness, and insight arises when there is a dynamic coordination between them. This contrasts with

the spotlight metaphor, which oversimplifies attention as focusing on a single object.

9.2 Transparency and Opacity in Attention

Apter expanded on Polanyi's work by discussing the shift between transparency and opacity in attention.

- **Transparency**: When you are using something as a tool and are aware of what you are perceiving through it.
- **Opacity**: When you become aware of the tool itself.

The ability to shift between transparency and opacity is crucial for insight, allowing for cognitive leaping and breaking of frames. This process is not captured by the spotlight metaphor of attention, which assumes a fixed focus.

9.3 Attention as Dynamic Coordination

Attention, as described by Vervaeke and Ferraro (2016), is not a single spotlight but a dynamic process involving scaling up and down between focal and subsidiary awareness. Cognitive leaping, breaking frames, and de-automatization are all part of this process, which enhances insight and problem-solving.

9.4 Salience and Sizing Up

Matson introduced the idea of *salience landscaping*, where attention is used to size up and prioritize relevant information in a nested structure. This process, known as *figuration*, helps individuals foreground and background different aspects of a problem, which frames the problem's structure (SFO) and aids in problem solving.

9.5 Attention as Prioritization

Watzl proposed that attention functions as a hierarchy of prioritization, directly pointing to relevance realization. Attention is not just a spotlight but a process that handles multiple constraints simultaneously, allowing for insight to occur by balancing these constraints.

9.6 Challenges to the Single Difficulty Hypothesis

Kershaw and Ohlsson (2004) challenged the single difficulty hypothesis, which was based on the spotlight metaphor. They identified three factors influencing problem-solving:

- **Perceptual factors**: Involves attentional focus, figural processing, and pattern recognition.

- **Processing factors**: Involves framing and formulating the problem.
- **Knowledge factors**: Involves issues of transfer and applying relevant knowledge.

Training in all three factors resulted in a 40

9.7 Entropy and Self-Organizing Criticality (SOC)

Stephan and Dixon (2009) explored the relationship between entropy and insight. They found that insight is associated with a particular kind of entropy: as mutual prediction decreases (increased entropy), the system becomes more disorganized, leading to self-organizing criticality (SOC). After SOC, a new order emerges, facilitating insight.

This idea can be understood using the metaphor of an hourglass: as entropy breaks down the old frame, it creates space for a new frame to form, leading to insight.

9.8 Wiring Theory and Small World Networks

Wiring theory, based on network theory, examines how different types of networks are structured:

- **Regular networks**: Low efficiency but high resiliency.
- **Random networks**: High efficiency but low resiliency.
- **Small world networks**: More efficient than regular networks and more resilient than random networks.

Schilling (2005) proposed that insight involves restructuring semantic networks, shifting from regular to small world networks. This reorganization optimizes the balance between efficiency and resiliency, reflecting the process of cognitive leaping and insight.

9.9 Conclusion

Attention and insight go beyond the simple spotlight metaphor. Instead, attention involves a dynamic coordination between focal and subsidiary awareness, transparency and opacity, and the balancing of multiple constraints. Insight occurs through processes such as de-automatization, breaking frames, and reorganizing semantic networks. Understanding attention in this way highlights the complexity of cognitive processes involved in problem solving and insight.

10 Procedural Knowledge and Insight

10.1 Transfer Failure and Procedural Knowledge

Gick and Holyoak's experiments on problem transfer showed that insight problems often fail to transfer solutions, even when they are propositionally similar. This failure is due to the procedural differences between tasks. Lockhart, Lamon, and Fick (1988) distinguished between propositional and procedural knowledge, highlighting that procedural knowledge involves knowing how to perform tasks, while propositional knowledge focuses on knowing that something is true.

Their study on the "multiple marriage problem" showed that participants who approached the problem procedurally (trying to find an alternative meaning of marriage) performed significantly better than those who approached it propositionally (reading comprehension). This finding suggests that procedural knowledge is crucial for insight and problem-solving, especially when transferring solutions between tasks.

10.2 De-Automatization and Insight

One of the key aspects of insight is de-automatization, which involves converting automatic processes back into conscious, effortful ones. Automatization allows tasks to be performed without cognitive load, but when problems are novel or ill-defined, automatic processes fail. Insight requires bringing these tasks back into working memory and altering perspectival knowledge to solve the problem.

Weisberg and Alba's Search Inference Framework initially proposed that insight problems were simply memory searches. However, they later admitted that insight involves more than just retrieving information. Instead, insight requires breaking automatic processes and reevaluating the problem. This de-automatization allows for a shift in perspective, which is critical for solving insight problems.

10.3 The Role of Attention and Constraint Relaxation

Attention plays a key role in de-automatization. Kaplan and Simon (1990) proposed that insight occurs when individuals notice invariants—features that remain unchanged across different problem formulations. However, attention is not just about directing focus; it involves adjusting the way problems are framed. Knoblich's research on the matchstick problem showed that people who are good at insight are able to break tight chunks of information and relax constraints that limit problem solving.

Constraint relaxation is a procedural understanding of how to approach problems. It involves loosening the rules or assumptions that govern the problem, allowing for new perspectives and solutions. This procedural version of constraints highlights the importance of cognitive flexibility in insight.

10.4 Cognitive Flexibility and Insight Problem Solving

Research by Gilhooly and Murphy (2005) identified cognitive flexibility as the most predictive factor of insight problem solving. Individuals who are good at breaking and re-forming cognitive frames are better at solving insight problems. Tasks like figural fluency (connecting dots to form patterns) and alternative use tests (finding different uses for common objects) are strong predictors of insight performance.

The ability to make cognitive leaps—breaking existing frames and creating new ones—is essential for solving insight problems. Cognitive flexibility allows individuals to move beyond rigid thinking and discover creative solutions.

10.5 Neurological Basis of Insight: Left and Right Hemisphere Shift

The process of insight has also been linked to a neurological shift from the left hemisphere (responsible for step-by-step problem solving) to the right hemisphere (responsible for holistic, Gestalt processing). Jung-Beeman and Kounios (2003, 2004) found that insight problems often involve a shift from left to right hemisphere activity. The left hemisphere is adept at solving well-defined problems through incremental steps, while the right hemisphere excels at solving ill-defined problems through leaps in understanding.

This shift between hemispheres reflects the dual nature of problem solving—moving between primary search space (incremental problem solving) and meta search space (reformulating the problem). Understanding this neurological basis helps explain why insight feels sudden and often ineffable.

10.6 Conclusion

Insight is a complex cognitive process that involves more than just memory search. It requires de-automatization, attention, and cognitive flexibility to break existing frames and create new ones. Procedural knowledge, particularly in the form of constraint relaxation, plays a critical role in solving insight problems. The neurological shift between left and right hemispheres further explains the leap from incremental problem solving to insight, highlighting the dual nature of cognitive processing.

11 Introduction to the History of AI

11.1 The Aristotelian Worldview

The Aristotelian worldview is built on two mutually supportive philosophies:

- **Epistemology**: the study of how we know the world.
- **Ontology**: the study of the structure of the world (Structural Functional Organization, or SFO, of reality).

In this worldview, Aristotle's *Conformity Theory* of knowing states that to know something is not just to represent it in language through propositions but to conform to it. For example, a person who knows how to make a chair understands the form of the chair (SFO) and can create it. Knowing is about sharing the same form as the object of knowledge.

Aristotle also proposed a *Geocentric Theory*, where objects, including the Earth, move because of their inertial drive or internal purpose. Systems of objects —including the cosmos which contains planets— are also inertial and maintain harmonic movements of its inhabitants. This can take the form of natural cycles such as air, fire, water, and earth returning to their "natural" places. Finally, according to Geocentric Theory, the Earth is at the center of the universe system.

11.2 The Scientific Revolution

The scientific revolution represented a rejection of the Aristotelian worldview. Copernicus proposed a heliocentric theory, where the sun is at the center of the universe. This shift undermined the Conformity Theory because it challenged the belief that human perception alone leads to knowledge of reality.

Galileo further revolutionized thinking by asserting that mathematics is the language of the universe. He introduced the concept of inertial motion, suggesting that not all motion has purpose, as Aristotle believed. This transition led to the idea that primary qualities (objective, measurable properties) are real, while secondary qualities (subjective experiences) are illusory.

11.3 Descartes, Hobbes, and the Rise of Computation

Descartes was deeply troubled by the scientific revolution and the possible implication that we could not trust our senses. Among his solutions, he introduced the Cartesian coordinate system and argued that the mind should operate through mathematical propositions. However, he also rejected materialism, arguing that cognition is not purely mechanical. According to Descartes, reasoning involves purpose, value, and meaning—qualities that cannot be reduced to mere mechanical processes.

Thomas Hobbes, on the other hand, embraced the idea that cognition is computation. He proposed that machines could perform computations automatically and that this mechanistic process could explain cognition. However,

Descartes countered that while machines can manipulate logic, they cannot reason because they lack purpose and meaning.

11.4 Cartesian Dualism and the Problems of AI

Descartes' arguments against Hobbes led to the formulation of Cartesian dualism, which holds that the mind and body are fundamentally different kinds of things. Descartes also raised several key problems that any theory of artificial intelligence must address:

- **The Problem of Purpose**: How can machines, which are inert, perform actions with purpose?
- **The Problem of Normativity**: How can machines make judgments about what is true or valuable?
- **The Problem of Original Meaning**: How can machines, which manipulate symbols, attach meaning to those symbols?
- **The Problem of Consciousness**: How can machines, which operate through mechanical processes, achieve consciousness?

11.5 Artificial Intelligence: Weak and Strong AI

The philosophical-scientific project of AI aims to resolve these problems and show that Hobbes was correct about cognition as computation. There are two main types of AI:

- **Weak AI**: Machines that can perform tasks previously done by humans, such as playing chess or diagnosing diseases.
- **Strong AI (Artificial General Intelligence)**: Machines that can create cognition and help us understand human intelligence. The ultimate goal of strong AI is to place cognition within the scientific worldview, answering the questions raised by Descartes.

The development of AI must address the problems of purpose, normativity, original meaning, and consciousness. If successful, AI could revolutionize our understanding of human cognition and demonstrate that cognition is, in fact, a computational process.

11.6 Conclusion

The history of AI is deeply rooted in philosophical debates about the nature of knowledge, purpose, and reasoning. From Aristotle's Conformity Theory to Descartes' dualism and Hobbes' mechanistic view of cognition, the quest to understand intelligence has shaped the development of AI. As AI continues to evolve, it must confront the fundamental challenges posed by the problems of purpose, meaning, and consciousness.

12 The Formal System and Computation

12.1 What Defines a Computer?

Computation involves the manipulation of logical and mathematical abstract propositions, a concept closely related to the functioning of modern computers. Hobbes raised an important question: how can one be a computer? While this question might not be addressed from an engineering perspective, we are interested in the abstract version. A computer, in this context, is defined as an interpreted automatic formal system.

12.2 Formal Systems and Token Manipulation

A formal system, much like a formal game such as chess, is governed by strict rules for manipulating tokens. These tokens, whether physical pieces on a chessboard or symbols in a computational system, follow well-defined manipulation protocols:

- Tokens are identified by their physical properties.
- Manipulation of tokens follows rules, and the meaning of the tokens is irrelevant during the process.
- The process of moving, altering, and replacing tokens creates new configurations in a self-contained system.

The formal system emphasizes that the meaning of the tokens is not a relationship within the system but with the external world.

12.3 Medium Independence and Multiple Realizability

A formal system operates under the principle of medium independence, meaning that the specific physical properties of a system are irrelevant as long as the rules of the formal system are followed. This gives rise to the concept of multiple realizability, where a formal system can be implemented in various mediums. AI, for example, presupposes that cognitive systems can be realized in different physical forms, much like how a chess game can be played with different sets of physical pieces.

12.4 Digitality and Finite Playability

Formal systems, to be effective, must be digital and finitely playable:

- **Digitality** ensures the self-contained nature of the formal system, using reliable read-write cycles to maintain its integrity.
- **Finite playability** means that the formal system must operate within finite constraints, allowing rules to be followed without leading to infinite regress or combinatorial explosion.

A formal system can either follow rules or simply describe them. For example, the solar system can be described by calculus but is not "following" the rules of calculus. This raises philosophical questions about whether formal systems can truly "follow" rules or simply describe them.

12.5 Dennett's Solution and the Decomposition of Intelligence

Daniel Dennett proposed that intelligence can be decomposed into a series of mindless, primitive operations. These operations, following simple rules, can be linked together to form complex algorithms. This process of decomposition explains how intelligent systems, including computers, can perform complex tasks in a purely mechanical manner.

The creation of automatic formal systems, where referees ensure rule-following, is based on the idea that meaning is not intrinsic to the system but must be integrated through other means. Until we can show how formal systems can handle meaning, Descartes' challenge to explain cognition remains unanswered.

12.6 Conclusion

The exploration of formal systems reveals that computation, at its core, involves manipulating tokens under a structured set of rules. These systems are abstract, medium-independent, and can be realized in various physical forms—highlighting the principle of multiple realizability. However, while formal systems excel at processing information mechanically, they fall short in addressing questions of meaning and consciousness. Dennett's work on the decomposition of intelligence offers a path forward, but the challenge of embedding meaning within purely formal operations remains unresolved. This raises fundamental philosophical questions about the nature of cognition, computation, and whether machines can truly replicate human understanding.

13 Algorithmic Systems and Cognitive Science

13.1 The Implementation-Use Distinction

In computational systems, every program is algorithmically built and follows the principles of the Naturalistic Imperatives. However, this does not mean the program is always used algorithmically. For example, chess programs are built on algorithms but often run heuristically during gameplay. These programs, being algorithmic in nature, are formal systems because they manipulate abstract propositions.

13.2 Resolving the Paradox of Mechanical Reasoning

One key challenge in cognitive science is resolving the paradox of mechanical reasoning. Classical interpretations suggest that interpreting information in-

volves making sense of it. However, a machine that translates languages (e.g., French to English) may map one code to another without truly understanding the meaning behind the translation. This distinction between syntactic relations (token-to-token) and semantic relations (token-to-world) highlights the difference between merely following rules and genuinely understanding the content.

When a token has a semantic relationship, it becomes a symbol, which is essential for cognition. The paradox lies in the fact that while syntactic processes can create coherent patterns, they often fail to capture the true meaning behind those patterns.

13.3 First, Second, and Third Generations of Cognitive Science

The development of cognitive science can be divided into three generations:

1. **First-Generation Cognitive Science**: Rooted in the idea that cognition is computation, exemplified by GOFAI (Good Old-Fashioned AI), which focused on formal systems for manipulating abstract propositions. Early cognitive scientists, like Hobbes, proposed that cognition could be fully understood as computation.

2. **Second-Generation Cognitive Science**: This generation shifted focus to neural networks and connectionist models, where cognition is seen as distributed and emergent from simple interactions between units.

3. **Third-Generation Cognitive Science**: Known as 4E cognition, this perspective views cognition as embodied, embedded, enacted, and extended, emphasizing the importance of environmental and bodily interactions in shaping cognitive processes.

13.4 The Limitations of GOFAI and SHRDLU

GOFAI's limitations became evident when its models, such as SHRDLU, failed to scale up to handle real-world complexity. SHRDLU performed well in simplified environments but struggled with issues such as reconstructive memory, development, and ill-defined problems. It could position objects according to queries, but its success in micro-worlds did not translate to complex, dynamic environments. This highlighted the difficulty of creating AI systems that can generalize beyond controlled settings.

13.5 The Frame Problem and Computational Limitations

Herbert Dreyfus famously outlined the limitations of computational systems in his work, "What Computers Can't Do." One key issue is the frame problem, which refers to the challenge of coming up with a representation of the world that deals with change. Dan Dennett expanded on this, pointing out two main problems:

1. **Logical Problem of Change**: How do we represent and reason about a world that is constantly changing?

2. **Theoretical Problem**: How can we design a cognitive agent, such as a robot, that adapts to an ever-changing environment without predefined rules?

These problems continue to challenge the notion that cognition can be fully understood as computation, requiring more nuanced approaches to understanding intelligence.

13.6 Conclusion

There are many challenges in defining and implementing computational systems as models of cognition. While GOFAI provided an early foundation for cognitive science, it struggled to address real-world complexity, leading to the development of new models, such as connectionism and 4E cognition. These newer approaches emphasize the importance of context, embodiment, and adaptability in understanding how cognitive systems operate. Despite advancements, the frame problem and issues of scalability continue to pose significant obstacles to developing AI that truly mirrors human cognition.

14 Cognition Beyond Language: Challenges to GOFAI

14.1 Fodor and Dennett's Critique of Common Sense

Jerry Fodor and Daniel Dennett highlighted a critical issue: the blinding effect of the obviousness of common sense. This problem stems from the assumption that cognition is computation—a belief rooted in the manipulation of abstract propositions. Underlying this assumption is the deeper belief that cognition is like language, where subject-predicate logic forms the foundation of thought.

Fodor and Dennett argued that this language-like view of cognition underpins early GOFAI (Good Old-Fashioned AI). The common-sense assumption that cognition operates like language blinded early researchers to alternative possibilities. While GOFAI contributed to refining psychological theories and helped break away from dualism, its reliance on the assumption that cognition is linguistic would eventually be refuted by second and third-generation cognitive science.

14.2 Connectionism as an Alternative Model

In contrast to GOFAI, connectionism models cognition based on the brain rather than language. Instead of relying on logical and grammatical principles, connectionist models abandon the idea that cognition uses discrete symbols or operates inferentially. The brain is seen as a network of neurons that signal each other,

with signal strength and valence (positive or negative) influencing the weight of connections between neurons.

This shift represents a departure from the linguistic model of cognition, emphasizing non-linguistic brain functions. The connectionist approach offers explanations for phenomena that GOFAI struggled with, such as:

- Explaining procedural knowledge (distinct from propositional knowledge).
- Handling the multiple simultaneous constraint problem.

Connectionist models provide a framework for understanding these issues by moving away from the presupposition that cognition is linguistic.

14.3 Neural Networks and Learning

Neural networks operate through a layered structure: input, hidden layers, and output. The hidden layers play a crucial role in the learning process. Although early work by Minsky and Pappert suggested that adding hidden layers wouldn't significantly change outcomes, it was later shown that these layers are vital for learning.

Learning in neural networks is often explained through the backpropagation of errors. This process adjusts the weights of connections between neurons based on feedback, allowing the network to learn from mistakes. However, challenges such as correlation error and sampling bias can affect the learning process, highlighting the need for data augmentation techniques.

Neural networks have been particularly successful in explaining procedural knowledge, which is different from propositional knowledge. This success has been attributed to their ability to handle multiple constraints simultaneously, providing a more robust model of cognition than GOFAI.

14.4 Implicit Learning and Neural Networks

One key strength of neural networks is their ability to explain implicit learning, which occurs without explicit awareness. Experiments using artificial grammar, such as Markov grammars, have shown that people can learn patterns without knowing the underlying rules. Neural networks model this form of learning effectively, performing well even when humans struggle to articulate the knowledge they've acquired.

Developmental neural networks also show potential in understanding human-like learning processes. Unlike traditional computational models, which often fail to replicate human developmental learning, neural networks exhibit flexibility and adaptability that mimic human cognition more closely.

14.5 Conclusion

The evolution of cognitive science from GOFAI to connectionism marks a shift from viewing cognition as language-like to seeing it as brain-like. Neural net-

works offer a powerful alternative by abandoning linguistic principles and focusing on non-linguistic brain functions. Their ability to explain procedural knowledge, implicit learning, and multiple constraint problems has solidified their role in modern cognitive science. However, challenges remain, particularly in refining neural network models to handle real-world complexity and developmental learning processes.

15 Challenges in Cognitive Science: Connectionism vs. GOFAI

15.1 The Weakness of Connectionism

Connectionism emerged as an alternative to GOFAI (Good Old-Fashioned AI), but it faced criticism for its disconnection from language-like structures, which are necessary for specificity and inferential coherence in thought. Fodor pointed out that while connectionism focuses on causal connections between nodes in a neural network, it fails to capture the structural relationships that are critical for explaining cognition scientifically. In particular, connectionism lacks a clear theory of truth preservation, which is fundamental to explaining how cognition is systematic and truth-preserving.

15.2 The Limitations of Connectionism and GOFAI's Strengths

Fodor emphasized the need for a theory of cognition that aligns with the scientific worldview, where truth preservation, inferential coherence, and productivity are key. While neural networks may co-activate nodes representing concepts like "Loves," "John," and "Mary," the addition of more nodes can lead to ambiguity in representing relationships. GOFAI, on the other hand, maintains causal and structural relations that preserve meaning through a subject-predicate relationship.

Connectionism struggles to account for the specificity of thought and the systematicity of cognition, as it lacks the compositionality needed to explain how cognitive systems generate an unlimited number of inferences. This is a key advantage of GOFAI, which explains systematicity through structural patterns and logical operations.

15.3 Truth Preservation and Structural Patterns

Fodor argued that cognitive systems must be truth-preserving, and this can only be achieved through structural patterns that ensure logical consistency. While GOFAI provides an explanation for truth preservation by relying on formal systems, connectionism's associative models do not guarantee that associations between nodes preserve truth. The challenge for connectionism is developing an alternative theory of truth preservation that does not rely solely on structural patterns but still accounts for the systematic nature of thought.

15.4 The Systematicity of Cognition

Human cognition exhibits tremendous systematicity, meaning that our thoughts and inferences are coherent and productive. Fodor linked this systematicity to language-like structures, which provide inferential coherence and productivity. GOFAI is well-suited to explain systematicity because it relies on formal systems with clear subject-predicate relationships. In contrast, connectionism lacks the necessary structural elements to explain how thoughts are systematically related to one another.

15.5 The Hybrid Approach and Dual Processing

As a response to the limitations of both GOFAI and connectionism, cognitive science has moved toward a hybrid approach, incorporating elements of both. This dual-processing theory of cognition suggests that there are two kinds of cognitive processes: one that is fast and intuitive and another that is slower and more deliberate. This hybrid model acknowledges that no single system can explain the full complexity of human cognition.

15.6 Conclusion

The debate between GOFAI and connectionism highlights fundamental challenges in cognitive science. While connectionism offers a model based on causal connections between neural nodes, it struggles to account for truth preservation, specificity of thought, and systematicity. GOFAI, with its reliance on formal systems, provides explanations for these cognitive features but is limited in its applicability to real-world problems. The hybrid approach, which combines elements of both, offers a promising path forward in understanding the complexity of human cognition.

16 Criteria for Cognition and the Mind-Body Problem

16.1 The Need for a Cognitive Criterion

In cognitive science, it is essential to establish a clear criterion for what constitutes cognition. Without a natural instance of cognition, we cannot reach definitive scientific conclusions about it. Steven Harnad proposed various tests to evaluate cognitive capabilities, creating a scale of progressively stronger tests:

- **Toy Modeling (T1)**: A weak model that avoids complex problems but cannot scale up to real-world applications.
- **Turing Test (T2)**: A widely recognized but insufficient test that faces issues such as the symbol grounding problem and the Chinese Room argument.

- **Total Turing Test (T3)**: A machine capable of replicating human cognition across multiple domains over a lifetime, exhibiting development. This encompasses artificial general intelligence (AGI) with unsupervised learning.
- **Neuromimetic Models (T4)**: Models that attempt to replicate human neural processes but are deemed overly strict, as machines do not need to mimic every physical aspect of human biology.
- **Neuroquantum Models (T5)**: A hypothesis suggesting that cognition operates at the quantum level, which could address the combinatorial explosion problem.

Harnad advocated for the T3 test as the ideal compromise—while it is difficult to achieve, it does not fall into the impractical extremes of the other tests.

16.2 The Role of Explanatory Power and Generalization

Chiappe and Kukla emphasized the importance of explanatory power in identifying a cognitive criterion. They advocated using J.S. Mill's approach to ensure that the criterion supports scientific inquiry by selecting stable and intrinsic categories. These categories must be generalizable, allowing past experience to inform future understanding. For example, learning about horses helps inform future interactions with other horses, unlike learning about random categories, such as "white things," which do not provide systematic explanatory power.

16.3 Relevance Realization as a Criterion for Cognition

The problem of establishing a criterion for cognition is closely linked to the concept of relevance realization (RR). RR is essential for cognition, as it allows cognitive systems to focus on what is meaningful in a given context. However, RR itself faces challenges in becoming scientifically measurable. Despite these difficulties, Harnad and others argue that RR might serve as the necessary criterion for defining cognition, as it ties cognitive processes to explanatory power and systematic import.

16.4 The Mind-Body Problem and Cartesian Dualism

The mind-body problem is central to understanding cognition. Historically, materialism, as proposed by Hobbes, argued that everything is matter, including the mind. However, this view faced challenges with explaining immaterial properties such as normativity and consciousness.

Descartes introduced dualism, claiming that the mind and body are two distinct substances. While the body is physical, the mind is immaterial and possesses properties such as purpose, normativity, and qualia. However, Cartesian dualism has been criticized for its inability to explain how the mind and body interact causally.

16.5 Challenges to Cartesian Dualism

One major problem with Cartesian dualism is the individuation problem—how can we separate the mind from the body when the two seem to share no common properties? Descartes suggested that the mind resides in the pineal gland, a proposal later refuted by Malebranche, who argued that God mediates the connection between the mind and body. Leibniz introduced the idea of pre-established harmony, where the mind and body correlate but do not causally interact.

These historical attempts to solve the mind-body problem ultimately failed, as they could not reconcile the immaterial mind with the physical body. Modern cognitive science has largely moved away from Cartesian dualism in favor of materialist or emergentist models of cognition.

16.6 Emergentism and Property Dualism

Property dualism, which suggests that mental properties emerge from physical processes, provides a modern alternative to Cartesian dualism. Emergentism posits that the mind emerges from complex interactions within the brain, similar to how water emerges from the combination of hydrogen and oxygen. There are two forms of emergence:

- **Weak Emergence**: Mental properties arise from physical processes but are still reducible to them.
- **Strong Emergence**: Mental properties are irreducible and possess unique causal powers, existing as fundamental properties of the universe.

Epiphenomenalism, a related view, proposes that mental properties exist but lack causal power, functioning more as byproducts of physical processes.

16.7 Conclusion

In the quest to define cognition and understand the mind, cognitive scientists must grapple with both the philosophical and empirical challenges of defining a clear criterion. Whether through relevance realization or emergentist models, the challenge remains to bridge the gap between the physical brain and the immaterial aspects of cognition, such as normativity and consciousness. While Cartesian dualism has been largely abandoned, the debate over the nature of the mind continues to shape modern cognitive science.

17 The Fall of Dualism and the Rise of Materialism

17.1 Rejection of Dualism in Cognitive Science

Dualism, once considered a plausible explanation for the mind-body problem, has been largely rejected by modern cognitive science. Cognitive scientists now favor monistic approaches, particularly materialism, which posits that everything, including mental states, has a physical basis. Dualism's core problem lies in its inability to explain how immaterial entities (like the mind) interact with physical entities (like the body).

17.2 Monism and Idealism

Idealism, a form of monism, claims that everything is mental in nature. While philosophically intriguing, idealism faces significant challenges:

- **Deep past**: How did the world exist before there were minds to perceive it?
- **External world**: If the external world is mind-dependent, how do we explain phenomena that exist independently of human observers?

Idealism also struggles with explaining the reality of non-human minds and events that occurred long before human consciousness emerged. While idealism presents a radical alternative to materialism, its lack of scientific grounding makes it an impractical framework for understanding cognition.

17.3 Logical Behaviorism and the Concept of Mind

In response to the failings of dualism and idealism, logical behaviorism emerged as a more scientific approach to understanding the mind. Inspired by philosophers like Wittgenstein and Gilbert Ryle, logical behaviorism posits that mental states are nothing more than dispositions to behave in certain ways. Ryle famously argued that the mind-body problem was a category mistake—that mental states are not things but rather dispositions tied to behavior.

Ryle's argument aimed to dissolve the mind-body problem by showing that the question itself was posed incorrectly. However, logical behaviorism faces several problems:

- **Super actor problem**: It is possible for someone to perfectly mimic behaviors without possessing the corresponding mental states.
- **Super Spartan problem**: Individuals can possess mental states without exhibiting any observable behavior.
- **Combinatorial explosion**: Logical behaviorism requires an exhaustive list of behaviors to explain all possible mental states, leading to combinatorial explosion.

These problems indicate that logical behaviorism, while useful for understanding certain aspects of mental states, cannot fully replace mental terms with behavioral descriptions.

17.4 Identity Theory and the Reduction of Mental States

Identity theory suggests that all mental states can be reduced to brain states, arguing that psychological terms can be replaced by neuroscientific ones. Proponents of identity theory argue that as neuroscience advances, we will eventually be able to fully explain mental states in terms of brain activity.

However, identity theory also faces significant criticism:

- **Multiple realizability**: Mental states can be realized in different physical systems, suggesting that they cannot be reduced solely to brain states.
- **Subjectivity of consciousness**: Identity theory struggles to explain qualia, or the subjective experience of consciousness. This leads to an explanatory gap between physical processes and mental experiences.

Despite its limitations, identity theory remains influential in discussions of the mind-body problem, although it has increasingly been replaced by functionalist approaches.

17.5 Computational Functionalism and the Problem of Consciousness

Computational functionalism offers an alternative to identity theory by suggesting that mental states are functional states, defined by the roles they play in cognitive systems rather than by their physical substrate. Functionalism claims that cognition is a form of computation, where mental processes are the result of information processing.

However, computational functionalism faces significant challenges in explaining consciousness:

- **Inverted spectrum argument**: Individuals with the same functional states could experience different qualia (e.g., seeing red as green and vice versa).
- **Absent qualia**: Functional systems could operate without any subjective experience (e.g., the Chinese Room argument).

These challenges highlight the difficulty of explaining consciousness within a purely computational framework. While functionalism offers a promising approach to understanding many cognitive processes, it struggles to account for the subjective nature of consciousness.

17.6 Conclusion

The rejection of dualism and the rise of materialism in cognitive science have shifted the focus to understanding the mind as a physical process. However, both identity theory and computational functionalism face challenges in explaining the subjective aspects of consciousness. As cognitive science continues to evolve, new approaches will be needed to bridge the gap between physical processes and mental experiences, while addressing the limitations of current theories.

18 Relevance Realization and the Limitations of Computational Functionalism

18.1 The Problem of Relevance Realization (RR)

Relevance realization (RR) seems to be at the core of all cognitive processes, yet it remains a challenge that both GOFAI and neural networks have failed to solve. RR is dynamically self-organizing and plays a crucial role in insight and decision-making processes. However, one major issue is the detection of relevance—if relevance is seen as an objective property, it implies that categories are stable and homogeneous. This is not the case; relevance is not intrinsic but is instead a relationship between the objective and subjective aspects of the world.

18.2 Representation and the Frame Problem

Searle (1992) argued that all representations are aspectual; they always grasp only an aspect of things. This perspective suggests that RR is deeply tied to consciousness, as representations rely on a point of view. Representational theories face significant challenges when explaining RR because they rely on presupposing RR to explain it, leading to circular reasoning (the homuncular problem). This indicates that attempts to explain RR purely in computational terms will fail, as seen with GOFAI and neural networks.

18.3 Challenges of Rule Following and Inference

Jerry Fodor highlighted the complexity of inference, noting that while meaning stays the same, what something means to an individual changes depending on relevance. Relevance cannot be captured in semantic syntax. Furthermore, rules themselves, such as "be kind," rely on conditions of application that cannot be fully captured by another set of rules. Rule-following presupposes RR, which means that any attempt to define RR purely in terms of rules is doomed to fail.

18.4 The Limits of Computational Functionalism

Computational functionalism, while useful in explaining many cognitive processes, struggles to account for the subjective nature of relevance and conscious-

ness. Philosophical zombies, such as blindsight (where individuals accurately perceive their environment without visual qualia), demonstrate that intelligent behaviors can occur unconsciously, raising questions about the role of qualia in cognition.

18.5 The Evolutionary Analogy for Relevance Realization

An analogy can be drawn between RR and evolutionary processes. Just as pre-Darwinian naturalists could not define "fittedness" based on any essential features, relevance cannot be defined as an intrinsic property. Darwin's insight was that fittedness is an evolving relationship between organisms and their environments—similarly, RR is an evolving relationship between the cognitive system and the world. This analogy highlights that RR is a dynamically self-organizing process, akin to how biological systems evolve through constraints and selective pressures.

18.6 Virtual Engines and Dynamical Systems

Relevance realization can be understood as part of a virtual engine, where selective constraints (like scarcity) act as a virtual governor, and enabling constraints (like variation) act as a virtual generator. These elements combine to create a feedback cycle, akin to a dynamical system. Deep learning, for example, demonstrates this kind of system, where constraints help shape learning and evolution in AI systems.

18.7 Autopoiesis and Information Processing

For cognitive systems to process information meaningfully, they must be autopoietic—they need to care about the information they process. This concept leads to the question of how artificial systems can achieve autopoiesis. Autopoietic systems turn information into meaningful information, a relationship between probability and events. Kolchinsky and Wolpert proposed that information is syntactic data that an organism uses to maintain itself, with causal relevance emerging from this self-maintenance.

18.8 Conclusion

Relevance realization is a fundamental yet elusive aspect of cognition that cannot be fully captured by computational models or rules. The analogy to evolutionary processes provides a useful framework for understanding RR as a dynamically self-organizing system. However, for artificial systems to truly engage in meaningful information processing, they must become autopoietic, capable of caring about the information they process. This presents a significant challenge for developing AI systems that can replicate the nuanced, self-organizing nature of human cognition.