

Artificial Intelligence & Machine Learning in Cognitive World:

Fundamentals of Artificial Intelligence: Various areas of AI:

a) Knowledge: Text Analytics, Topic Modelling, Natural Language Processing (NLP), Natural Language Generation (NLG), Natural Language Understanding (NLU), Named-entity recognition (NER), Role of Grammar in Language Processing, Connectionist Models, Physical symbol systems and language of thought-Applying the Symbolic Paradigm

b) Perception: Cognitive Bias Problem, Is AI Biased? Image Analytics, Video Analytics & Audio Analytics

c) Memory: Cognitive Engagement: BOTs, Virtual & Digital Assistants, Augmented Reality, Virtual Reality, Mixed Reality

d) Learning: Intelligent Automation, Reinforcement Learning, Intelligent Agents

Spectrum of AI

a) Reactive Machine: Low memory, works on Known rules, such as Object Detection/Games/Recommendations specific to known Rules

b) Limited Memory: Memory used to learn and improve continuously such as Most ML Models, Automated Vehicles

c) Theory of Mind: Machine Understands and responds such as BoTs/Virtual/Digital Assistants

d) Self-Aware: Human-like intelligence such as Super Robots in Space etc.

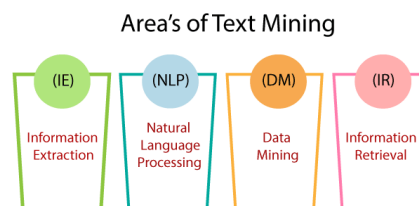
- a) **Knowledge:** Text Analytics, Topic Modelling, Natural Language Processing (NLP), Natural Language Generation (NLG), Natural Language Understanding (NLU), Named-entity recognition (NER), Role of Grammar in Language Processing, Connectionist Models, Physical symbol systems and language of thought-Appling the Symbolic Paradigm

BACKGROUND

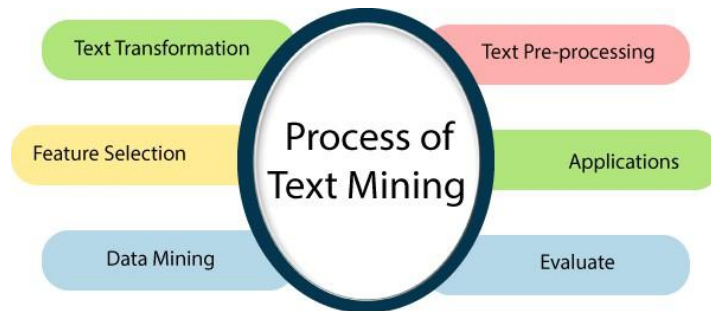
Text Data Mining

- ✚ *Text data mining can be termed as the process of extracting essential data from standard language text. All the data that we generate via text messages, documents, emails, files are written in common language text. Text mining is primarily used to draw useful insights or patterns from such data.*
- ✚ The text mining market has experienced exponential growth and adoption over the last few years and also expected to gain significant growth and adoption in the coming future.
- ✚ With increasing completion in business and changing customer perspectives, organizations are making huge investments to find a solution that is capable of analyzing customer and competitor data to improve competitiveness.
- ✚ *The primary source of data is e-commerce websites, social media platforms, published articles, survey, and many more. The larger part of the generated data is unstructured, which makes it challenging and expensive for the organizations to analyze with the help of the people. This challenge integrates with the exponential growth in data generation has led to the growth of analytical tools. It is not only able to handle large volumes of text data but also helps in decision-making purposes. Text mining software empowers a user to draw useful information from a huge set of data available sources.*

Areas of text mining in data mining:



Text Mining Process: The text mining process incorporates the following steps to extract the data from the document.



- **Text transformation**

A text transformation is a technique that is used to control the capitalization of the text.

Here the two major way of document representation is given.

1. *Bag of words*
2. *Vector Space*

- **Text Pre-processing** Pre-processing is a significant task and a critical step in Text Mining, Natural Language Processing (NLP), and information retrieval(IR). *In the field of text mining, data pre-processing is used for extracting useful information and knowledge from unstructured text data. Information Retrieval (IR) is a matter of choosing which documents in a collection should be retrieved to fulfill the user's need.*
- **Feature selection:** Feature selection is a significant part of data mining. *Feature selection can be defined as the process of reducing the input of processing or finding the essential information sources.* The feature selection is also called variable selection.
- **Data Mining:** Now, in this step, the text mining procedure merges with the conventional process. Classic Data Mining procedures are used in the structural database.
- **Evaluate:** Afterward, it evaluates the results. Once the result is evaluated, the result abandon.

Text Mining Approaches in Data Mining:

These are the following text mining approaches that are used in data mining.

1. Keyword-based Association Analysis:

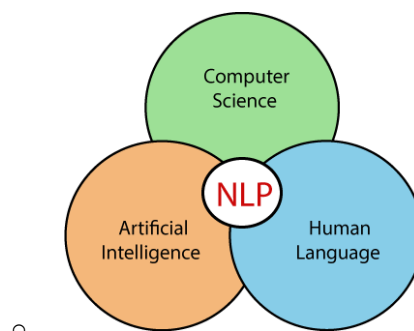
- ✚ *It collects sets of keywords or terms that often happen together and afterward discover the association relationship among them. First, it preprocesses the text data by parsing, stemming, removing stop words, etc. Once it pre-processed the data, then it induces association mining algorithms. Here, human effort is not required, so the number of unwanted results and the execution time is reduced.*

2. Document Classification Analysis:

✚ **Automatic document classification:** *This analysis is used for the automatic classification of the huge number of online text documents like web pages, emails, etc.* Text document classification varies with the classification of relational data as document databases are not organized according to attribute values pairs.

Numericizing text:

- **Stemming algorithms** A significant pre-processing step before ordering of input documents starts with the stemming of words. *The terms "stemming" can be defined as a reduction of words to their roots.* For example, different grammatical forms of words and ordered are the same. *The primary purpose of stemming is to ensure a similar word by text mining program.*
- **Support for different languages:** There are some highly language-dependent operations such as stemming, synonyms, the letters that are allowed in words. Therefore, support for various languages is important.
- **Exclude certain character:** Excluding numbers, specific characters, or series of characters, or words that are shorter or longer than a specific number of letters can be done before the ordering of the input documents.
- **Include lists, exclude lists (stop-words):** *A particular list of words to be listed can be characterized, and it is useful when we want to search for a specific word.* It also classifies the input documents based on the frequencies with which those words occur. Additionally, "stop words," which means terms that are to be rejected from the ordering can be characterized. Normally, a default list of English stop words incorporates "the," "a," "since," etc. These words are used in the respective language very often but communicate very little data in the document.
- NLP stands for **Natural Language Processing**, which is a part of **Computer Science**, **Human language**, and **Artificial Intelligence**. It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages. It helps developers to organize knowledge for performing tasks such as **translation**, **automatic summarization**, **Named Entity Recognition (NER)**, **speech recognition**, **relationship extraction**, and **topic segmentation**.



COMPONENTS OF NLP

1. Natural Language Understanding (NLU)

- ✚ Natural Language Understanding (NLU) helps the machine to understand and analyse human language by extracting the metadata from content such as concepts, entities, keywords, emotion, relations, and semantic roles.
- ✚ NLU mainly used in Business applications to understand the customer's problem in both spoken and written language.

NLU involves the following tasks -

- It is used to map the given input into useful representation.
- It is used to analyze different aspects of the language.

2. Natural Language Generation (NLG)

- ✚ Natural Language Generation (NLG) acts as a translator that converts the computerized data into natural language representation. It mainly involves Text planning, Sentence planning, and Text Realization.

Note: The NLU is difficult than NLG.

Difference between NLU and NLG

NLU	NLG
NLU is the process of reading and interpreting language.	NLG is the process of writing or generating language.
It produces non-linguistic outputs from natural language inputs.	It produces constructing natural language outputs from non-linguistic inputs.

How to build an NLP pipeline?

There are the following steps to build an NLP pipeline -

Step1: Sentence Segmentation

Sentence Segment is the first step for building⁵ the NLP pipeline. It breaks the paragraph into separate sentences.

Example: Consider the following paragraph -

Independence Day is one of the important festivals for every Indian citizen. It is celebrated on the 15th of August each year ever since India got independence from the British rule. The day celebrates independence in the true sense.

Sentence Segment produces the following result:

1. "Independence Day is one of the important festivals for every Indian citizen."
2. "It is celebrated on the 15th of August each year ever since India got independence from the British rule."
3. "This day celebrates independence in the true sense."

Step2: Word Tokenization

Word Tokenizer is used to break the sentence into separate words or tokens.

Example:

JavaTpoint offers Corporate Training, Summer Training, Online Training, and Winter Training.

Word Tokenizer generates the following result:

"JavaTpoint", "offers", "Corporate", "Training", "Summer", "Training", "Online", "Training", "and", "Winter", "Training", "."

Step3: Stemming

Stemming is used to normalize words into its base form or root form. For example, celebrates, celebrated and celebrating, all these words are originated with a single root word "celebrate." The big problem with stemming is that sometimes it produces the root word which may not have any meaning.

For Example, intelligence, intelligent, and intelligently, all these words are originated with a single root word "intelligen." In English, the word "intelligen" do not have any meaning.

Step 4: Lemmatization

Lemmatization is quite similar to the Stemming. It is used to group different inflected forms of the word, called Lemma. The main difference between Stemming and lemmatization is that it produces the root word, which has a meaning.

For example: In lemmatization, the words intelligence, intelligent, and intelligently has a root word intelligent, which has a meaning.

Step 5: Identifying Stop Words

In English, there are a lot of words that appear very frequently like "is", "and", "the", and "a". NLP pipelines will flag these words as stop words. Stop words might be filtered out before doing any statistical analysis.

Example: He is a good boy.

Note: When you are building a rock band search engine, then you do not ignore the word "The."

Step 6: Dependency Parsing

Dependency Parsing is used to find that how all the words in the sentence are related to each other.

Step 7: POS tags

POS stands for parts of speech, which includes Noun, verb, adverb, and Adjective. It indicates that how a word functions with its meaning as well as grammatically within the sentences. A word has one or more parts of speech based on the context in which it is used.

Example: "Google" something on the Internet.

In the above example, Google is used as a verb, although it is a proper noun.

Step 8: Named Entity Recognition (NER)

Named Entity Recognition (NER) is the process of detecting the named entity such as person name, movie name, organization name, or location.

Example: Steve Jobs introduced iPhone at the Macworld Conference in San Francisco, California.

Step 9: Chunking

Chunking is used to collect the individual piece of information and grouping them into bigger pieces of sentences.

APPLICATIONS OF NLP

There are the following applications of NLP -

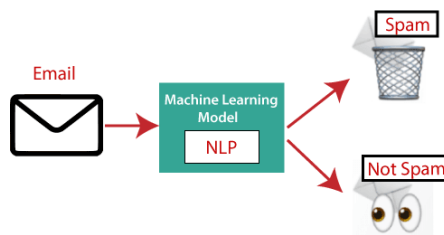
1. Question Answering

Question Answering focuses on building systems that automatically answer the questions asked by humans in a natural language.



2. *Spam Detection

Spam detection is used to detect unwanted e-mails getting to a user's inbox.



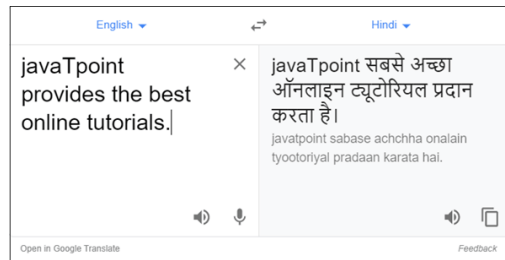
3. Sentiment Analysis

Sentiment Analysis is also known as opinion mining. It is used on the web to analyse the attitude, behaviour, and emotional state of the sender. This application is implemented through a combination of NLP (Natural Language Processing) and statistics by assigning the values to the text (positive, negative, or natural), identify the mood of the context (happy, sad, angry, etc.)



4. Machine Translation

Machine translation is used to translate text or speech from one natural language to another natural language.



Example: Google Translator

5. Spelling correction

Microsoft Corporation provides word processor software like MS-word, PowerPoint for the spelling correction.



6. Speech Recognition

Speech recognition is used for converting spoken words into text. It is used in applications, such as mobile, home automation, video recovery, dictating to Microsoft Word, voice biometrics, voice user interface, and so on.

7. Chatbot

Implementing the Chatbot is one of the important applications of NLP. It is used by many companies to provide the customer's chat services.

8. Information extraction

Information extraction is one of the most important applications of NLP. It is used for extracting structured information from unstructured or semi-structured machine-readable documents.

9. Natural Language Understanding (NLU)

It converts a large set of text into more formal representations such as first-order logic structures that are easier for the computer programs to manipulate notations of the natural language processing

ROLE OF GRAMMAR IN LANGUAGE PROCESSING

UNGRAMMATICALITY

Let us begin by looking at some very simple but real problems in language processing. Consider, for example, how the human language processor might be presumed to deal with a minimal utterance like (1):

(1) Mary snores.

- ✚ Given what we know about the world (and the meaning of lexical items), we can assign an interpretation to Mary (a name designating a female human who can be the actor or agent in an utterance) and one to snore (an action requiring an animate agent involving the production of a particular noise during sleep). We can assume that these meanings are combined to provide a full understanding of the utterance—that there is some female human named Mary who makes noise while she sleeps (see chapter 10 for discussion). Next, consider how the language processor will react to (2):

(2) *Snores Mary.

- ✚ It is clear that (2) is not a sentence of English. By this we mean that native English speakers will reliably reject (2) as well formed even when they have only a rough-and-ready conscious conception of grammaticality. Furthermore, there is evidence suggesting that an English speaker will not process (1) and (2) in the same fashion. Indeed, quick introspection should convince you that (2) is not assigned a normal sentential meaning at all, or, at best, the meaning emerges only after a moment of recognition that the utterance is not "normal" in some sense. There is also experimental evidence that ungrammatical sentences involve longer processing times than grammatical sentences. *Flores d'Arcais (1982) found that syntactic ungrammaticality prolonged reading time even when subjects were not aware of the syntactic violation.*
- ✚ *What makes (2) difficult to process is that the normal syntactic order of simple English sentences is violated.* But it is important to note that since languages do exist where the verb normally precedes the subject (for instance, Tagalog), we cannot assume that the language processor automatically (as a general species characteristic) rejects verb-first utterances as impossible or treats them as intrinsically difficult to process. Rather, we infer from these data that grammaticality supports normal processing and that a determination of ungrammaticality forces the processor into an unusual (and time-consuming) mode.

In a similar vein, consider the two syntactically well-formed sentences in (3):

- (3) a. Selma disqualified Harry.
b. Harry disqualified Selma.

An approach based purely on knowledge about the meaning and referents of the words in (3a) and (3b) will not be able to distinguish between their quite distinct meanings. The discourse context might provide enough information to determine the appropriate meanings, but it might not. Imagine a conversation like the one in (4):

- (4) Speaker A: You'll never guess who disqualified who in the first round!
Speaker B: I wasn't there. How could I know?
Speaker A: Well, I'll tell you . . . Selma disqualified Harry.

Who disqualified whom is a function of the order of linguistic elements. Our knowledge about which linguistic element corresponds to which thematic role (actor, agent, patient,) is not part of our knowledge about things in the world: it is a function of normal syntactic word order, that is, a property of the grammar of the particular language that we speak.

AMBIGUITY

- ✚ In the simple cases above, the grammatical utterances each admitted of only a single interpretation. Even those simple cases required us to recruit information about the syntactic structure of the sentence.
- ✚ Utterances that can be interpreted in more than one way provide an additional perspective on the kind of knowledge the processor requires.

Consider, for example, a simple sentence like *Time flies*. Under one interpretation, it has the idiomatic meaning that "time goes by quickly." But it can also be construed as a command to use a stopwatch to measure insect behavior. *Glue sticks* can be a statement about glue's adhesive qualities or a command to use glue to join pieces of wood. As Waltz and Pollack (1985) note, these kinds of sentences have the feel of a linguistic "Necker cube"—a visual illusion in which we perceive a two-dimensional representation of a cube to flip back and forth between two interpretations of its depth characteristics. As with the Necker cube, we can "see" both meanings, but not at the same time.

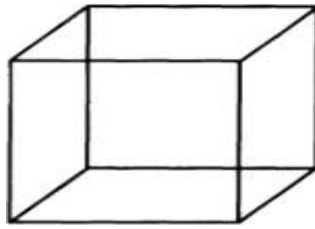


Figure: A Necker cube

It is not difficult to see how these linguistic ambiguities arise. In the cases discussed above both words can belong to more than one syntactic category.

- ✚ In each case *the first word, time or glue, can be interpreted as a noun, which can serve as the simple subject of the sentence, or as a verb, serving to mark the (subjectless) imperative.*
- ✚ *The second word, flies or sticks, can likewise be interpreted as a third-person verb or as a plural noun. Our multiple interpretations of these utterances can be understood quite simply as a function of those alternate categorical properties.*
- ✚ *Interestingly, a third interpretation is also possible for each utterance.* If we take both words to be nouns, glue sticks can be construed as a compound noun meaning "sticks of glue," and time flies can be construed (like dragon flies) as yet another species of insect. But it is not possible, in either case, to construe both of the words in these utterances as verbs. Why should this be? The answer is straightforward if we assume that the mental language processor is paying attention to the grammar of English.
- ✚ *The grammar provides possible structures for the simple declaratives, the imperatives, and the compound nouns*—but it does not countenance sequences of verbs. These words can mean something to us, normally, only within the confines of what is grammatically possible (for English).
- ✚ The above cases strongly suggest that normal interpretation requires at least some purely grammatical information.
- ✚ In other situations where more than a single meaningful interpretation is possible, the roles of grammatical and other types of information may be considerably more complex.
- ✚ Consider, for instance, a sentence like (5), taken from the early work of Winograd (1972), whose SHRDLU program was a pioneering attempt to model natural language processing in the AI framework: (

5) Put the red cube on the block in the box.

Such a sentence can be read in two ways, as suggested by the alternative bracketings in (6):

(6) a. Put [the red cube on the block] in [the box].

b. Put [the red cube] on [the block in the box].

- ✚ In SHRDLU a highly restricted domain of discourse is established in which a "robot" is presumed to inhabit a "blocks world." In this world variously shaped and colored objects exist in a variety of relationships to one another.
- ✚ ***The robot is presented with English sentences that refer to the blocks world and is asked to perform manipulations of objects that are consistent with the properties of that world.*** In this way the robot can be said, in a reasonable sense, to "understand" the sentences that deal with such properties.
- ✚ Winograd's model presumed a certain degree of syntactic knowledge: for example, the ability to parse an utterance like (5) for the purpose of identifying noun phrases. But, as (6a-b) demonstrate, there is more than one such parsing of (5). It is the common, introspective experience of human language users that, in ordinary discourse, ambiguities of this sort do not seem to interfere with the smooth course of processing; we somehow find a single, appropriate reading of ambiguous sentences, and often we must be led explicitly to recognize the existence of a second interpretation.
- ✚ On a largely grammar-driven view of processing, we might well try to explain this perception by discovering general syntactic principles that favor one reading over another; in a more interactive framework, we might expect to uncover properties of the domain of discourse—the immediate context, and knowledge about the world—that direct the processor to favor some particular reading.
- ✚ Winograd's solution utilizes such "real-world" knowledge (or "blocks-world" knowledge, in this limited case). The robot proceeds by examining its simulated world and determining whether in that world there exists a unique reference to some configuration of red cubes and boxes. The program determines whether the noun phrase (NP) [red cube] has a referent. If there is a unique referent (only one red cube), the reading in (6b) is invoked, but if the blocks world contains more than one red cube—say, one on a block and one on the table—the program invokes the reading that is most successful in uniquely identifying a referent, given the information in the sentence.
- ✚ In this case it will be the red cube on the block, the reading in (6a). This principle, called the Principle of Referential Success by Crain and Steedman (1982) in the context of a model of human language processing, is clearly interactive: it is information about the sentence's meaning and its use, rather than principles operating over its structure, that determines the action of the processor.
- ✚ Now it may at first blush seem reasonable that a human being, in the blocks world, would follow a similar strategy. It seems right that a human observer who actually perceived a red cube on a block under these circumstances would ultimately hit on reading (6a). If the task can be accomplished without recourse to any additional intermediate syntactic analysis, it would be more computationally efficient to accomplish it without such a step in the process.

- ✚ Such an assumption may be perfectly appropriate to an enterprise whose goal is to simulate language understanding in a computer: the measure of the program's success is its ability to manipulate its world appropriately, and to do so in a way that places the least burden on the computing device.
- ✚ In the case of human beings, however, there is an additional empirical burden: we cannot simply assume that an approach that is computationally effective for a particular class of contemporary computers is the one that guides human information processing in this domain. The assumption may well be a reasonable default position, in the absence of further evidence. But, as we will now see, there is a body of experimental research that argues against such a view for human language processing.

GARDEN-PATH EFFECTS

Consider the sentence in (7), first discussed by Bever (1970):

(7) The horse raced past the barn fell.

On first hearing or reading it, native speakers of English typically reject this kind of utterance as ungrammatical. It is a classic example of a garden-path sentence in which the listener or reader seems to be "led down a garden path" in assuming that *raced* is the main verb of a simple sentence. Unlike the earlier cases we were considering, this kind of sentence is not ambiguous in the sense of having two overall possible interpretations. Rather, (7) exhibits a kind of temporary or local ambiguity that is resolved as processing progresses. Thus, only when it encounters the final verb *fell* does the garden-pathed processor realize that a mistake has been made: (7) is not a grammatical sentence of English under the assumption that the first six words constitute a simple sentence in which the horse was actively racing. The only (grammatical) interpretation is that the horse was being raced (by an unspecified agent). Consequently, (7) can only mean the same as (8):

(8) The horse that was raced (by someone) past the barn fell.

- ✚ Moreover, the ambiguity disappears when a different past-participle verb form (like *ridden*) is employed. As the processor proceeds to parse a sentence like *The horse ridden past the barn fell*, it is not garden-pathed because *ridden* cannot be analyzed as the main verb of the sentence whose subject is the horse.

A GRAMMATICAL PARSING ANALYSIS

- ✚ Why, then, are we confounded when we try to interpret (7)? Frazier (1987), and others have proposed a processing model that assumes that the speaker initially constructs a syntactic representation of the utterance, quite independently of meaning or context. The model works on input sentences from left to right, assigning each incoming word to some syntactic structure according to the following language-independent principle:

- (9) Minimal Attachment Principle (Clifton and Ferreira 1987)
 Each new incoming item is added to the phrase structure representation with the least possible number of new syntactic phrase nodes needed at the moment of its arrival.

✚ The Minimal Attachment Principle directs the processor to construct the single simplest syntactic constituent that is consistent with incoming data. Assuming the processor is working from left to right, the approach will unfortunately produce an incorrect result in this kind of case: the simplest analysis (part (b) of figure 11.2) presumes that there is no embedded sentence (hence, no additional S node) in the subject NP (compare part (a)) and goes on to make the incorrect assumption that (7) begins with a simple sentence—only to fail when it comes across *fell*, which cannot be attached to the tree, since there is no phrase structure rule of English that will allow a verb as the final element in a prepositional phrase.

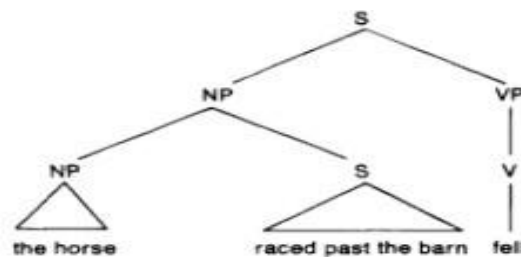


Figure 11.2
 Syntactic structure for the well-formed interpretation of (7)

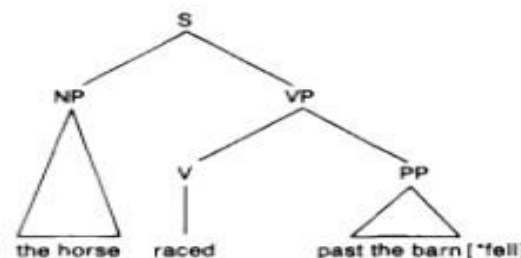


Figure 11.3
 Syntactic structure of the "garden-path" interpretation of (7)

✚ Since the structure in figure 11.3 is "minimal" in that it contains just four major category nodes, whereas the first structure (the correct one) contains five nodes, *the Minimal*

Attachment Principle correctly predicts that the processor will incorrectly initially choose the second structure over the first.

✚ We can discern further evidence for a model incorporating the Minimal Attachment Principle by examining another class of structurally ambiguous sentences. Consider a sentence like (10):

(10) Al played the records on the stereo.

(10) is normally first understood to mean that Al used the stereo as an instrument to play the records. There is also a second, less likely interpretation, namely, that Al played records that were located on the stereo (say, on its plastic cover).

What determines the preferred "instrumental" reading of (10)? The Minimal Attachment Principle makes just the right prediction. Examine the tree structures in figure 11.4.

Observe that in part (a) of figure 11.4, the structure for the instrumental reading, the Prepositional Phrase (PP) on the stereo is attached to (and modifies) the Verb Phrase (VP). It is a simple matter to show that the PP is not inside the NP constituent. For instance, it cannot be passivized as a whole with the NP: The records on the stereo were played by Al is not a paraphrase of the first interpretation. By contrast, the PP in part (b) of figure 11.4, the structure underlying the second interpretation, lies within the (object) NP, and the whole NP is subject to passivization. Indeed, The records on the stereo were played by Al is a proper paraphrase of the non-instrumental reading.

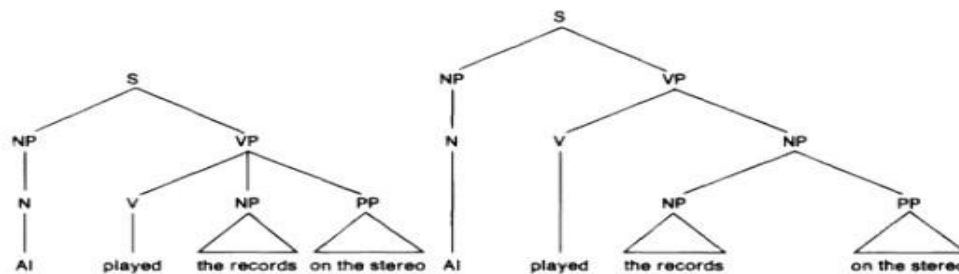


Figure 11.4
Alternative syntactic structures for *Al played the records on the stereo*. (a) is the structure for the instrumental reading, and (b) is the structure for the noninstrumental reading

Furthermore, the structure in part (a) is minimally attached since it contains one node fewer than the structure in part (b), and this explains why it corresponds to the preferred first interpretation of the utterance.

CONNECTIONIST MODELS

- ✚ In the connectionist paradigm it is assumed that knowledge—be it information about the world, memory, or the representation of linguistic generalizations—resides in the complex patterns of connectivity that can arise in the mind as a consequence of the interaction of myriad very simple processing units.
- ✚ In most connectionist theories, knowledge does not take the form of the kind of abstract rules and higher-order structures that are typical of linguistic theory. Notions like Minimal Attachment, or even phrase structure, may have no independent status in a connectionist model, since whatever knowledge they embody is ultimately built out of simple, relatively homogeneous processing units that, by definition, are not uniquely or modularly dedicated to special tasks like language processing.
- ✚ Connectionist models have been developed in a number of domains of interest to the student of NLP. Waltz and Pollack (1985), for example, argue that a form of "semantic" garden-pathing can be insightfully understood within a strong interactive model of processing designed on connectionist principles. They consider sentences like those in (13):
(13) a. The astronomer married the star.
 b. The sailor ate a submarine.
- ✚ Each of these sentences is (lexically) ambiguous. But Waltz and Pollack assert that hearers at first generally experience only one reading, indeed, the semantically bizarre interpretation (the astronomer took a celestial body as a spouse and the sailor consumed a boat). It takes some time and thought to recognize that (13a) can mean that the astronomer married a movie celebrity or that (13b) can mean that the sailor ate a sandwich. In a sense, the processor is led down the garden path once again.
- ✚ Once it assigns a semantic interpretation to sailor or astronomer, it is now primed to make further semantic interpretations within the sentence that are in some sense associated with the meaning of that subject NP. But this priming effect confounds the processor, which expects a verb like eat to have an edible object.

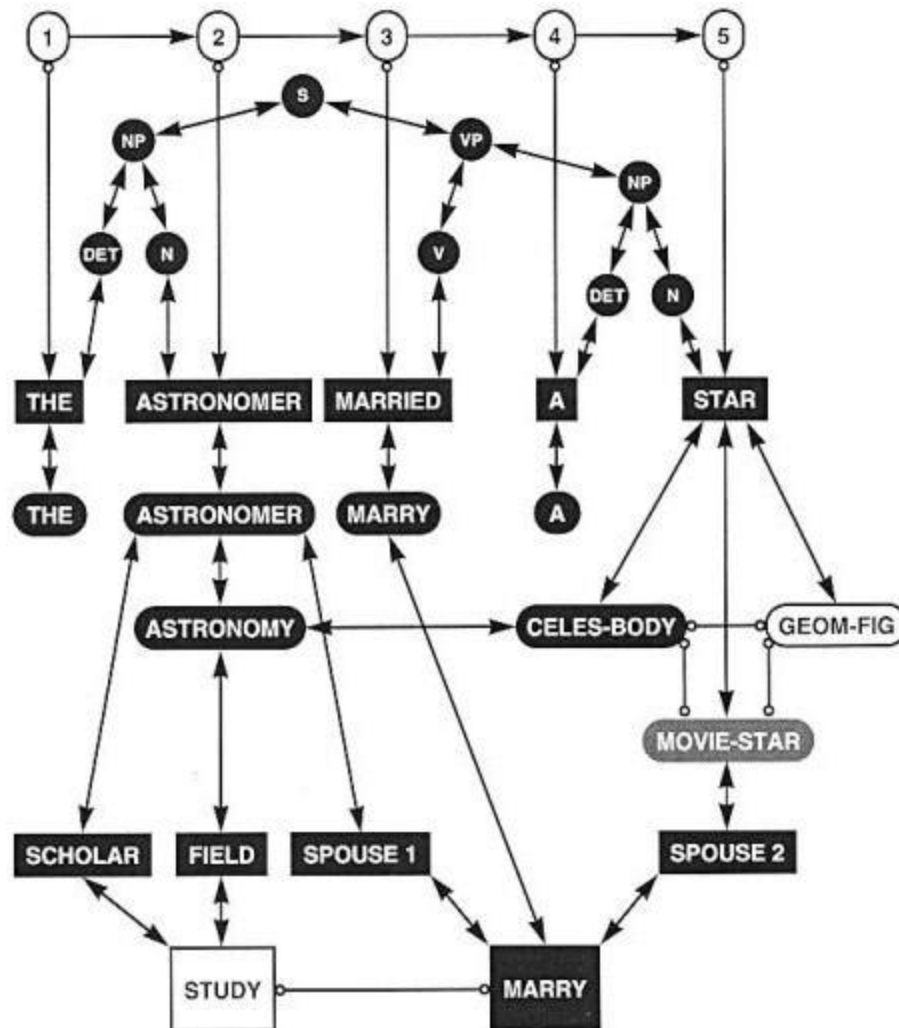


Figure 11.5
A cycle of a connectionist net for the sentence *The astronomer married the star*. (Adapted from Waltz and Pollack 1985.)

- ✚ Accounts of processing in which each lexical item has a discrete, self-contained meaning are hard pressed to account for this effect. But in a distributed connectionist model of the lexicon, Waltz and Pollack argue, the garden-path effect has a straightforward explanation.
- ✚ The subject NP *sailor* can be presumed to have a strong connection to a set of terms with nautical meanings, including *submarine*, and a much weaker connection to the food terms that include the sandwich meaning of *submarine* (see figure 11.5).
- ✚ In such a model the activation of *sailor* will produce a high level of activation for the nautical meaning of *submarine*.
- ✚ Indeed, it will do so relatively quickly, accounting for the nature of the initial interpretation. Ultimately, however, the processor will have to deal with the fact that *eat* also expects an edible object, and the activation level of the sandwich meaning of

submarine will eventually rise, while the processor searches for a connection to a node with foodlike features.

- ✚ An association-rich distributed model of the lexicon also provides a ready framework in which to make sense of the context effects on processing that have been adduced by Marslen-Wilson, Tyler, and others.
- ✚ If PDP modelers ultimately were also able to account directly for the syntactic effects observed by workers like Frazier, Clifton, and their colleagues, the connectionist paradigm would be a major contender as a broadly empirically adequate model of natural language processing.

THE PHYSICAL SYMBOL SYSTEM HYPOTHESIS

✚ In 1975 the Association of Computing Machinery gave their annual Turing Award to two very influential pioneers of artificial intelligence – **Herbert Simon and Allen Newell**. As a great example of the interdisciplinary nature of cognitive science, Simon was actually an economist and political scientist, rather than a computer scientist (as Newell was). Their joint *contributions to computer science included the Logic Theory Machine (1957) and the General Problem Solver (1956), two early and very important programs that developed general strategies for solving formalized symbolic problems.*

✚ Newell and Simon gave a public lecture as one of the conditions of receiving the award. That lecture proposed a bold strategy both for the study of the human mind and for the emerging field of artificial intelligence (AI). *Their manifesto hinged on what they called the physical symbol system hypothesis, which they proposed as a fundamental law for studying intelligence.* For Newell and Simon, the physical symbol system hypothesis is as basic to AI as the principle that the cell is the basic building block of all living organisms is to biology. Here is how they phrased it:

The physical symbol system hypothesis: A physical symbol system has the necessary and sufficient means for general intelligent action.

✚ There are two separate claims here.

- *The first (the necessity claim) is that nothing can be capable of intelligent action unless it is a physical symbol system.* So, since humans are capable of intelligent action, the human mind must be a physical symbol system.
- *The second (the sufficiency claim) is that there is no obstacle in principle to constructing an artificial mind, provided that one tackles the problem by constructing a physical symbol system.*

The significance of these two claims depends on what a physical symbol system is. Here are Newell and Simon again:

- ❖ *A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure).*
- ❖ *Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another).*
- ❖ *At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate*

on expressions to produce other expressions: processes of creation, modification, reproduction, and destruction.

- ❖ *A physical symbol system is a machine that produces through time an evolving collection of symbol structures.*

This passage illustrates four distinctive features of physical symbol systems. Here they are:

1. *Symbols are physical patterns.*
2. *These symbols can be combined to form complex symbol structures.*
3. *The physical symbol system contains processes for manipulating symbols and complex symbol structures.*
4. *The processes for generating and transforming complex symbol structures can themselves be represented by symbols and symbol structures within the system.*

You might have noticed that a *physical symbol system looks very much like an abstract characterization of a digital computer*. That is absolutely correct, as we'll now see.

Symbols and Symbol Systems

- ❖ Newell and Simon make clear in their paper how Turing's work on Turing machines in the 1930s *was the first step toward the physical symbol system hypothesis*. Even though a computer has an alphabet composed of the digits 0 and 1, we will not find any 0s and 1s in it if we open it up. If we dig down deep enough, all that there is to a computer is electricity flowing through circuits. If an electrical circuit functions as an on/off switch, then we can view that switch in symbolic terms as representing either a 0 (when it is off) or a 1 (when it is on). But there are no digits to be found in the circuit.
- ❖ (2) *Symbols can be combined to form complex symbol structures*. Just as letters can be put together to form words, the symbols in any physical symbol system can be combined to form word-like symbol structures. Those word-like structures can then be put together to form sentence-like structures. Both types of combination are governed by strict rules.
- ❖ You can think of these strict rules as telling the symbol system which combinations of symbols count as grammatical. These rules are likely to be recursive in form. That means that they will show how to get from an acceptable combination of symbols to a more complex combination that is still acceptable.
- ❖ The rules for how to define what counts as a sentence in the branch of logic known as sentential logic or propositional logic provide a good illustration of recursive rules and how they work. See Box 4.1.

- ❖ Turing machines can scan only a single cell at a time, but they are still capable of working with complex symbol structures because those complex symbol structures can be built up from individual symbols in adjacent cells.
- ❖ The Turing machine needs to know two things: It needs to know what symbols can follow other symbols. And it needs some way of marking the end of complex symbols. The first can come from instructions in the machine table, while for the second there are symbols that serve as punctuation marks, effectively telling the scanner when it has arrived at the end of a complex symbol.

BOX 4.1 Defining Sentences in Propositional Logic

Propositional logic studies the logical relations holding between whole sentences, or propositions. The language of propositional logic is very simple. It contains basic symbols for sentences (such as "P," "Q," and "R"), together with a small set of logical connectives.

A typical formulation of propositional logic might have three connectives (the so-called Boolean connectives). These are " \neg ," read as "not-"; " \vee ," read as "or"; and " \wedge ," read as "and."

These logical connectives allow sentence symbols to be combined to form more complex sentences. So, for example, " $P \wedge Q$ " is a sentence. It is true just when the two sentences P and Q are both true.

Propositional logic has clear and unambiguous rules for determining what counts as a legitimate sentence. These rules fix when the rules governing the connectives have been correctly applied. The legitimate combinations of symbols in the alphabet might typically be defined as follows.

- (a) Any sentence symbol is a sentence.
- (b) If " φ " is a sentence then " $\neg \varphi$ " is a sentence.
- (c) If " φ " and " ψ " are sentences, then " $\varphi \wedge \psi$ " is a sentence.
- (d) If " φ " and " ψ " are sentences, then " $\varphi \vee \psi$ " is a sentence.

These are examples of what are called *recursive rules*. They show how, starting with a basic set of sentences (the sentence symbols), you can construct arbitrarily complex formulas that will count as genuine sentences.

Note that " φ " and " ψ " can stand here for any formula, not just for sentence symbols. So you can apply the recursive definition to show that $\neg(P \wedge \neg P)$ is a genuine sentence of propositional logic.

Can you see how? (Hint: If P is a sentence symbol, then it is a sentence, by (a). If P is a sentence, then so is $\neg P$, by (b). Continue in this vein.)

Transforming Symbol Structures

The third feature of physical symbol systems involves transformation.

(3) The physical symbol system contains processes for manipulating symbols and complex symbol structures. Here we have the distinctive claim of the physical symbol system hypothesis. Thinking is no more (and no less) than transforming symbol structures according to rules. Any system that can transform symbol structures in a sophisticated enough way will qualify as intelligent. ***According to Newell and Simon when we fully understand what is going on in intelligent agents***

(such as human beings), all we will ultimately find is symbol structures being transformed in rule-governed ways.

✚ In the background here is Newell and Simon's *fundamental idea that the essence of intelligent thinking is the ability to solve problems. Intelligence consists in the ability to work out, when confronted with a range of options, which of those options best matches certain requirements and constraints. Intelligence cannot be applied without what might abstractly be called a search-space.* The notion of a search-space is very general. Consider, for example, the position of one of the players halfway through a chess match. Each chess player has a large number of possible moves and a clearly defined aim – to checkmate her opponent. The possible moves define the search-space and the problem is deciding which of the possible moves will move her closest to her goal.

✚ Another example (much studied by computer scientists and mathematicians) is a traveling salesperson who starts in a particular city (say, Boston) and has to visit twenty other cities as quickly and efficiently as possible before eventually returning to Boston. Here we can think about the search-space in terms of all the possible routes that start and end in Boston and go through the twenty cities (perhaps visiting some more than once). The diagram at the top in Figure 4.1 illustrates a simpler traveling salesperson problem with only five cities (a, b, c, d, and e).

Search-spaces are typically represented in terms of states. There is an initial state (the start state) and a set of permissible transformations of that start state. The search-space consists of all the states that can be reached from the start state by applying the permissible transformations. The transformations can be carried out in any order.

- ✓ In the chess example, the start state is a particular configuration of the chess pieces and the permissible transformations are the legal moves in chess.
- ✓ In the traveling salesman problem, the start state might be Boston, for example, and the permissible transformations are given by all the ways of getting directly from one city to another. This means that each state of the traveling salesman problem is given by the current city, together with the cities already covered and the cities still left to visit.
- ✓ Computer scientists standardly represent search-spaces in terms of trees. So, for example, the search-space for the traveling salesperson problem is given by a tree whose first node is the starting city.
- ✓ The diagram at the bottom of Figure 4.1 illustrates a part of the search-space for our five-city version of the traveling salesperson problem. A branch from the first node (a, the start city) goes to a node representing each city to which the start city is directly connected – i.e., cities b, c, d, and e. From each of those nodes, further branches connect each city to all the other cities to which it is directly connected. And so on.

- ✓ What counts as solving a problem? Basically, you've solved a problem when you've found the solution state in the search-space. In chess, the solution state is any configuration of the board in which the opponent's king is in checkmate. In the traveling salesperson problem, the solution is the shortest branch of the tree that ends with Boston and that has nodes on it corresponding to each city that the salesman needs to visit. But how is that done?
- ✓ Obviously, you have to search through the search-space until you find a solution state. But this can be much harder than it sounds. Brute force searches that follow each branch of the tree typically only work for very simple problems. It does not take long for a problem space to get so big that it cannot be exhaustively searched in any feasible amount of time.

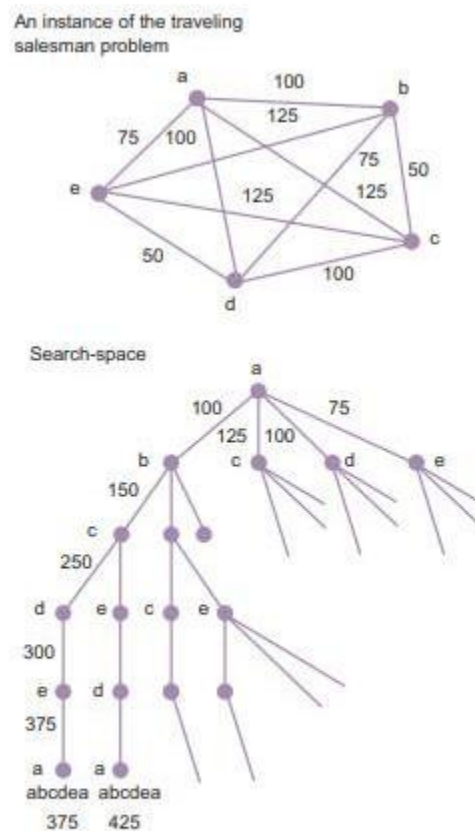


Figure 4.1 A typical traveling salesperson problem. The top diagram depicts the problem. A traveling salesperson has to find the shortest route between five cities. The diagram below depicts part of the search-space. A complete representation of the search-space would show twenty-four different routes.

- ✓ The traveling salesperson problem is a great example. If there are n cities, then it turns out that there are $(n-1)!$ possible routes to take into account, where $(n-1)! = (n-1) \times (n-2) \times (n-3) \times \dots$. This number of routes is not too many for the five-city version of the problem depicted in Figure 4.1 (it gives twenty-four different routes). But the problem gets

out of control very quickly. A twenty-city version gives approximately 6^{20} ways for a traveling salesperson to start in Boston and travel through the other nineteen cities visiting each exactly once. Checking one route per second, we would need more than the entire history of the universe to search the problem space exhaustively.

Newell and Simon developed their General Problem Solver (GPS) program as a way of solving problems of this type (although they focused on much simpler problems than the traveling salesperson problem, which still has no general solution).

The basic idea behind the GPS program is relatively straightforward. The program uses means–end analysis. Here are simple instructions for applying means–end analysis.

- 1. Evaluate the difference between the current state and the solution state.*
- 2. Identify a transformation that reduces the difference between the current state and the solution state.*
- 3. Check that the transformation in (2) can be applied to the current state.*
 - 3a. If it can, then apply it and go back to step (1).*
 - 3b. If it can't, then return to (2).*

Means–end analysis is an example of what Newell and Simon call heuristic search. Heuristic search techniques are techniques for searching through a search-space that do not involve exhaustively tracing every branch in the tree until a solution is found. Heuristic search techniques reduce the size of the search-space in order to make the search process more manageable.

Exercise 4.1 Explain how means–end analysis makes it more manageable to search the search-space.

- Here is the problem of the foxes and the chickens – a type of problem that Newell and Simon showed could be solved by their GPS program. Imagine that there are three chickens and three foxes on one side of a river and they all need to get over to the other side. The only way to cross the river is in a boat that can take only two animals (or fewer) at a time. The boat can cross in either direction, but if at any point the foxes outnumber the chickens then the outnumbered chickens will be eaten. How can you get all the chickens and foxes onto the other side of the river without any of the chickens being eaten?
- Here each state specifies which animals are on each bank and which in the boat (as well as the direction in which the boat is traveling). The start state obviously has all six on one bank (say the left bank) with nobody in the boat or on the other bank. The solution state is the state that has all six on the right bank, with nobody in the boat or on the other bank. The permissible transformations are defined by the rule that the boat cannot carry more than two animals at a time.

- The foxes and chickens problem is a great example of how the GPS program works. If we feed into the GPS program representations of the start state and the solution, or goal state, the program employs various transformation strategies to minimize the difference between the start state and the goal state. The eventual solution is a series of representations, whose first member is a representation of the start state and whose final member is a representation of one of the goal states, and where each member is derived from its predecessor by a permissible transformation.

Exercise 4.2 Find a solution to the foxes and chickens problem.

- A final point. These rule-governed transformations are algorithmic. An algorithm is a finite set of unambiguous rules that can be applied systematically to transform an object or set of objects in definite and circumscribed ways. Algorithms are purely mechanical procedures. They can be followed blindly, without any exercise of judgment or intuition. Elementary school arithmetic provides plenty of examples of algorithms, such as the algorithms for multiplying pairs of numbers and for long division.

Intelligent Action and the Physical Symbol System

- ✓ *The final feature of physical symbol systems is what really makes symbol systems capable of intelligent action.*

(4) *The processes for generating and transforming complex symbol structures can themselves be represented by symbols and symbol structures within the system. A fundamental feature of modern computers – so familiar that most of us never think about it – is the fact that a single computer (a single piece of hardware) can run many different programs, often simultaneously.* This capability is what distinguishes a general-purpose computer from a specialized computing machine such as a pocket calculator. Computers can be programmed in this way because they can contain symbol structures that encode information about, and instructions for, other symbol structures.

Alan Turing proved that it is possible to construct a special kind of Turing machine (a universal Turing machine) that can mimic any specialized Turing machine implementing a particular algorithm. The universal Turing machine is a general-purpose computer. You can think of the specialized computers as software programs that run on the more general operating system of the universal Turing machine. *The universal Turing machine is possible because Turing machine tables can be encoded as numbers, and hence can serve as inputs to Turing machines. The physical symbol system hypothesis builds something like this feature into the characterization of an intelligent system.*

FROM PHYSICAL SYMBOL SYSTEMS TO THE LANGUAGE OF THOUGHT

- The physical symbol system hypothesis tells us that intelligent agents solve problems by physically transforming symbolic structures. But we still need to know what these symbolic structures are, how they are transformed, and how those transformations give rise to intelligent action of the sort that human beings might carry out. This section looks at a proposal for answering these questions. This is the language of thought hypothesis developed by the philosopher and cognitive scientist Jerry Fodor (1935–2017).
- According to Fodor’s language of thought hypothesis, the basic symbol structures in the mind that carry information are sentences in an internal language of thought (sometimes called *Mentalese*). Information processing works by transforming those sentences in the language of thought.
- Our starting point for exploring this idea is the basic fact that the mind receives information about its environment. Some of this information is carried by light waves arriving at the retina or sound waves hitting the eardrum. But in general, our behavior is not determined by the information that we receive.
- Different people, or the same person at different times, react differently to the same situation. There is no standard response to the pattern of sound waves associated (in English) with a cry of “Help!” for example. It is a basic assumption of cognitive science that information processing is, at bottom, a matter of transforming these representations in a way that finally yields the activity in the nervous system that “instructs” my limbs to jump into the water.
- Among all these different types of representation, ***Fodor is particularly interested in the ones that correspond to beliefs, desires, and other similar psychological states. These psychological states are often called propositional attitudes by philosophers. They are called this because they can be analyzed as attitudes to propositions.***
- ***Propositions are the sorts of thing that are expressed by ordinary sentences.*** So, there is a proposition expressed by the sentence “That person will drown” or by the sentence “It is snowing in St. Louis.” Thinkers can have different attitudes to those propositions. I might fear the first, for example, and believe the second.
- ***Fodor’s starting point in thinking about propositional attitudes is that*** we are, by and large, pretty good at explaining and predicting other people’s behavior in terms of what they believe about the world and what they want to achieve. He thinks that ***this success is something that itself needs explanation.*** Why is the vocabulary of beliefs and desires (our belief–desire psychology or propositional attitude psychology) so deeply ingrained in us? Why does it seem so indispensable in our social interactions and social coordination? How and why do explanations that appeal to beliefs and desires actually work? And, in particular, why are these explanations so successful?

*According to Fodor, there can be only one possible answer. **Belief–desire psychology is successful because it is true. There really are such things as beliefs and desires. They are physical items that cause us to behave in certain ways. Belief–desire explanations are successful when they correctly identify the beliefs and other states that caused us to act in the way that we did.***

If we say that someone jumped into the water because she believed that a child was drowning and wanted to save him, then what we are really claiming is that that person's bodily behavior was caused by internal items corresponding to the belief that someone is drowning and the desire to save her. This view is often called intentional realism.

Fodor's argument for the language of thought hypothesis is, in essence, that the hypothesis of intentional realism is the only way of explaining how belief–desire explanations can work. We will examine his argument in the next two subsections.

Intentional Realism and Causation by Content

- ✓ Intentional realism treats beliefs and desires as the sorts of things that can cause behavior. But this is a special type of causation. There is a fundamental difference between my leg moving because I am trying to achieve something (perhaps the journey of a thousand miles that starts with a single step) and my leg moving because a doctor has hit my knee with his hammer. In the first case, what causes my movement is what the desire is a desire for, namely, the beginning of the journey of a thousand miles. This is what philosophers call the content of the desire. There is nothing corresponding to a desire with content (or any other state with content) when a doctor hits my knee with a hammer. The movement that I make is simply a response to physical stimulus. It is not a response to something that I want to achieve.
- ✓ Intentional realism treats beliefs and desires as the sorts of things that can cause behavior. But this is a special type of causation. There is a fundamental difference between my leg moving because I am trying to achieve something (perhaps the journey of a thousand miles that starts with a single step) and my leg moving because a doctor has hit my knee with his hammer. In the first case, what causes my movement is what the desire is a desire for, namely, the beginning of the journey of a thousand miles. This is what philosophers call the content of the desire. There is nothing corresponding to a desire with content (or any other state with content) when a doctor hits my knee with a hammer. The movement that I make is simply a response to physical stimulus. It is not a response to something that I want to achieve.

Let us call the physical properties that can be manipulated within brains formal properties. We call them this because they have to do with the physical form (i.e., the shape) of the representation.

And let's use semantics for the properties that enable representations to represent – just as semantics is the branch of linguistics that deals with the meanings of words (how words represent).

This gives us another way of putting the problem. How can the brain be an information-processing machine if it is blind to the semantic properties of representations? How can the brain be an information-processing machine if all it can process are the formal properties of representations?

At this point, we can see the particular slant that Fodor is putting on the physical symbol system hypothesis. Computers essentially manipulate strings of symbols. A computer programmed in binary, for example, manipulates strings of 1s and 0s. This string of 1s and 0s might represent a natural number, in the way that in binary 10 represents the number 2 and 11 represents the number 3. Or it might represent something completely different. It might represent whether or not the individual members of a long series of pixels are on or off, for example.

In fact, with a suitable coding, a string of 1s and 0s can represent just about anything. As far as the computer is concerned, however, what the string of 1s and 0s represents is completely irrelevant. The semantic properties of the string are irrelevant. The computer simply manipulates the formal properties of the string of 1s and 0s. In fact, it would be more accurate to say that the computer operates on numerals rather than numbers. Numerals are just symbols with particular shapes. Numbers are what those numerals represent.

But here's where the computer program comes in. The computer is programmed to manipulate strings of 1s and 0s in certain ways that yield the right result, even though the computer has no idea what that right result is. Take an adding machine, for example. Suppose it is given two strings of 0s and 1s and in response outputs a third string of 1s and 0s. If the first two strings represent the numbers 5 and 7, respectively, then (if the machine is well designed) the third string will be a binary representation of the number 12.

But even though all the computer is doing is mechanically manipulating 1s and 0s (numerals not numbers), operating on their formal properties, it nonetheless comes up with the right answer, all. So, although the computer itself is not concerned with what "12" means, the computer program must respect the rules of addition in order for the computational result – "12" – to have meaning.

In essence, what computer programmers do when they are programming an adding machine, is writing code so that purely mechanical manipulations of numerals will correctly track arithmetical relations between the numbers that the numerals represent. So, the adding machine must manipulate the numerals "7" and "5" in such a way that taking them as inputs to the machine yields the numeral "12," because it is an arithmetical fact that $7 + 5 = 12$ (which is a statement about numbers, not numerals).

Fodor thinks that way of thinking about computer programs is a great model for the human brain. Brains are physical systems that can be sensitive only to the formal properties of mental representations. But nonetheless, as information-processing machines, they (like computers) have to respect the semantic properties of mental representations. The language of thought is what makes this possible.

The Language of Thought and the Relation between Syntax and Semantics

here are the three main claims of Fodor's language of thought hypothesis.

- 1. causation through content takes place through causal interactions between physical states.*
- 2. these physical states have the structure of sentences, and their sentence-like structure determines how they are made up and how they interact with each other.*
- 3. causal transitions between sentences in the language of thought respect the rational relations between the contents of those sentences (what they mean).*

- According to Fodor, we think in sentences, but these are not sentences of a natural language such as English. The language of thought is much closer to a logical language, such as the propositional calculus. It is supposed to be free of the nuances, ambiguities, and multiple layers of meaning that we find in English and other natural languages.

The analogy between the language of thought and logical languages is at the heart of Fodor's solution to the problem of causation by content. It is what lies behind claim

(3). The basic fact about formal languages that Fodor exploits is the clear separation that they incorporate between syntax and semantics.

Syntax has to do with symbols and the rules for combining them. You can think of it as the logical equivalent of grammar.

Semantics, on the other hand, has to do with what the symbols actually mean and, relatedly, to what makes sentences true (or false).

- *To illustrate this general distinction, we can use the predicate calculus. This is a logical language more powerful and sophisticated than the propositional calculus we looked at in Box 4.1. Unlike the propositional calculus (which only allows us to talk about complete sentences or propositions) the predicate calculus allows us to talk directly about individuals and their properties.*

So, for example, the predicate calculus allows us to formalize inferences such as:

Hubert is laughing

Therefore, someone is laughing

Or:

Everyone is laughing

Therefore, Hubert is laughing

- *In order to represent these inferences, the predicate calculus has special symbols. These special symbols include individual constants that name particular objects, and predicate letters that serve to name properties.* The symbols are typically identifiable by simple typographical features (such as uppercase for predicate letters and lowercase for individual constants) and they can be combined to make complex symbols according to certain rules. *It also includes quantifiers, which are logical expressions corresponding to the English words “some” and “all.”*
- *From a syntactic point of view, a formal language such as the predicate calculus is simply a set of symbols of various types together with rules for manipulating those symbols according to their types.* These rules identify the symbols only in terms of their typographical features.
- An example would be the rule that the space after an uppercase letter (e.g., the space in “F–”) can be filled only with a lowercase letter (e.g., “a”). This rule is a way of capturing at the syntactic level the intuitive thought that properties apply primarily to things – because uppercase letters (such as “F–”) can only be names of properties, while lowercase letters (such as “a”) can only be names of objects. The rule achieves this, however, without explicitly stating anything about objects and properties. It just talks about symbols. It is a matter purely of the syntax of the language.

The connection between the formal system, on the one hand and what it is about, on the other, comes at the level of semantics. When we think about the semantics of a formal language, we assign objects to the individual constants and properties to the predicates. We identify the particular object that each individual constant names, for example. To provide a semantics for a language is to give an interpretation to the symbols it contains – to turn it from a collection of meaningless symbols into a representational system.

Fodor’s basic proposal is that we understand the relation between sentences in the language of thought and their content (or meaning) on the model of the relation between syntax and semantics in a formal system. Sentences in the language of thought can be viewed purely syntactically. From the syntactic point of view, they are physical symbol structures composed of basic symbols arranged according to certain rules of composition. Or they can be viewed semantically in terms of how they represent the world (in which case they are being viewed as the vehicles of propositional attitudes).

So now, suppose we think that the causal transitions between sentences in the language of thought are essentially syntactic, that is, sensitive only to the formal properties of the relevant symbols, regardless of the symbols' meanings. Then we need to ask the following question:

Why do the syntactic relations between sentences in the language of thought map onto the semantic relations holding between the contents of those sentences?

If we take seriously the idea that the language of thought is a formal system, then this question has a perfectly straightforward answer. Syntactic transitions between sentences in the language of thought track semantic transitions between the contents of those sentences for precisely the same reason that syntax tracks semantics in any properly designed formal system.

Fodor can (and does) appeal to well-known results in meta-logic (the study of the expressive capacities and formal structure of logical systems). These results establish a significant degree of correspondence between syntax and semantics. So, for example, it is known that the first-order predicate calculus is sound and complete. That is to say, *in every well-formed proof in the first-order predicate calculus the conclusion really is a logical consequence of the premises (soundness) and, conversely, for every argument in which the conclusion follows logically from the premises and both conclusion and premises are formulable in the first-order predicate calculus there is a well-formed proof (completeness).*

The combination of soundness and completeness has the following important consequences. If a series of legitimate and formally definable syntactic transformations lead from formula A to a second formula B, then one can be sure that A cannot be true without B being true – and, conversely, if A entails B in a semantic sense then one can be sure that there will be a series of formally definable inferential transitions leading from A to B.

Here's an example. Suppose that we have two complex symbols, "Fa" and "Ga." Each of these symbols is a sentence in the language of thought with a particular syntactic shape. We know that "F–" and "G–" are symbols for predicates. Let us say that "F–" means "– is tall" and "G–" means "– has red hair." We also know that "a" is a name symbol. Let us say that "a" names Georgina. The meaning of "Fa" is that Georgina is tall, while the meaning of "Ga" is that Georgina has red hair.

Table 4.1 shows how a very simple piece of thinking might be analyzed by the language of thought hypothesis.

TABLE 4.1 Syntax and semantics in the predicate calculus		
SYMBOLS	TRANSFORMATION RULE	MEANING
1. Fa		1. Georgina is tall
2. Ga		2. Georgina has red hair
3. $(Fa \ \& \ Ga)$	If complex symbols "S" and "T" appear on earlier lines, then it is legitimate to write "(S & T)"	3. Georgina is tall and has red hair
4. $\exists x (Fx \ \& \ Gx)$	If on an earlier line there is a complex symbol containing a name symbol, then it is legitimate to replace the name symbol by "x" and write " $\exists x -$ " in front of the complex symbol [NOTE: " $\exists x -$ " is the symbol for "there is at least one x such that -"]	4. At least one person is tall and has red hair

The table shows how two physical symbols: " Fa " and " Ga " are transformed in two inferential steps into the more complex physical symbol " $\exists x (Fx \ \& \ Gx)$." Here " $\exists x -$ " is the symbol for "there is at least one x such that -," so that this sentence means "There is at least one thing that is both F and G."

The rules that achieve this transformation are purely syntactic. They are simply rules for manipulating symbol structures. But when we look at the relation between the meanings of " Fa " and " Ga ," on the one hand, and the meaning of " $\exists x (Fx \ \& \ Gx)$," on the other, we see that those purely syntactic transformations preserve the logical relations between the propositions that the symbols stand for. If it is true that Georgina is tall and that Georgina has red hair, then it is certainly true that at least one person is tall and has red hair.

In sum, beliefs and desires are realized by language-like physical structures (sentences in the language of thought), and practical reasoning and other forms of thinking are ultimately just causal interactions between those structures. These causal interactions are sensitive only to the formal, syntactic properties of the physical structures. Yet, because the language of thought is a formal language with analogs of the formal properties of soundness and completeness, these purely syntactic transitions respect the semantic relations between the contents of the relevant beliefs and desires. This is how (Fodor claims) causation by content can take place in a purely physical system such as the human brain. And so, he argues, *common-sense psychological explanation is vindicated by thinking of the mind as a computer processing sentences in the language of thought.*

The line of reasoning that leads to the language of thought hypothesis is fairly complicated. To make it easier to keep track of the different steps, I have represented them diagrammatically in Figure 4.2.



Exercise 4.7 Use the flowchart in Figure 4.2 to explain Fodor's argument in your own words.

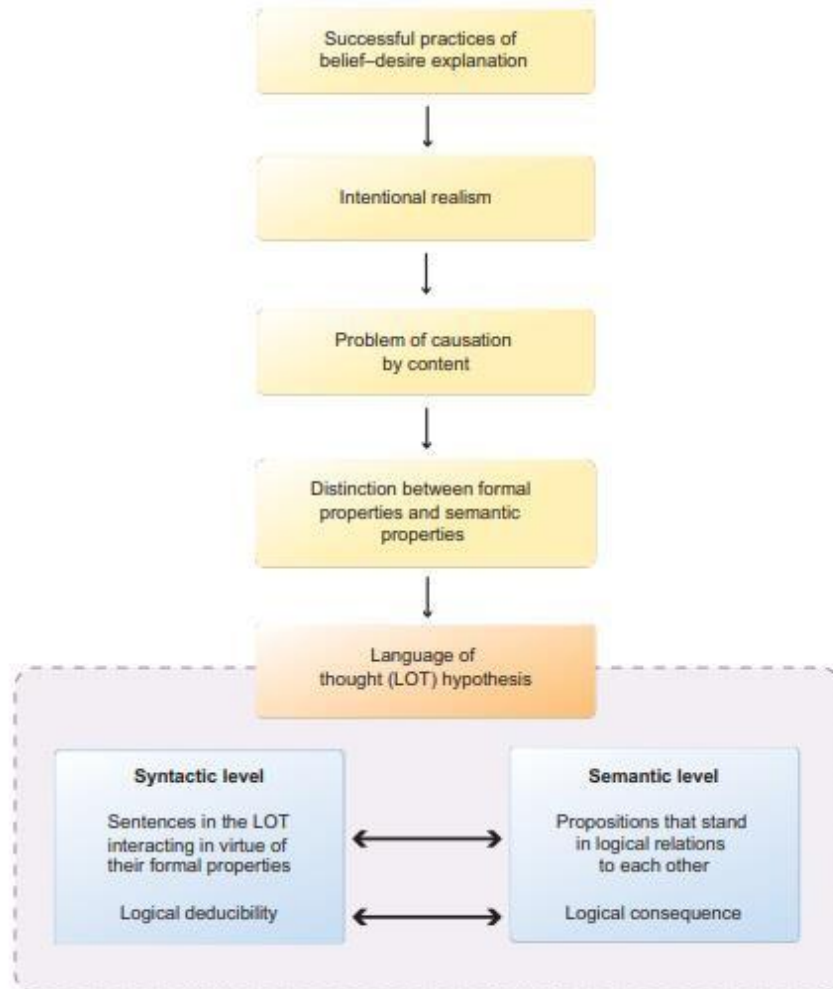


Figure 4.2 The structure of Fodor's argument for the language of thought hypothesis.

PERCEPTION: COGNITIVE BIAS PROBLEM, IS AI BIASED? IMAGE ANALYTICS, VIDEO ANALYTICS & AUDIO ANALYTICS

COGNITIVE BIAS

- ✚ This section describes what is understood by cognitive biases and how this universal psychological phenomenon, which most prominently occurs in human judgment and decision making, can be explained. ***People are constantly forming judgments and making decisions, both consciously and unconsciously, without certainty about their consequences.***
- ✚ The decision to invest money in shares, to start a new project, or to move to a new house is generally made without knowing beforehand whether stock prices will rise, how internal and contextual success-factors will develop, or what is it like to live in that other place.
- ✚ Making conscious decisions in uncertainty is based on two main characteristics of the possible outcome: the (un)desirability of this possible outcome, and the probability of this outcome.
- ✚ ***Rational choice theory is concerned with the development of methods (algorithms) that provide an optimal choice for given problems and probabilities. The rational decision-maker is assumed to take as much as possible account of all available information, probabilities of events, potential costs, and benefits of possible outcomes, and select the optimal choice of action*** (Homans, 1961; Scott, 2000).
- ✚ ***In most situations, however, people make decisions in a more “intuitive” way, substantially deviating from the assumptions of formal choice models. Human decision making, in general, does not meet the criteria prescribed by the rules of logic, probability reasoning, formal cost-benefit models, and prudent considerations*** (e.g., Gigerenzer and Gaissmaier, 2011; Kahneman, 2011; Kahneman et al., 1982; Shafir and LeBoeuf, 2002; Tversky and Kahneman, 1974).
- ✚ There are controversies as to whether this deviation from the tenets of rational reasoning should be considered as “irrational”.
- ✚ Given limitations in the available information and in our human information processing capacities, intuitive information processing can be quite effective, leading to acceptable or satisfying outcomes, for example under time pressure or when the relevant or available information is too extensive or detailed, or when no optimal solution is evident (e.g., “Bounded Rationality”; Simon, 1955). In these cases, we use pragmatic decision routines, characterized by a high ratio of benefits to cost in terms of the quality of the outcomes relative to invested time, effort, and resources (e.g., Gigerenzer and Todd, 1999).
- ✚ The terms “heuristic” or “mental shortcut” are typically used to qualify this supposed smart and efficient way of dealing with decision situations that are complex, uncertain or time-critical (e.g., Gigerenzer and Gaissmaier, 2011; Gigerenzer and Goldstein, 1996; Hardman, 2009; Myers, 2010). According to this heuristic vision, people can be considered “rational” in the sense that²⁵ they use their limited information processing capabilities adaptively and as efficiently as possible to achieve their goals.
- ✚ ***This heuristic way of dealing with our information processing limitations and data***

limitations does not alter the fact that many of our decisions substantially deviate from what may be considered optimal, advisable, or utile (in relation to the goals we pursue). In addition, it has been shown that these “irrational” tendencies are not random, but very specific and systematic: in a wide range of different conditions, people show the same, typical tendencies in the way they pick up and process information in order to judge and decide (e.g., Hastie and Dawes, 2001; Kahneman, 2011; Kahneman and Tversky, 2000; Korteling et al., 2018; Shafir and LeBoeuf, 2002).

- ✚ ***The concept of “cognitive biases” is most commonly used to describe and qualify these deviations in human decision making*** (Tversky and Kahneman, 1974). Cognitive biases can generally be described as systematic and universally occurring tendencies, inclinations, or dispositions that skew or distort information processes in ways that make their outcome inaccurate, suboptimal or simply wrong (e.g., Lichtenstein and Slovic, 1971; Tversky and Kahneman, 1981).
- ✚ So, these deviations are not random, but specific and systematic. The intuitive processes that precede biased judgments and decisions, and that are at the basis of the behavior that is ultimately shown, are largely implicit and unconscious. That is why they can also, and maybe better, be termed or qualified as “a-rational” instead of “ir-rational”.
- ✚ Many cognitive biases seem robust and universal psychological phenomena, extensively demonstrated, described and analyzed in the scientific literature. They apply to (almost) everybody, at all levels and in all parts of society, not only in our daily life, but also in professional institutions like politics, government, business, and media. They are also pervasive and persistent. We typically feel quite confident about our decisions and judgments, even when evidence is scarce and when we are aware of our cognitive biases (Eigenauer, 2018; Pronin et al., 2002; Risen, 2015).
- ✚ We also notice the biased reasoning in others more than in ourselves. This is termed the Bias blind spot (which could be compared with the visual blind spot: Korteling et al., 2018). Pronin et al. (2002) demonstrated in a sample of over 600 US residents that more than 85% believed they were less biased than the average American. Only one participant believed that he or she was more biased than the average American. This effect remained even after receiving information about the general nature and operation of biases. Many different biases have been identified so far (see Glossary, giving brief descriptions of the biases mentioned in this section).

Psychological Frameworks

- ✚ ***The psychological*** literature distinguishes many different biases. Most publications provide demonstrations of biases and describe their phenomenology, together with the factors and conditions that trigger them. Since many biases resemble each other, there is a high degree of overlap between them, and one bias is often a specific example of another (more generic) bias.
- ✚ For example, wellknown tendencies and biases, such as Prejudice, Conservatism, Familiarity bias, Confirmation bias, Status quo bias, System justification, Normalcy bias, and Default effect all have in common that we prefer information that is compatible with or confirms the existing state of affairs. The lack of theory formation on the origin of the multitude of biases calls for a more unifying and binding framework of underlying mechanisms in which biases are categorized based on their

similarities and their underlying cause.

- ✚ *Psychological framework makes the structure and relationships between the abundance of bias phenomena clearer and more manageable, thereby enabling the development of methods and tools to handle them. For that purpose, we will first discuss the recent insights about the origin of biases from a psychological point of view.*

Capacity-Based Frameworks

- ✚ *Capacity-based frameworks (“Heuristics & Biases” perspectives) attribute cognitive biases to limitations in the available data and in the human information processing capacity (Broadbent, 1958; Evans, 2008; Kahneman, 1973, 2003; Morewedge and Kahneman, 2010; Norman and Bobrow, 1975; Simon, 1955). In this view, decision-makers tend to use simple heuristics in complex, unfamiliar, uncertain, and time-constrained situations because they can only process a limited amount of the available information (“limited-” or “bounded rationality”: Gigerenzer and Selten, 2002; Kahneman, 1973; Norman and Bobrow, 1975; Simon, 1955). In most cases, this may yield nevertheless acceptable outcomes.*
- ✚ However, capacity limitations may result in decision errors when people select simple or improper cues and ignore or inappropriately weigh relevant information (Evans, 2008; Kahneman, 2003; Kahneman and Klein, 2009). This view has resulted in the Dual Processing framework (Evans, 2008; Kahneman, 2003, 2011; Stanovich and West, 2001).
- ✚ This model assumes that people can process information in two ways: the first way (Type 1 thinking) works fast, automatically, unconsciously and requires little effort. This is the default mode that is sensitive to biases.
- ✚ *Type 1 thinking feels natural and obvious and quickly provides intuitive, self-evident answers to problems as they arise. Type 1 processes do not demand executive working memory resources and operate implicitly, in parallel, and are highly accessible (De Neys, 2006; Kahneman, 2003). The terms “heuristic” or “intuitive” are most often used to describe this way of information processing.*
- ✚ *The second way of thinking (Type 2 thinking) is slower, conscious and requires concentration and attention. Type 2 processes rely on time- and resource-consuming serial operations on the available data and are constrained by the limited capacity of the central working memory (Baddeley, 1986; Baddeley and Hitch, 1974; Evans and Stanovich, 2013). Type 2 thinking requires effort and thereby processing capacity. Hence, we cannot simultaneously perform multiple Type 2 operations.*
- ✚ *The terms “deliberate” or “analytical” are often used to describe this type of information processing. Under some circumstances (e.g., after training) humans can deliberately use Type 2 processing to monitor a solution that is initially provided by intuition.*
- ✚ If this monitoring results in alerts that intuition may be wrong, then Type 2 processing can correct or override the automatic judgments initially resulting from

Type 1 processing (Kahneman and Klein, 2009; van den Bosch and Toet, 2018). In general, biases occur when deliberate processing either (1) fails to successfully engage (Kahneman, 2003) or (2) fails to override the biased heuristic response (Evans and Stanovich, 2013). In the literature, these thinking processes are sometimes labeled as “Systems”.

- ✚ This “System” terminology should be taken as a figure of speech rather than representing physical entities in the brain and is adopted to make these processes more easily comprehensible (e.g., Kahneman, 2011; Stanovich and Toplak, 2012).

Expertise-Based Frameworks

- ✚ Heuristic thinking can be quite effective (i.e., “unbiased”) in many practical or familiar situations (Gigerenzer, 2000; Gigerenzer and Gaissmaier, 2010; Goldstein and Gigerenzer, 2002).
- ✚ From a heuristic viewpoint, people don’t have to weight every conceivable strand of evidence. They only have to consider the information at hand, and then decide on the basis of previously acquired simple heuristics (Gigerenzer, 2007). This view is known under various names, like “fast and frugal” (Gigerenzer and Goldstein, 1996), “satisficing” (Simon, 1987), “take the best” (Newell and Shanks, 2003), and “blinking” (as opposed to thinking: Gladwell, 2005).
- ✚ ***According to Klein (Klein, 1997, 1998) in regular and familiar situations this fast and efficient way of decision making is not much more than recognition, i.e. the activation of domain knowledge, established and stored in memory by prior learning and experience.***
- ✚ This is termed “skilled intuition” (Klein, 1997, 1998; Simon, 1992). Klein explicated this view in his Recognition Primed Decision Making model (Klein, 1997).
- ✚ In this view, correct decisions that are made intuitively are the result of experience and expertise. Most of the time, in normal and familiar situations, we can either trust our first impulse (which is quite often right) or we can correct our initial decisions if we turn out to be wrong. Incorrect heuristics and biases result from a lack of expertise. People may then apply experience-based heuristics in unknown or unfamiliar conditions that do not match their mental model.
- ✚ This occurs for example in experiments that are performed in artificial- or laboratory contexts (Klein, 1993, 1998, 2008). The overall problem is that we have no clue where our intuitions come from. There is no subjective marker that distinguishes correct intuitions from intuitions that are produced by highly imperfect heuristics (Kahneman and Klein, 2009).
- ✚ ***A prerequisite for the development of adequate heuristics (i.e., for effective learning) is that the environment should be sufficiently consistent and predictable, providing adequate feedback and a high number and variety of learning experiences (Dane and Pratt, 2007; Klein, 1998; Shanteau, 1992). So, decision errors are made in environments that are irregular or noisy, without sufficient stable relationships between clearly identifiable cues and subsequent outcomes (of possible actions). In such environments learning and the development of expertise is difficult or even impossible (Kahneman and Klein, 2009).***

- ✚ After sufficient and adequate training, experts will usually effectively rely on heuristic processing but may switch to deliberate reasoning when they notice that they are relatively unfamiliar with a given issue (“adaptive rationality”). Checking one’s intuition is a System 2 operation, which people do not always perform, either because it is difficult and requires effort or because they simply do not bother or recognize the need.
- ✚ A related notion concerns the “smart” nonconscious, which may be superior to conscious decision making if the problem at hand is complex enough. This non-conscious system performs highly complex operations, integrating information automatically and often without any noticeable effort (Bargh and Morsella, 2008; Dijksterhuis and Nordgren, 2006). In contrast, conscious thought is hierarchical and inherently constrained by low capacity with a tendency toward stereotypes and jumping to conclusions based on plausible, accessible or verbalizable attributes

Neuro-Evolutionary Explanations

- ✚ *Neural-Network Explanations As a biological neural network of flesh and blood, necessary for survival, our brain has undergone an evolutionary process over millions of years. Over this extensive period, it has turned into a highly effective and efficient system for regulating essential biological functions and for performing perceptual-motor and pattern-recognition tasks, such as searching for food, fighting and flighting, and mating. Almost during our entire evolution, the neural networks of our brain have been optimized for these basic biological and perceptual-motor processes that lie at the basis of our daily practical skills, like cooking, gardening, or household jobs. Possibly because of the resulting proficiency for these kinds of tasks, we tend to forget that the processes involved are characterized by extremely high computational complexity (e.g., Moravec, 1988).*
- ✚ For example, when we tie our shoelaces, millions of signals flow in and out through many different sensor systems, from tendon bodies and muscle spindles in our extremities to our retina, otolithic organs and the semi-circular channels in the head (Brodal, 1981, Kahle, 1979). This enormous amount of information from many different perceptual-motor systems is continuously, parallel, effortlessly and even unconsciously processed in the neural networks of our brain (Minsky, 1988; Moravec, 1988; van de Grind, 2004).
- ✚ *To efficiently process this incoming information, the brain has a number of universal (inherent) working mechanisms, such as association and associative learning (Bar, 2007; Shatz, 1992), potentiation and facilitation (Bao et al., 1997; Katz and Miledi, 1968), saturation and lateral inhibition (Isaacson and Scanziani, 2011; Korteling et al., 2018).*
- ✚ *These basic biological and perceptual-motor capacities have already been established and optimized during our entire evolution. However, our cognitive abilities and rational functions have only recently started to develop. They are probably less than 100 thousand years old, and may, therefore, be qualified as “embryonal” on the time scale of evolution (e.g., Henshilwood³⁹ and Marean, 2003; McBrearty and Brooks, 2000; Petraglia and Korisettar, 2003).*

- ✚ In addition, evolution could not develop these new cognitive functions from scratch, but instead had to build this new, and thin layer of human achievement from its “ancient” neural heritage for essential survival functions (Moravec, 1988). So, our cognitive capacities are not only built on top of but also from and with these inherent (neuro) biological regulation mechanisms (Damasio, 1994).
- ✚ As a result, they are less optimized for cognitive functions that have become essential for “survival” in our present modern civilizations. Using this ancient biological intelligence for cognitive functions may, therefore, be compared to eating soup with a fork, or to plowing land with a racing car: although possible, it evidently does not work very well.
- ✚ From this perspective it is not surprising that the capacities of our brain for carrying out “higher” cognitive tasks still show specific tendencies and distortions. Korteling et al. (2018) pointed out how these tendencies and distortions may originate from these inherent characteristics and mechanisms of the brain as a neural network.
- ✚ Basically, these mechanisms – such as association, facilitation, adaptation, or lateral inhibition – result in a modification (e.g., weighing) of the original or available data and its processing. For instance, lateral inhibition is a universal neural process resulting in the magnification of differences in neural activity (contrast enhancement), which is very useful for perceptual-motor functions.
- ✚ However, for higher cognitive functions, that require exact calculation and proper weighing of data, this kind of data transformation may be detrimental. In this view, biases are seen as the outcome of inherent or “hard-wired” system properties (Korteling et al., 2018).
- ✚ They can be seen as cognitive deviations or distortions, resulting from the same kind of underlying neural mechanisms that also cause many types of perceptual phenomena, like Mach bands (Ratliff, 1965; Reeves and Pinna, 2017).
- ✚ ***The neural mechanisms and processes causing these perceptual illusions under specific, mostly degraded, circumstances (Gibson, 1966, 1979) have been established and proven effective during millions of years. For cognitive tasks, however, they are probably less generally valid. While these underlying mechanisms may sometimes yield fast and effective solutions in cognitive tasks (e.g., heuristics in the case of well-trained experts), they may produce solutions that are inadequate (biases) in other circumstances.***
- ✚ ***Korteling et al. (2018) proposed a neural network framework for cognitive biases based on the basic mechanisms and characteristics of neural “wetware” (Kosslyn and Koenig, 1992) that are inherent to (all) neural networks and therefore occur throughout all nervous systems. The basic unifying construct of this framework is the “biological neural network”.***
- ✚ ***This network has “association” as its most fundamental and basic “binding” principle, together with some neural network characteristics that are intrinsically connected to associative information processing.***
- ✚ This forms the basis for the following tendencies:

- 1. The tendency to associate (unrelated) information in order to find relationships and patterns. Examples are Superstition, Story bias, Representativeness bias, or Stereotyping.
- 2. The tendency to give priority to information and choices that are compatible and consistent with our present expectations, knowledge, and choices. Examples are Confirmation bias, Sunk-cost fallacy, Belief bias, Cognitive dissonance, Familiarity bias, Status quo bias, or System justification.
- 3. The tendency to retain given information that sometimes better could be ignored. Examples are Hindsight bias, Anchoring bias, Outcome bias, or Framing bias.
- 4. The tendency to focus on activated or dominant information while neglecting relevant information that is not directly available or recognized (“Blind spot”). Examples are the Availability bias, the Survivorship bias, the Ego-centric bias, the Focusing illusion, the Priority heuristic, and the Bias blind spot.

Evolutionary Explanations

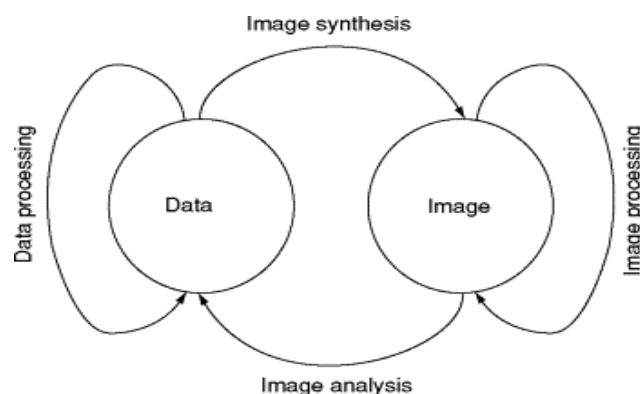
- 1. The neuroscientific perspective, based on the inherent characteristics of biological neural networks, cannot provide a plausible explanation for all biases.
- 2. For example, why should an associative neural network exhibit socio-psychological or motivational biases (e.g., Haselton et al., 2009; Haselton et al., 2005), such as Reciprocity (Cialdini, 1984; Fehr and Gächter, 2000), Loss aversion (Kahneman and Tversky, 1984), Conformity (Cialdini and Goldstein, 2004), Groupthink (Janis, 1982; Turner and Pratkanis, 1998), the numerous phenomena of Herd behavior (Raafat et al., 2009) and all kinds of (more subtle) forms of selfishness and pro-social behavior (Buss, 2005). Why are we so eager to follow (or copy) the decisions of people around us (Social proof: Cialdini, 1984) and why do we attribute greater accuracy to the opinion of authority figures and celebrities (unrelated to content)?
- 3. In this, the evolutionary view provides a unifying and consistent explanation for various (more socially oriented) biases which are difficult to explain from inherent neural-network characteristics.
- 4. ***Evolutionary psychology explains psychological phenomena and human psychological traits from evolved adaptation, i.e. the functional products of natural or sexual selection in human evolution.***
- 5. Much like physiological mechanisms, such as liver, limbs, blood circulation, and immune system, the mind comprises adaptations serving different functions that have evolved to solve recurrent problems in human primal environments (Buss, 2005; Dunbar and Barret, 2007; Durrant and Ellis, 2003; Pinker, 2002; Tooby and Cosmides, 2005; van Vugt et al., 2014). Without such a mechanism, the chances of survival for human beings would have been relatively low and the human species as we know it nowadays would not exist.
- 6. With regard to human decision making, evolutionary psychology explains biases from ingrained evolutionary principles that were of value for the survival of our ancestors, who, being hunter-gatherers, lived in small, close-knit groups (Haselton et al., 2009; Korteling et al., 2018; Tooby and Cosmides, 2005).

- ✚ During the largest part of evolution, certain tendencies and behavior patterns were beneficial for survival and reproduction. However, the psychological tendencies that optimized the chances of survival of our ancestors in their (natural) environment (“evolutionary rationality”: Haselton et al., 2009) are not always beneficial in the modern (artificial) world. In some cases, they may lead to suboptimal results under current living conditions.
- ✚ People are usually unaware of the evolutionary causes of their behavior. From this perspective cognitive biases can originate from a mismatch between evolutionarily adaptive psychological tendencies) and the current context or environment (Haselton et al., 2009; Tooby and Cosmides, 2005).
- ✚ For example, our ancestors, living in an uncertain and threatening environment, were smart to quickly flee when information suggested a potential threat: it was probably wiser to flee when seeing a silhouette that could indicate the presence of a saber- toothed tiger than to first establish the exact nature of its origin through careful observation.
- ✚ In modern civilized societies it is often more sensible, in the event of imminent threat, to first find out exactly how a situation works, instead of acting directly. Our evolutionary imprinting nevertheless still urges us to act immediately, now known as the Action bias (Baron and Ritov, 2004; Patt and Zeckhauser, 2000). We are sensitive to all kinds of peer pressure, and we are also focused on maintaining or strengthening our status and position within the social group. This can lead for example to parroting or copying the behavior of others, to faithful following of leaders, to striving for status relative to peers, to having difficulty with being indebted to others, and to showing pro-social and altruistic behavior (e.g., Buss, 2005). This way the evolutionary view provides a simple, plausible, and consistent explanation for many biases, especially in the social-psychological domain. Examples are the Social comparison bias, Authority bias, Group think, Ingroup-bias, Bandwagon effect, and Reciprocity.
- ✚ ***These cognitive biases in the socio-psychological domain are difficult to explain directly from neural basic mechanisms. Other biases that are not easily explained by the generic (inherent) characteristics of neural networks relate to evolutionary survival principles of self-interest and self-preservation. According to Hardin (1968) and van Vugt et al. (2014) the human mind is shaped to prioritize personal interests over collective interests because natural selection favors individuals who gain a personal benefit at the expense of unrelated others. The resulting prevalence of narrow self-interests over the common good of the group is described by a social dilemma, called the Tragedy of the Commons. Searching for direct individual profit, individuals tend to increase their exploitation of common resources (e.g., water and land) thereby unintentionally causing their depletion.*** Other examples of ingrained biases originating from our ancestral self-interest and self-preservation are Loss aversion, Hedonic adaptation, Hyperbolic discounting, Optimism bias and the Scarcity heuristic.

IMAGE ANALYTICS

- ✚ *Image analysis is the extraction of meaningful information from images. Precision medicine demands precision diagnostics. Image analysis (computer vision) is an ideal application of digital pathology to address this need, helping pathologists to transition from providing just traditional qualitative (descriptive) reports to more quantitative results.*
- ✚ *Specialized image analysis software platforms are commercially available (e.g., Visiopharm, Definiens, Indica Labs, Virtuoso from Roche, and Genie from Leica). Analyses have shifted from FOV (static) images to WSI and from one to multiple markers (multiplexing), often even using many colors (multispectral imaging).*
- ✚ *Image algorithms include several steps such as image preprocessing (e.g., color normalization), detection, segmentation, feature extraction, classification, and quantification. Software algorithms have been developed to identify rare events (e.g., screening for microorganisms, counting mitoses, and detecting micrometastases in lymph nodes) and quantify stains (e.g., most commonly breast biomarkers) and various features (e.g., extent of tissue fibrosis and degree of liver steatosis). They can also analyze spatial patterns and feature distribution better than humans.*

There are several benefits of image analysis. Algorithms offer better accuracy, because they provide more precise and even continuous quantitative measurements compared with humans. Using image algorithms also permits standardization due to more reproducible results, especially for intermediate scoring categories and complex scoring systems. Automated image analysis can reduce time consumption for pathologists, especially for performing mundane tasks like counting. Image analysis promises to introduce CAD, helping pathologists find, diagnose, and grade cancer. With deep learning and convoluted neural networks, some algorithms have been shown to even assist with predicting cancer diagnoses.



- ✚ *Image analysis is a computer-based process of extracting quantitative information from images. The process begins with the input of an image and ends with the output of numerical data (Figure 1). This distinguishes it from image processing where both*

input and output are in the form of an image.

- ✚ ***Image processing is the means by which the input image is modified by one or more mathematical algorithms to generate an output image that is enhanced in some way.*** For example, edges may be enhanced or the noise reduced. Image processing is often used to prepare images prior to analysis.
- ✚ ***Image analysis requires specialized computer equipment fitted with an imaging device, such as a television camera, coupled to a microscope or a macroviewer.***
- ✚ The first commercial image-analyzing computers were constructed in the early 1960s for grading steel by measuring nonmetallic inclusions.
- ✚ The value of using an automated method for the rapid, objective measurement of images was soon recognized for many different applications in life and materials sciences.

VIDEO ANALYTICS

- ✚ ***Video analytics is a technology that processes a digital video signal using a special algorithm to perform a security-related function. There are three common types of video analytics:***
 - ✚ Fixed algorithm analytics
 - ✚ Artificial intelligence learning algorithms
 - ✚ Facial recognition systems
- ✚ ***The first two of these try to achieve the same result. That is, they try to determine if an unwanted or suspicious behavior is occurring in the field of view of a video camera and the algorithm notifies the console operator of the finding. However, each takes a dramatically different route to get to its result. Fixed algorithm analytics use an algorithm that is designed to perform a specific task and look for a specific behavior.***

For example, common behaviors that fixed algorithm analytics look for including the following:

- a) Crossing a line,
- b) Moving in the wrong direction down a corridor ,
- c) Floating face down in a swimming pool

Each fixed algorithm looks for a very specific behavior. The client must pay for each individual algorithm for each individual video camera in most cases.

- ✚ ***Artificial intelligence learning algorithms operate entirely differently. Learning algorithm systems begin as a blank slate. They arrive completely dumb. After connecting to a given camera for several weeks, they begin to issue alerts and alarms.***
- ✚ During that time period the system is learning what is normal for that camera's image during the day, night, weekday, weekend, and hour by hour. After several weeks, the system begins to issue alerts and alarms on behavior in the screen that it has not seen before or that is not consistent with what it has seen during that time period for that day of week.

An example illustrates the usefulness of this ⁴⁴approach. In one early installation at a major international airport that was intended to spot children climbing on a baggage carousel, the system alerted on a man who picked up a small bag from the carousel and placed it inside an

empty larger bag. The man was intercepted and interrogated only to discover that the luggage inside his did not belong to him and that he was part of a ring who came to the airport regularly to steal baggage in this way. The airport had no idea this was even occurring, so there was no way they could have purchased a fixed behavior algorithm for this, even if such existed (which it did not). This approach to video analytics is most useful.

✚ *The third type of analytic is facial recognition. Facial recognition systems can be used for access control or to help identify friend or foe. Facial recognition systems can also be used to further an investigation.*

✚ *Typical facial recognition systems match points on a face with a sample stored in a database. If the face does not match a record, it will try to create a new record from the best image of that person available. These are capable of making real-time matches of one image against many. The latest version of facial recognition systems constructs 3-D maps of faces in real time and compares those to a truly vast database. One manufacturer claims to be able to match individuals in real time against a country-sized database of images (millions of records).*

Traditional facial recognition systems require well-lit scenes and fairly static backgrounds. The latest versions are reported to work under fair to poor lighting and with dynamic backgrounds.

AUDIO ANALYTICS

Audio analytics is about analyzing and understanding audio signals captured by digital devices, with numerous applications in enterprise, healthcare, productivity, and smart cities.

Applications include customer satisfaction analysis from customer support calls, media content analysis and retrieval, medical diagnostic aids and patient monitoring, assistive technologies for people with hearing impairments, and audio analysis for public safety.



✚ Extracting non-verbal cues from human speech. This refers to analyzing a human voice to extract information beyond speech recognition, including speaker identification and verification: age, gender, and emotional state.

- ✚ *Audio understanding. This aims to analyze and extract insights from audio signals such as detecting audio events, recognizing audio backgrounds, and detecting audio anomalies.*
- ✚ Audio search. This focus area refers to audio data search mechanisms, essential for navigating through large amounts of raw audio data and metadata. Audio search includes description and annotation of audio data, querying and indexing, and ranking and retrieval.
- ✚ *A typical audio processing process involves the extraction of acoustics features relevant to the task at hand, followed by decision-making schemes that involve detection, classification, and knowledge fusion. We use various approaches ranging from Gaussian mixture model-universal background model (GMM-UBM) and support vector machine (SVM) to the latest deep learning neural network architectures.*

Emotion recognition is one of the first areas of audio analytics that we started to explore. We designed a series of neural network architectures and worked with both public (IEMOCAP, eINTERFACE, Berlin) and private (Xbox, Cortana, XiaoIce) datasets. Labeling typically happens by using crowdsourcing, which gives us an understanding of how well humans perform on the task. Speech emotion recognition is a challenging problem, as different people express their emotions in different ways—even human annotators sometimes cannot agree on the exact emotion labels.

- ✚ *A multimodal approach, based on both audio and text (the output of a speech recognizer), can provide significant improvement in model performance. When it comes to other non-verbal cues extraction tasks, such as age and gender detection from speech, these are easier tasks and here our neural networks perform as well as human labelers.*
- ✚ *Audio events detection was an early project in our audio understanding research. Using pre-trained CNN models as feature extractors, we enable knowledge-transfer from other data domains, which can significantly enrich the in-domain feature representation and separability. We combined knowledge-transfer with fully connected neural networks and achieved classification results close to human labelers, when testing on a noisy dataset with more than ten types of audio classes.*
- ✚ Sound event detection results. Audio search is the most complex and difficult of all three research directions. We have built a prototype of an audio search engine that uses joint text-audio embeddings for query indexing and personalized content search, and it shows promising results when compared to equivalent search approaches that rely on text information alone. We also proposed a first deep hashing technique for efficient audio retrieval, demonstrating that a low-bit coding representation combined with very few training samples (~10 samples per class) can achieve high mAP@R values for audio retrieval.

Audio Retrieval Accuracy

	Audio-only Index (MFCC)	Text-only Index (Word2Vec)	Audio-Text Joint Index (CNN)
Search with Audio	49.7	2.4	50.3
Search with Text	2.4	85.5	86



Memory: Cognitive Engagement: BOTs, Virtual & Digital Assistants, Augmented Reality, Virtual Reality, Mixed Reality

COGNITIVE ENGAGEMENT

- ✚ Cognitive engagement involves the psychological investment of the student in the learning process.
- ✚ *Cognitive engagement is defined as the extent to which students' are willing and able to take on the learning task at hand. This includes the amount of effort students are willing to invest in working on the task (Corno and Mandinach 1983), and how long they persist (Richardson and Newby 2006; Walker et al. 2006).*
- ✚ *Cognitive engagement has traditionally been operationalized by measuring the extent of students' homework completion, class attendance, extra-curricular participation in activities, or their general interactions with the teachers, and how motivated they seem while engaging in classroom discussions (Appleton et al. 2006).*
- ✚ We suggest that cognitive engagement is more or less dependent on the task at hand because the task determines the extent of students' autonomy. For instance, working with groups and engaging in discussions, searching for information on the internet, or listening to a lecture is likely to result in different levels of cognitive engagement because of different levels of autonomy.
- ✚ Listening to a lecture is arguably the least cognitively engaging since under such circumstances there is little to no student autonomy. On the other hand, when students independently search for information on the internet—that is, when students engage in self-initiated information-seeking behaviors—the level of autonomy should be relatively high and thus lead to more cognitive engagement.
- ✚ Working in groups and engaging in discussions could result in either high or low feelings of autonomy, depending on the group dynamics.

- ✚ *The new situational cognitive engagement measure is composed of three overlapping facets:*
 - ✚ (1) *how students perceive their present engagement with the task,*
 - ✚ (2) *how they rate their effort and persistence while working on the task, and*
 - ✚ (3) *how much they feel absorbed by the learning task, for instance, whether it makes them forget everything around them. It is important to note that all three facets measure ongoing cognitive engagement and try to capture the activity of being engaged.* This is conceptually different from existing measures of cognitive engagement, which are typically administrated at the end of the course or a semester and require students to make summative judgments of how engaged they generally were during a particular course spanning several weeks.

Cognitive engagement in the problem-based learning classroom

✚ ***The objective of the present study was to examine to what extent autonomy in problem-based learning (PBL) results in cognitive engagement with the topic at hand. To that end, a short self-report instrument was devised and validated. Moreover, it was examined how cognitive engagement develops as a function of the learning process and the extent to which cognitive engagement determines subsequent levels of cognitive engagement during a one-day PBL event. Data were analyzed by means of confirmatory factor analysis, repeated measures ANOVA, and path analysis. The results showed that the new measure of situational cognitive engagement is valid and reliable. Furthermore, the results revealed that students' cognitive engagement significantly increased as a function of the learning event. Implications of these findings for PBL are discussed***

✚ To test the extent to which this is the case we first devised and validated a short self-report instrument to measure students' situational cognitive engagement in the classroom. Subsequently, we examined how situational cognitive engagement develops as a function of the learning process in PBL and the extent to which situational cognitive engagement during the learning process determines subsequent levels of cognitive engagement. Data were collected from applied-science courses at a polytechnic in Singapore.

✚ The results of the construct validation and cross-validation study suggest that the four-item instrument is a reliable and valid measure to determine students' situational cognitive engagement in the classroom. As such, we used it for the subsequent analyses.

✚ ***Following Self-determination theory (Deci and Ryan 2004) under which autonomy is defined as the degree to which individuals feel volitional and responsible for their initiation of their behavior (Williams 2004), we hypothesized that students would have the highest feelings of autonomy (and thus would engage more) during self-study because during self-study they are expected to feel most volitional in their actions and are most responsible for their***

✚ The results of our analyses did however not support this hypothesis. Students' cognitive engagement did not progress in a wave-like pattern, but it increased significantly and consistently over the day. Strongest evidence against our hypothesis is that students' situational cognitive engagement increased significantly not during the first self-study phase, but when students met with the group thereafter to discuss their findings. Situational cognitive engagement increased significantly again during the second self-study phase. Our data suggest that students' situational cognitive engagement is not influenced by changes in task demand and associated feelings of

autonomy, but situational cognitive engagement is more a function of the learning event itself: if students progress with their learning in PBL, their situational cognitive engagement increases.

- ✚ Considering this outcome, we offer an alternative hypothesis. We propose that students' feelings of autonomy and situational cognitive engagement *are a direct function of a students' knowledge construction*.
- ✚ During the early stages of the problem day (i.e. during the problem-definition phase), students struggle to come up with adequate theories to explain the phenomena described in the problem. Struggling to explain the problem is expected (and intended) because students lack relevant knowledge as they are supposed to engage in theory construction.
- ✚ At this stage students largely depend on the elaborations with the other team members and the questioning of the tutor. As such, choice and autonomy are expected to be generally low. However, as students gain a deeper understanding of the topic, they gradually depend less on the support of their peers and the tutor, because they have gained more knowledge to direct their own learning.
- ✚ With increasing knowledge, the knowledge of possible (learning) choices also increases, which translates into a feeling of autonomy and consequently higher levels of situational cognitive engagement.

This finding opens up new areas of research for self-determination theory. Reeve (2004) has proposed that there is empirical evidence to support two conclusions about self-determination theory and its significance for education: ***(1) autonomously-motivated students thrive in educational settings; and (2) students benefit when teachers support their autonomy. Indeed, there is considerable evidence linking these two factors to positive educational outcome, such as higher academic achievement higher rates of retention , higher perceived competence , greater conceptual understanding greater creativity , and higher self- esteem. However, largely missing from the current research agenda is the consideration of the significant role knowledge may play in autonomy and autonomy-supportive behavior.***

Needless to say, further research needs to be carried out to empirically test whether and how autonomy and knowledge (development) are interlinked. We suggest a fruitful approach would be to include, besides the situational cognitive engagement measure, a measure of students' autonomy and their factual knowledge in the investigation to examine how these three factors are related and how they develop during the stages of student learning in PBL.

ADVANCES IN ARTIFICIAL INTELLIGENCE, AUGMENTED REALITY, AND VIRTUAL REALITY MEAN FRESH AR/VR DESIGN OPPORTUNITIES FOR DESIGNERS.

The digital and technological landscape is constantly changing, and in many ways accelerating. Designers tasked to come up with innovative ideas have to keep track of what is trending and where the creative opportunities are. A New Technology Landscape: Human- assisted AI, Augmented Reality, Virtual Reality, and Mixed Reality

What Is Artificial Intelligence?

- ✚ *Artificial intelligence (AI) is giving a computer or computer-controlled device the ability to perform intelligent tasks such as reasoning or learning from experience. Cognitive computing was the term used when it all started a decade ago—the idea that, facing cognitive overload, we are going to need assistance making decisions. Citing obvious advantages of current artificial intelligence in use, we are already relying on AI digital assistants like Siri and Alexa (among others) to help us make simple decisions.*
- ✚ Product designers will have an increasing volume of data information to inform their design decisions. Artificial intelligence experts leverage AI technology to create dynamic user interfaces. These technologies personalize each user encounter by examining every interaction within an app or a website, integrating elements of AR/VR design and analytics for a more immersive experience.

VIRTUAL ASSISTANTS – AI TECHNOLOGY IN ACTION

- ✚ Virtual assistants like Siri, Alexa, Cortana, Watson, and others rely on artificial intelligence to help do everything from finding a restaurant to sending an email or scheduling a meeting. Many of these are operated by voice, a more natural interface for human than a traditional visual UI of windows and buttons. These technologies are likely to reshape the current design of search engines.
- ✚ Speaking is a natural and straightforward method of interaction for humans. Thus, it's no surprise that voice-controlled virtual assistants have received a favorable adoption rate in our homes. Users can simply talk to the system rather than “operate” it with hands or gestures. For example, with an iPhone, using only one quick request, Siri allows us to transfer money via PayPal to another person
- ✚ As AI is integrated into apps, it will help intelligent virtual assistants (IVAs) become more and more intelligent. Nevertheless, regardless of how smart IVAs get, users with poor eyesight or other physical disabilities—who use screen readers and other assistive technologies—could still face significant challenges. It will be up to UX designers working with user experience researchers to bridge such critical accessibility gaps.

How Machine Intelligence Is Powering a Driverless World

- ✚ *The transit systems of the future will have traffic flow through our streets smoothly and efficiently. Design innovations like modular, detachable buses, flying taxis, and networks of suspended magnetic pods will make the dream of a dynamic, driverless world into a reality.*
- ✚ *What was science fiction news some years ago is now becoming a reality. Wheelys' Moby-Mart is a self-driving store with a holographic shop assistant that is likely to change the world of eCommerce design.*
- ✚ The next generation of eCommerce designers will need multi-specialization in both physical and digital spaces. Future AR/VR designers will need to design an experience where many of the familiar human elements are removed or remote. Amazon Go, a cashier-less store, is a great example of what designing the future looks and feels like. They designed a shopping experience where anyone can enter the store, browse, shop, and leave with their products in hand, friction-free. No lines at the checkout.
- ✚ *The intelligence built within each autonomous vehicle allows it to make all the right decisions for us—everything from the most optimal routing to human safety concerns. The creation of such complex design systems is creating a product template that will start a new wave of products designed outside the transportation vertical.*
- ✚ It will become necessary for automotive interior designers to create designs making interior space comfortable and adaptive. The interfaces will need to be intuitive and reassuring in order to encourage trust from the passenger while focusing on human interactions between people sharing the ride.

AUGMENTED REALITY

Augmented reality (AR) acts as a virtual layer on top of the world in front of you, and is in a new age of discovery and innovation. While early AR applications depended on a smartphone or tablet, they will soon expand into wearable devices like smart glasses requiring a different kind of UX and design process—thinking outside the confines of a smaller screen. While the debate over AR vs. VR vs. MR is ongoing, augmented reality graphic design is a unique opportunity for designers to create apps using image, object, and color recognition. Just like AI leverages data to make decisions, AR leverages the recognition of key “anchor” objects and points within a given space.

- ✚ A great example of a perfect combination between AI and AR is with Apple's Animoji, animated emoji for the iPhone X. According to Apple, Animoji is “custom animated messages that use your voice and reflect your facial expressions.” The rendering and animation work in real time, thanks to iPhone X face-scanning features.

Meta – Building Immersive AR Experiences

- ✚ In the next five years, AR will be better integrated with our everyday lives, merging our physical and digital worlds with new wearable products. A product like Meta glasses, which overlays augmented reality on top of the user's reality, is likely to be popular as a medium for game designers.
- ✚ Based on holographic technology, Meta glasses can identify users' gestures to let them manipulate 3D projections of objects. With so many devices based on gestures, standardization and cultural gaps could become a main challenge for designers working in the field. "The future of productivity is spatial," said Meta when introducing their AR Workspace. In the future, wearable hardware will enable employees to work hands-free.
- ✚ The devices of the future could work on voice commands or gestures. The increasing success of AR experiences may predict the end of physical controllers such as a mouse cursor or a keyboard. In augmented reality, natural gestures could be the key drivers of the experience.

AR-enhanced Living

As part of its concept kitchen, IKEA is developing a smart table that suggests recipes based on the ingredients placed on it. This is a great example of AR technology working in the real world. The interactive table has a camera-equipped projector that shows recipes on its surface and recognizes ingredients, giving you an idea of what to make with what you have on hand. Such immersive experiences are in direct competition with mobile apps that will soon seem outdated due to their more restricted capabilities. Mobile app designers will soon be creating augmented applications that integrate with physical spaces to deliver a new and exciting way of interacting with our physical environments.

The Future of Work with AR Augmented reality can also be used for knowledge transfer, such as training in industrial environments. Instead of explaining, imagine showing a worker how to do a specific task via in-context video. This opens doors for designers in the e-learning area to create next-generation AR interfaces.

VIRTUAL REALITY APPLICATIONS.

- ✚ The novelty of virtual reality content offers a unique opportunity to engage large audiences as demonstrated by a VR documentary from National Geographic: The Protectors, Walk in the Ranger's Shoes. It received more attention compared to other nature themed documentaries. Beyond transporting viewers into new places, the next opportunity for designers is helping them experience life as another person. Offering a

first-person view can create unique and effective campaigns like “A Walk Through Dementia,” centered on empathy.

- ✚ A new set of design patterns will ultimately emerge for VR designers, but a flat design approach is unlikely to succeed. In VR, our brain uses spatiotemporal orientation and problem-solving capabilities; therefore, it’s essential to use textures, lighting, and finer details for a more believable and immersive experience.
- ✚ VR gives users the ability to get up close and personal with virtual objects and also move back from them. Since the eye is very good at picking out depth information, rendering of objects by designers needs to be realistic and detailed. Sound and music are other important considerations for VR designers.

MR – MIXING REAL AND VIRTUAL REALITIES

Microsoft HoloLens mixed reality blends 3D holographic content with the physical world, giving holograms real-world context and scale. MR technology mixes and blends a person’s physical surroundings with objects from the physical and digital world together. Using this MR technology, we can interact with both digital content and the world around us as well as interact with 3D holographic projections.

MR and VR applications give new creative freedom to designers. We can create experiences where users can interact with products virtually and visit distant environments, historical events, concerts, and much more. Within the virtual or mixed realities that we create, we can bend the laws of physics with endless possibilities and experience the world like never before. From a design and usability standpoint, a key principle to keep in mind is to put the users at ease and avoid discomforting sensations by keeping them grounded in the real world. To that end, we can leverage sound, motion, and a good sense of scale—the virtual environments and interactions we create should feel natural and ergonomic, not necessarily realistic. When designing a visual interface that uses a headset and needs a close examination of an object like a VR organ, we must keep in mind that people are much more susceptible to motion sickness when wearing a VR headset than when looking at a computer screen.

An effective use of virtual reality design and marketing allows brands to leverage the following design components:

- 1. A well thought-out creative strategy*
- 2. Application of usability best practices*
- 3. Strong narrative and storytelling*
- 4. Ergonomic information presentation*
- 5. Compelling visual design and simulations*

Learning: Intelligent Automation, Reinforcement Learning, Intelligent Agents

TECHNOLOGIES FOR INTELLIGENT AUTOMATION

- ✚ Research in AI has been undertaken since the 1970s with early developments in decision support systems (DSS) and expert systems (ES) (El-Najdawi and Stylianou, 1993). However, in recent years, game-changing progress has been made in addressing some of the fundamental challenges of the AI discipline.
- ✚ Advances have been made in Natural Language Processing, Machine Learning, and Computer Vision (Brynjolfsson and McAfee, 2016). The rapid growth in the availability and accessibility of big data combined with vast computing power, readily available through the cloud, have aided these developments (Davenport and Kirby, 2016b). These recent advances in AI are creating a new generation of systems that are distinct from the early DSS and knowledge-based systems in three respects.
 - First, the old systems could not automatically learn and improve their methods and results and were reliant on human programmers to make adjustments.
 - Second, the old systems functioned as assistants or advisors to human professionals providing recommendations or advice, but they required a human worker to apply the decision.
 - Third, while these systems were designed to help managers with repetitive decisions and complex unstructured problems, they were not designed to remove cognitive tasks from the workload of the human.

Thus, although the recent advances in AI may be considered a further evolution and extension of the original AI field, their widespread adoption in organisations presents a fundamentally different landscape to what came before. Table 1 summarises these early AI applications and their limitations.

Table 1. Early applications of AI and associated limitations.

Application	Description	Limitations
<i>Decision Support System (DSS)</i>	<i>A set of tools utilising models and or analytic techniques to assist managers in their decision-making (El-Najdawi and Stylianou, 1993).</i>	<i>It contains the knowledge of human experts in solving particular problems but is not able to automatically enhance that knowledge based on the experience of results without human intervention.</i>
	55	<i>System designed to support, and not replace, humans (Mallach, 2000).</i>

Expert (ES)	Systems	<i>Systems designed to capture the knowledge of human experts in a narrow problem domain, and help solve</i>	<i>Aimed to achieve the flexibility to remove experts from repetitive decisions but not to replace human counterparts (El-Najdawi and</i>
--------------------	----------------	--------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------

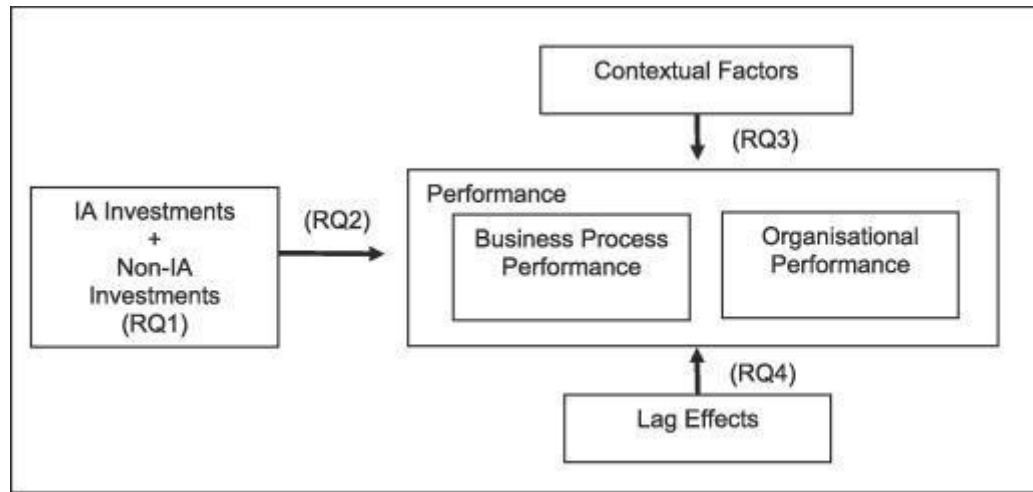
Application	Description	Limitations
	<i>problems (El-Najdawi and Stylianou, 1993, Ye and Johnson, Stylianou, 1993).</i>	<i>(1995).</i>
Knowledge Management Systems (KMS)	<i>Support the creation, transfer, and application of knowledge in an organisation (Alavi and Leidner, 2001).</i>	<i>Provides professionals with a support tool to find organisational knowledge to solve business problems or locate the relevant internal expertise, but do not implement decisions (Alavi and Leidner, 2001).</i>
Recommendation Agent (RA)	<i>Software agents that capture the preferences of customers and make recommendations based on these preferences (Xiao and Benbasat, 2007).</i>	<i>A human is still required to review the recommendations presented by an RA and decide which recommendation to apply (Alavi and Leidner, 2001, Xiao and Benbasat, 2007).</i>

- ✚ Advances in ML are enabling the development of algorithms that allow cognitive tasks found in knowledge and service work, to be automated (Frey and Osborne, 2017). Frey and Osborne (2017) emphasise the importance of advances in ML that develop algorithms that mimic human cognitive functioning. Therefore, ML advances are critical, as once progress has been made regarding machines' ability to build on their knowledge for decision making, real-world applications of automated decision making are likely to follow (Davenport and Kirby, 2016a).
- ✚ For example, ML algorithms are enabling the automation of cognitive tasks such as medical imaging analysis (Lee et al., 2017), or auditing tasks such as identifying accounting anomalies in unusually high sales figures (Kokina and Davenport, 2017).
- ✚ When reviewing our sample literature, we used four research questions to structure our review framework:
 - ✚ 1) What Intelligent Automation investments and non-Intelligent Automation investments have been studied?
 - ✚ 2) How have Intelligent Automation investments influenced business process performance or organisational performance?
 - ✚ 3) How have contextual factors influenced Intelligent Automation enabled business

process performance or organisational performance?

- ✚ 4) How have lag effects influenced Intelligent Automation enabled business process performance or organisational performance?

Fig. 1 presents the research framework we adopted with the associated research questions.



Note: IA = Intelligent Automation

Fig. 1. Framework for Intelligent Automation Literature Analysis (based on Schryen, 2013).

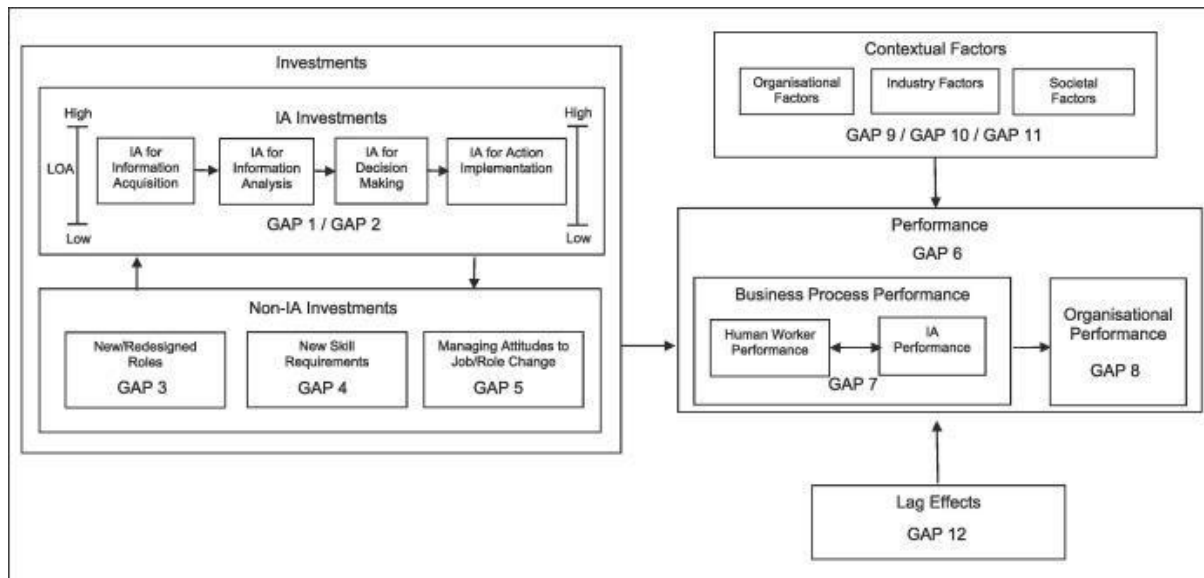
We followed the three-stage approach for undertaking the literature review advocated by Webster and Watson (2002).

- ✚ First, we selected four bibliographical databases (Scopus; Business Source Complete, PsycINFO and Web of Science) that between them encompass the academic literature in the fields of science, technology, medicine, social sciences, and arts and humanities, all disciplines of business, and interdisciplinary research in the behavioural and social sciences. These databases also index the AIS basket of eight journals.
- ✚ We developed two categories of search terms for interrogating the databases.
- ✚ *The first category concerned terms related to Intelligent Automation and associated technologies such as “automation”, “artificial intelligence”, “machine learning”, “cognitive computing”, “smart machines”, “mobile robots” and “robot”.*
- ✚ *The second category included search terms that were related to the effects or impacts of these technologies and included “innovation”, “business value”, “productivity”, “employment”, “ethics”, “social impact” as well as “knowledge work” and “service work”. Within the four databases, search combinations were performed where each search combined a technology search term with an impact term using truncation and wildcards.*

What Intelligent Automation investments and Non-Intelligent Automation investments have been studied?

- ✚ *Research on Intelligent Automation for knowledge and service work tends to present new artefacts or technological solutions that are designed to improve on human limitations in information processing for completing cognitive or manual tasks.*
- ✚ *Information processing comprises four stages (a) information acquisition, (b) information analysis, (c) decision making, and (d) action (Parasuraman et al., 2000).*

Two interconnected streams of research have driven these improvements: advances in AI and developments in mobile robotics.



Notes: IA = Intelligent Automation; LOA = Level of automation

Research questions, gaps, and agenda for intelligent automation research.

<i>Research questions</i>	<i>Gaps in literature</i>	<i>Future research agenda</i>
<i>What Intelligent Automation and non-Intelligent Automation investments have been studied?</i>	<p><i>GAP 1: Range and type of tasks targeted Intelligent Automation</i></p> <p><i>GAP 2: Level of Intelligent Automation implemented</i></p> <p><i>GAP 3: Impact of Intelligent Automation on jobs/work</i></p>	<p><i>How do organisations determine which tasks are appropriate for Intelligent Automation, and what factors shape this decision-making process?</i></p> <p><i>How can socially acceptable values be designed into Intelligent Automation to enable ethical decision-making, and how should this be tested?</i></p> <p><i>How do organisations determine an appropriate level of Intelligent Automation for tasks, what factors shape this decision-making process, and how does micro-level variation of Intelligent Automation within tasks influence task performance?</i></p> <p><i>How do new configurations of human-Intelligent Automation interactions emerge, and what are their impacts on working and organising?</i></p> <p><i>What are the positive/negative consequences of these new ways of working in organisations?</i></p> <p><i>How can organisations ensure Intelligent Automation is an enabler of meaningful work?</i></p>

GAP 4:
*Impact
of
Intelligent
Automation on
worker's skills*

*How are new forms of expertise,
skill requirements, and training
emerging to meet the demands of
using Intelligent Automation?*

*Does the use of Intelligent
Automation cause degradation or
enhancement of human worker
skills over time?*

GAP 5:
*Workers
attitudes
and
behaviours
in response to
the
implementation
of Intelligent
Automation*

*How do new configurations of
human-Intelligent Automation
interactions influence attitudes
and actions towards Intelligent
Automation?*

*What are the psychological,
emotional, and social aspects of
human-Intelligent Automation
collaboration?*

*What are the design implications
for Intelligent Automation
technologies for effective human-
Intelligent Automation*

<i>Research questions</i>	<i>Gaps in literature</i>	<i>Future research agenda</i>
		<i>collaboration in the workplace?</i>
<i>How have Intelligent Automation investments influenced business process performance or organisational performance?</i>	<p><i>GAP 6: Linkage of Intelligent Automation to organisational strategy</i></p> <p><i>GAP 7: Impact of Intelligent Automation on worker performance</i></p> <p><i>GAP 8: Impact of Intelligent Automation on organisational performance</i></p>	<p><i>How do organisations make strategic decisions regarding Intelligent Automation, why do they make these decisions, and what are the consequences of these decisions over time?</i></p> <p><i>How and why do new configurations of human-Intelligent Automation interactions influence human worker performance?</i></p> <p><i>How and why do investments in Intelligent Automation impact on organisational performance?</i></p>
<i>How have contextual factors influenced Intelligent Automation-enabled business process performance or organisational performance?</i>	<i>GAP 9: Influence of contextual factors on Intelligent Automation decision-making</i>	<p><i>How do organisations in different contexts devise policies and procedures for combining human judgement with Intelligent Automation decision-making?</i></p> <p><i>How do organisations decide where the balance of control lies in situations of human-Intelligent Automation conflict, and how do contextual factors influence this decision?</i></p> <p><i>How do organisations decide whether the human should be retained in the decision-making loop, and how do contextual factors influence this decision?</i></p>

<i>Research questions</i>	<i>Gaps in literature</i>	<i>Future research agenda</i>
	<i>GAP 10: Role of contextual factors on design and implementation of Intelligent Automation</i>	<i>How do contextual factors such as organisation size, industry type, or regulatory context shape the way Intelligent Automation technologies are developed and implemented?</i>
	<i>GAP 11: Stakeholders responsible for consequences of Intelligent Automation use</i>	<i>How do organisations decide who is responsible for the consequences of Intelligent Automation investments, and why?</i> <i>What constitutes a responsible Intelligent Automation system?</i> <i>How can an organisation identify bias in emerging decision-making capabilities of Intelligent Automation?</i>
<i>How have lag effects influenced Intelligent Automation-enabled business process performance or organisational performance?</i>	<i>GAP 12: Length of time between investment in Intelligent Automation, and positive return on investment</i>	<i>What is the return on investment from investments in Intelligent Automation, how should such returns be measured, and over what timescale should such investments be evaluated?</i>

Spectrum of AI

- a) **Reactive Machine:** Low memory, works on Known rules, such as Object Detection/Games/Recommendations specific to known Rules
- b) **Limited Memory:** Memory used to learn and improve continuously such as Most ML Models, Automated Vehicles
- c) **Theory of Mind:** Machine Understands and responds such as BoTs/Virtual/Digital Assistants
- d) **Self-Aware:** Human like intelligence such as Super Robots in Space etc.

a) **Reactive Machine:** Low memory, works on Known rules, such as Object Detection/Games/Recommendations specific to known Rules

+

b) **Limited Memory:** Memory used to learn and improve continuously such as Most ML Models,

JUVENILE ABSENCE EPILEPSY

- ✚ *Juvenile absence epilepsy is very similar to CAE except that the age at onset is in the second decade, with a peak between ages 10 and 12 (Wolf and Inoue, 2005). The absence seizures are not as frequent as in CAE.*
- ✚ In addition, the majority of patients also have generalized tonic-clonic seizures. This condition has a greater tendency for persistence of seizures into adulthood than is the case with CAE (CHILDHOOD ANSENSE EPILEPSY)

Epilepsies 1599

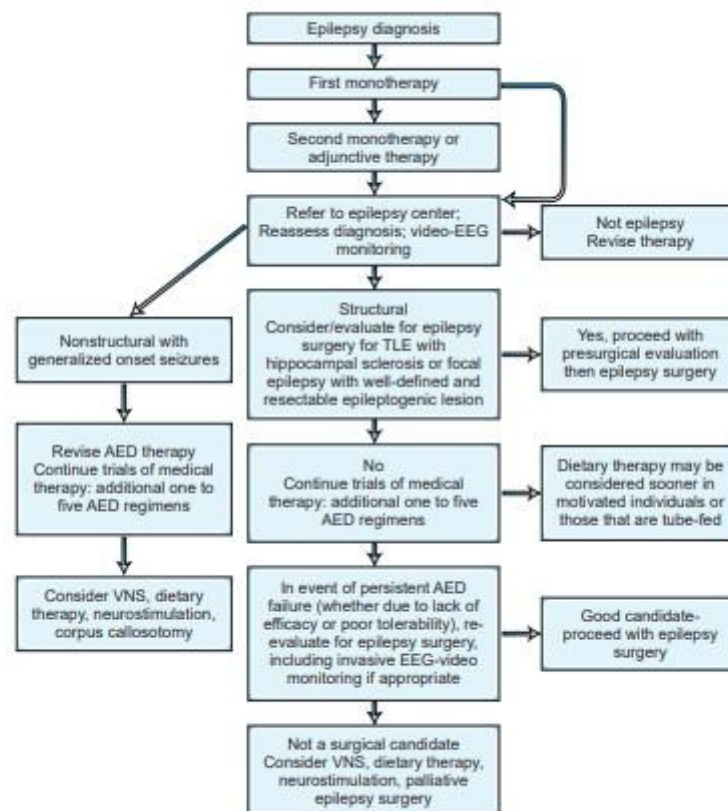


Fig. 101.17 Treatment algorithm for epilepsy. Additional steps assume that seizures are not controlled despite adequate trial of well-tolerated medication.

- ✚ Juvenile Absence Epilepsy (Table 4) Juvenile Absence Epilepsy (JAE) is characterized by absence seizures that typically occur less than daily in the untreated state and are associated with >3 Hz (range 3-5.5 Hz) generalized spike-wave in an otherwise normal adolescent. Generalized tonic-clonic seizures are seen in more than

90% of cases, most commonly beginning shortly after onset of absence seizures. Neurological examination is normal.

- ✚ Development and cognition are typically normal although ADHD and learning difficulties may occur. While seizures may be controlled with antiseizure medications, lifelong treatment is typically required. Epidemiology:
- ✚ JAE is less common than CAE, accounting for 2.4-3.1% of new-onset epilepsy in children and adolescents^{7, 8}. Clinical context: Typical age at onset is between 9-13 years, with a range of 8-20 years. Exceptional cases may present in adult life^{10, 41}. In cases with onset below 9 years of age, the distinction between JAE and CAE can be difficult (Table 1). Distinguishing features include the older age at onset and lower frequency of absence seizures in JAE. EEG features are similar however OIRDA is not seen and generalized discharges may be of slightly higher frequency and more irregular in JAE.
- ✚ ***Development and cognition prior to presentation are typically normal. A history of febrile seizures is seen in between 6-33% of cases^{19, 49}. Significant cognitive impairment should suggest an alternate diagnosis.*** Natural History JAE is often drug responsive but lifelong in the majority of cases^{41, 50, 51}.
- ✚ Ethosuximide as monotherapy is not recommended due to the high likelihood of generalized tonic-clonic seizures. Broad spectrum ASMs for generalized epilepsies should be used. Persons with JAE have higher rates of ADHD and learning problems, even if seizures are well controlled^{52, 53}.
- ✚ ***Seizure Types Absence seizures are mandatory. They have abrupt onset of impaired awareness, staring with loss of facial expression, interruption of activity, with/without oral automatisms, and immediate return to normal activity (Figure 3). Loss of awareness is often less complete than in childhood absence epilepsy⁵⁴. During absence seizures with incomplete loss of awareness, the person may be able to respond to commands but has difficulty doing complex tasks.***
- ✚ ***Typical duration is 5-30 seconds, with occasional longer seizures. Frequency is typically less than daily^{41, 54}. Subtle myoclonus may be seen during an absence seizure. Absence status epilepticus occurs in approximately 20% of patients⁵⁵. Generalized tonic-clonic seizures occur in more than 90% of cases⁴¹. They usually begin after onset of absences, but in 14-27% of cases, may precede absences^{41, 56}. The frequency of generalized tonic-clonic seizures is variable.***
- ✚ ***Myoclonic seizures are exclusionary, with the exception of subtle myoclonus occurring during an absence seizure. Independent myoclonic jerks, particularly in the morning or with sleep deprivation, should suggest Juvenile Myoclonic Epilepsy. Prominent myoclonus during an absence seizure would suggest Epilepsy with Myoclonic Absences. Prominent eyelid myoclonia during absence should suggest Epilepsy with Eyelid Myoclonia. Other seizure types are not expected in JAE. Staring spells lasting >30 seconds or with postictal impairment should suggest focal impaired awareness seizures.***

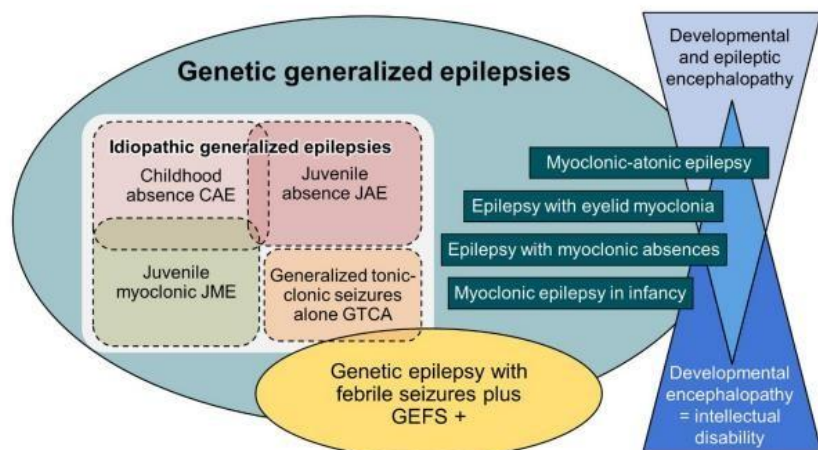


Table 1: Features seen in Childhood and Juvenile Absence Epilepsy

	Childhood Absence Epilepsy (CAE)	Juvenile Absence Epilepsy (JAE)
Age at onset		
-Usual	4-10 years	9-13 years
-Range	2-13 (caution if diagnosing <4 years of age)	8-20 years – exceptional cases may present in adulthood
Development	Typically normal, but may have learning difficulties or ADHD	Typically normal, but may have learning difficulties or ADHD
Absences		
-Frequency	At least daily to multiple per day but may be under-recognized by family	Less than daily
-Duration	Typical duration 3-20 seconds	Typical duration 5-30 seconds
-Impaired awareness	Severe loss of awareness	Less complete impairment of awareness
Other seizure types		
-Febrile	Occasional	Occasional
-Generalized tonic clonic seizure	Rarely precede or occur during period of frequent absences but may occur later with evolution to other IGE syndrome	May precede and commonly occur during the period of frequent absences
-Myoclonic	Prominent myoclonus exclusionary	Prominent myoclonus exclusionary
EEG Background	OIRDA in 21%	Normal

Epileptiform discharge -Awake -Asleep -Irregular generalized spike-wave	2.5-4 Hz generalized spike-wave Polyspike and wave may be seen in drowsiness and sleep only Uncommon	3-5.5 Hz generalized spike-wave Polyspike and wave may be seen in drowsiness and sleep only More common than CAE Discharges are more frequent than in CAE
Photoparoxysmal response	Rare IPS triggers generalized spike-wave in 15% but does not induce seizures	Rare IPS triggers generalized spike-wave in 25% but does not induce seizures
Hyperventilation induction	87%	87%

Ictal EEG	Regular 2.5-4 Hz generalized spike-wave If no generalized spike-wave is seen with hyperventilation x 3 minutes in an untreated patient, CAE can be excluded Disorganized discharges less frequent	Regular 3-5.5 Hz generalized spike-wave If no generalized spike-wave is seen with hyperventilation x 3 minutes in an untreated patient, JAE can be excluded Disorganized discharges 8x more frequent than CAE
-----------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Disorganized discharges are defined as either brief (<1 second) and transient interruptions in ictal rhythm or waveforms of different frequency or morphology during the ictal rhythm

ADHD: Attention deficit hyperactivity disorder

OIRDA: Occipital intermittent rhythmic delta activity

IPS: Intermittent photic stimulation

Table 3: Diagnostic Criteria for Childhood Absence Epilepsy

	Mandatory	Alerts	Exclusionary
Seizures	Typical absence seizures	<p>GTCS prior to or during the period of frequent absence seizures</p> <p>Staring spells with typical duration >30 seconds or with postictal confusion or fatigue</p> <p>Absences occurring <daily in an untreated patient</p>	<p>Any of the following seizure types:</p> <ul style="list-style-type: none"> • Prominent myoclonic seizures • Prominent eyelid myoclonia • Myoclonic-absence seizures • Atonic seizures • Tonic Seizures • Atypical absence seizures • Focal impaired awareness seizures
EEG	Paroxysms of 2.5-4 Hz GSW (may have been obtained historically)	<p>Consistently unilateral focal spikes</p> <p>Lack of HV activated 2.5-4 Hz GSW in untreated patient who performs HV well for 3 minutes or longer</p> <p>Recording a typical staring spell without EEG correlate in a child with a history of 2.5-4 Hz generalized spike-wave</p> <p>Persistent slowing of the EEG background in the absence of sedating medication</p>	Diffuse background slowing
Age at onset		2-3 or 11-13 years at onset	<2 or >13 years
Development at onset		Mild intellectual disability	Moderate to profound intellectual disability
Neurological exam		Potentially relevant neurological examination abnormalities, excluding incidental findings (see text)	
Comorbidities			Cognitive stagnation or decline
Imaging		Potentially relevant abnormal neuroimaging, excluding incidental findings (see text)	
Other studies – genetics, etc			Low CSF glucose and/or SLC2A1 pathogenic variant (testing not needed in most cases but strongly recommended in children with onset ≤3 years, with microcephaly and/or intellectual disability)

Are MRI or ictal EEG required for diagnosis?

An MRI is not required for diagnosis.

An ictal EEG is not required for diagnosis, provided the interictal study shows paroxysms of 2.5-3.5 Hz generalized spike wave discharge during wakefulness. However, most untreated patients will have a recorded absence seizure on routine EEG.

Syndrome without laboratory confirmation: In resource limited regions, CAE can be diagnosed in children without Alerts, who meet all other mandatory and exclusionary criteria if they have a witnessed typical absence seizure with hyperventilation.

Table 4: Diagnostic Criteria for Juvenile Absence Epilepsy

	Mandatory	Alerts	Exclusionary
Seizures	Typical absence seizures	Staring spells with typical duration >30 seconds or with postictal confusion or fatigue Absence seizure frequency of greater than 10 per day.	Any of the following seizure types: <ul style="list-style-type: none"> • Prominent myoclonic seizures • Prominent eyelid myoclonia • Myoclonic-absence seizures • Atonic seizures • Tonic Seizures • Atypical absence seizures • Focal impaired awareness seizures
EEG	Paroxysms of 3-5.5 Hz GSW (may have been obtained historically)	Lack of HV activated 3-5.5 Hz GSW in an untreated patient who performs HV well for 3 minutes or longer Persistent EEG background slowing in the absence of a sedating medication	Consistently unilateral focal spikes Diffuse background slowing Recorded typical staring spell without EEG correlate
Age at onset			<8 or >20 years
Development at onset		Mild intellectual disability	Moderate to profound intellectual disability
Neurological exam		Potentially relevant neurological examination abnormalities, excluding incidental findings (see text)	
Comorbidities			Cognitive stagnation or decline
Imaging		Potentially relevant abnormal neuroimaging, excluding incidental findings (see text)	
Other studies – genetics, etc			Low CSF glucose and/or SLC2A1 pathogenic variant (testing not needed in most cases but strongly recommended in those with microcephaly and/or mild intellectual disability)
Course of illness		Lack of GTCS over course of the epilepsy, in the absence of treatment with antiseizure medications which are effective for GTCS	
<p>Are MRI or ictal EEG required for diagnosis?</p> <p>An MRI is not required for diagnosis.</p> <p>An ictal EEG is not required for diagnosis, provided the interictal study shows paroxysms of 2.5-3.5 Hz generalized spike wave discharge during wakefulness. However, most untreated patients will have a recorded absence seizure on routine EEG.</p>			

Syndrome without laboratory confirmation: In resource limited regions, JAE can be diagnosed in persons without Alerts, who meet all other mandatory and exclusionary criteria if they have a witnessed typical absence seizure with hyperventilation.

AUTOMATED VEHICLES

Autonomous vehicle, automobile that employs driver assistance technologies to remove the need for a human operator. There are six stages of automation in automobiles, ranging from fully unassisted manual driving at stage 0 to fully automated self-driving cars at stage 5.

Though the terms self-driving or automated are commonly used interchangeably with autonomous, cars currently on the market are not capable of acting fully autonomously and cannot be operated without the intervention of a human driver. The industry standard is to use the term automated.

The stages of automation are defined by the Society of Automotive Engineers and were adopted by the U.S. Department of Transportation. The six stages follow.

Stages of Automation

- stage 0* *No automation. The vehicle is fully operated by the driver. Driver assistance is provided in the form of warnings; for example, blind spot or lane departure warnings.*
- stage 1* *The driver is fully in command of the vehicle with assistance from one automated feature. This may be in the form of automated acceleration and braking, as in the case of adaptive cruise control, in which the speed of the car adjusts automatically to keep up with the speed of traffic at a safe distance; or automated steering, in which the driver is assisted by features such as lane centring.*
- stage 2* *As in stage 1, the driver is fully in command of operating the vehicle. However, automation at stage 2 includes assistance from two automated features, for example, acceleration, braking, or steering.*
- stage 3* *Under specific conditions, automation at stage 3 enables a vehicle to operate autonomously, but a human driver must actively monitor conditions and immediately take control of the vehicle when the system alerts them.*
- stage 4* *In stage 4, a vehicle is fully self-operational within set boundaries, requiring no attention or assistance from a human driver, and indeed may not include features such as pedals or a steering wheel. Examples of stage 4 self-driving vehicles include local driverless taxis operating within “geofenced” boundaries.*
- stage 5* *Fully self-driving vehicles that require no driver assistance or monitoring and operate without boundaries or conditions. While there is work being done on this technology, experts’ predictions of the timing of its implementation and availability vary widely.*

Technology

- ✚ *Autonomous vehicles operate by using remote-sensing technology including radar, GPS, cameras, and lidar to monitor and create a 3-D map of their environment. This environment typically includes street infrastructure, other vehicles, pedestrians, traffic lights, and road signs.*
- ✚ Powerful computer systems process the gathered data and make decisions about vehicle operations, continually adjusting steering, cruising speed, acceleration, and braking, as sensors communicate constant changes about the vehicles' surroundings.
- ✚ Machine learning and artificial intelligence are foundational elements of automated vehicle systems. Through machine learning, vehicles are trained to learn from the complex data that they receive to improve the algorithms that they operate under and to expand their ability to navigate the road. Artificial intelligence enables vehicles' systems to make decisions about how to operate without needing specific instructions for each potential situation encountered while driving.

Connected vehicle technology

- ✚ Vehicles are able to “see” each other and their surroundings through radio signals, creating a fuller picture of their environment.(more)
- ✚ Connected vehicle technology enables communication with other vehicles and infrastructure. Through the use of radio signals, connected vehicles are able to “see” each other and their surroundings, creating a fuller picture of their environment—including infrastructure, vehicles, and other road users, whether in direct visual view or not. This leads to a safer environment for drivers, pedestrians, and cyclists.

The future of autonomous vehicles

- ✚ If full automation were to be achieved, advocates of self-driving technology predict that it would bring about increased road safety, as human error would have been eliminated from driving. Self-driving car technology also has the potential to reshape land-use patterns, increasing car sharing and eliminating the need for private parking spaces, as well as expanding mobility for children, the elderly, and those with physical disabilities.
- ✚ On the other hand, critics of autonomous technology predict that full automation could lead to increased vehicle miles traveled, with a resulting increase in congestion and environmental pollution. The elimination of driving could enable urban sprawl by making it easier for commuters to live farther from their places of work.
- ✚ By 2023 vehicles with partially automated systems, including lane-keeping assist, adaptive cruise control, and traffic jam assist, were on the market. Fully autonomous cars, however, are not available for purchase or use. Predictions for the availability of this technology vary. Stage 4 automation is predicted to be available to consumers by 2030–35.

C. THEORY OF MIND: MACHINE UNDERSTANDS AND RESPONDS SUCH AS BOTS/VIRTUAL/DIGITAL ASSISTANTS

- ✚ ***Theory of Mind*** *The term ‘theory of mind’ was coined by US psychologist David Premack in a famous article (Premack and Woodruff 1978) reporting experiments carried out on the chimpanzee, Sarah. The question of whether and in what sense animals other than humans have a ‘theory of mind’ ability remains unsettled.*
- ✚ ***Human beings, however, clearly attribute a wide range of mental states to each other, including intentions, hopes, expectations, imaginings, desires and beliefs. Psychologists have come to use the term ‘theory of mind’ to denote this everyday ability to attribute mental states to other people and to thereby interpret, explain, and predict their behavior. Theory of mind concerns our ability, not simply to have beliefs as such, but to have beliefs about mental states, including the recursive ability to have beliefs about beliefs.***
- ✚ Of the various mental states, at least three states appear to be basic to our commonsense explanations: desires (which identify an agent’s goals), beliefs (which reflect what an agent takes the state of the world to be), and pretence (because people do not always mean what they do or say).
- ✚ ***1. Theory of Mind and Cognitive Psychology*** *It is all too easy to overlook or take for granted our everyday cognitive abilities, even the most spectacular of them. The ability to think about thinking is a case in point.* While our knowledge of cause and effect, space, time, number, object, and so forth have been the focus of intense philosophical scrutiny since classical times, the highly abstract concepts that form cornerstones of our social intelligence have until recently received much less attention. One of the most surprising aspects of theory of mind ability is just how early in life it begins to develop.
- ✚ ***One might have thought that concepts as abstract as desire, pretense, or belief could only be acquired late in life, long after many other more concrete concepts, extensive general knowledge, and sophisticated reasoning abilities are in place.*** It would be very hard to explain to a young child what a belief is, given that one cannot point at a belief or pick one up or produce an intelligible definition. Fortunately, one does not need to do any of these things because young children acquire the relevant concepts spontaneously.
- ✚ In fact, mental state concepts emerge rapidly and without formal tuition during the preschool period (Leslie 1987, 2000, Perner et al. 1987, Wellman 1990), appear to be universal across cultures (Avis and Harris 1991), and can be acquired even by children with a clinically diminished IQ (Baron-Cohen et al. 1985, Tager-Flusberg et al. 1998).
- ✚ According to US psychologist, Henry Wellman, the concept of desire develops around two years of age (Wellman 1990). A related sensitivity to intention or goal of action seems to emerge even earlier, as shown by the 18-month-old child’s disposition to imitate the action an actor intended to perform but, due to error, did not actually perform (Meltzoff 1995).

✚ The Scottish psychologist, Alan Leslie, has shown that the concept of pretending begins to be used by children between 18 and 24 months of age. Strikingly, when children first begin to pretend play by themselves, they simultaneously acquire the ability to recognize that another person is pretending and can even share pretenses with others (Leslie 1987, 1994). The question at what age the concept of belief is first acquired has been more controversial and has attracted a great deal of study. It is generally accepted that to show a grasp of the concept of belief, the child should understand that what someone believes might be false. Working in Austria, Heinz Wimmer and Josef Perner developed a seminal task in which children are required to predict the behavior of a story character by attributing a false belief to the character (Wimmer and Perner 1983). They tested children between three and six years of age and found that it was not until the children were six years old that a majority of their subjects passed their task. Subsequently, a British team, working in London, and comprising Simon Baron-Cohen, Alan Leslie, and Uta Frith, simplified the Wimmer and Perner task and found that the vast majority of four year olds whom they tested passed by attributing a false belief to the story character (Baron-Cohen et al. 1985). Since then, such simplified tasks have been found to produce successful performance in four year olds.

✚ *The child is told a short story, with the aid of pictures or props, in which a character, Sally, has a marble. She places this marble in a basket and covers it. Sally then leaves and goes outside. While she is gone, naughty Ann enters the room, finds the marble in the basket, and moves it to a nearby box, and hides it there. The child is then asked two 'control' questions to test for basic comprehension: 'Where did Sally put the marble in the beginning?' and 'Where is the marble now?' Finally, the child is asked either directly about Sally's belief, 'Where does Sally think the marble is?' or the child is asked to predict Sally's behavior, 'Where will Sally look for the marble?' In order to answer either of the latter questions, the child must attribute to Sally a belief that is different from the child's own belief, and that the child considers to be false. Baron-Cohen et al. (1985) also found that older children with Down's syndrome who had an average IQ of 64 (in the mildly retarded range) mostly passed this test of false belief. However, the majority of a third group of children, children with autism, failed the false belief task. The group of children with autism had an average IQ of 82 and were thus in the low normal range. Their failure could not therefore be accounted for by mental retardation. Wimmer and Perner developed a further false belief task in which the child is shown a container which normally contains a well-known candy, and asked what is inside (Perner et al. 1987). After the child answers with the name of the candy, the container is opened and the child is shown that there is in fact a pencil inside. The pencil is replaced and the container closed once more. The child is then told that another person will be shown the container in the same way, and asked to predict what the other person will say is inside.*

✚ *Their findings showed that most four-year-olds could pass this task while most three-year-olds failed. Perner et al. (1989) found that a group of children with autism, despite having verbal mental ages in⁷³ excess of seven years, nevertheless*

mostly failed this task. Children with specific language impairment, matched on verbal ability with the autistic group, almost all passed. Subsequently, these patterns have been widely replicated: Normally developing four-year-old children and older moderate to mildly retarded children mostly pass false belief tasks, and most children with autism fail

2. Theories of Theory of Mind—Agreements and Disagreements

- ✚ *The existence of these early developing yet highly abstract concepts raises the question:*
 - ✚ *What is their origin?*
 - ✚ *Are these concepts innate and, if so, how is that possible?*
 - ✚ *Alternatively, are these concepts acquired and, if so, how is that possible?*
 - ✚ *Either way, we have a challenge.*
 - ✚ *If innate, how could such concepts be built into a cognitive system?*
 - ✚ *If acquired, how could such concepts be gained by children with limited experience and ability?*
- ✚ The concept of belief seems as abstract as the concept of the electron, and there seems to be as little direct everyday evidence for the existence of beliefs as there is for the existence of electrons.
- ✚ Thus, the child's theory of mind offers an intriguing case study of the nature of abstract concepts and their origins.
- ✚ *So far, two main answers to the above questions have been proposed. The first answer assumes that possession of abstract concepts depend upon possession of abstract knowledge. This reflects the traditional assumption that a concept is really a name for a packet of knowledge. In the case of theory of mind concepts, the knowledge packets are said to be commonsense theories (Gopnik and Meltzoff 1997, Perner 1991, Wellman 1990). This view of abstract concepts is referred to as the 'theory theory.' The second model assumes that possession of abstract concepts can depend upon a cognitive mechanism, rather than upon knowledge. This view is known as the theory of mind mechanism (ToMM) view (BaronCohen 1995, Frith and Frith 1999, Leslie 1987, 1992, 2000). The principle focus of debate has been around the concepts of pretense and belief. Before examining the knowledge versus mechanism dispute, we should outline points on which the two views agree.*
- ✚ First of all, while interpretations of the data differ, there are few disputes about the data themselves. Second, it is not disputed that even young children's theory of mind concepts are mentalistic rather than behavioristic in character.
- ✚ *Behaviorist philosophers and psychologists have argued that mental states do not really exist and that talk of mental states is really just a confused way of talking about behavior. However, our common sense, rightly or wrongly, is undeniably committed to the existence of mental states and, whether or not mental state concepts are 'sound,' we certainly possess and employ such concepts.*

- ✚ We naturally assume that Sally's behavior will be caused by her desire to find the marble in combination with her belief concerning its whereabouts. We use such belief-desire predictions ubiquitously and to good effect everyday. This suggests that our mental states concepts do refer to something real, probably to complex cognitive properties of agents. If concepts like pretending and believing are mentalistic rather than behavioral, then they cannot be a simple summary of sensory experience. And that will have implications for how these concepts can be acquired.
- ✚ **A third point that is not generally in dispute is that the child's theory of mind has some kind of innate basis in a disposition to employ mentalistic concepts.** And finally, there is no dispute that in the course of development, the child acquires more and more knowledge about people, about mental states, and about the role mental states play in causing people's behavior. Beyond the four points above, however, there is much controversy.
- ✚ As we have indicated, the disagreements center on the nature of abstract concepts and the role of concepts in development.
- ✚ *If a concept is the 'name' of a body of knowledge, then the only way to acquire a given concept is to acquire the appropriate body of knowledge. Another way to look at this is that possession of a given body of knowledge determines possession of a given concept. So, if this body of knowledge changes, then so too will the associated concept. Since everyone agrees that children's knowledge changes in the course of development, then so too, on this view, must their concepts change.*
- ✚ Hence, a major topic in the study of development is the study of conceptual change (Carey 1985). A further implication of this widespread view is that if a given concept is innate, then its associated body of knowledge must also be innate. Often, it is assumed that if something is innate then it cannot change with development.
- ✚ However, some theory theorists have proposed that some knowledge (specifically, knowledge of some theories) is both innate and revisable (Gopnik and Wellman 1994, 1995). On this account, both the child's innate and acquired concepts can and do change.
- ✚ *According to the ToMM view, concepts do not name bodies of knowledge but instead directly represent properties in the world. This means that the representational relation between a concept (in the head) and the property (in the world) that it designates is subserved by a causal mechanism of some sort.*
- ✚ For example, the concept 'red' is typically established and sustained by the neurocognitive mechanisms of color vision. Without color vision, it is hard to attend to and learn about redness. But nothing you will learn about redness changes the meaning of the concept 'red,' namely, the property of redness. On this view, concepts do not depend upon knowledge; on the contrary, they are a prerequisite for obtaining knowledge. Therefore, conceptual development is not the same thing as knowledge development.

- ✚ Consequently, innate concepts do not require innate knowledge, and acquired concepts do not depend upon acquiring knowledge. Instead, on this view, we have the job of inquiring into the nature of the neurocognitive mechanisms that establish and sustain conceptual representations. The above contrasting assumptions about the nature of concepts lead to very different interpretations of theory of mind development.

3. **Early Development, Normal and Abnormal** As mentioned earlier, 18-month-old infants will imitate the goal of an action rather than simply the movements comprising the action.

- ✚ *It is possible that they do so because infants have knowledge of a theory of goal-directed action. If so, the 'theory of goals' they have knowledge of needs to be discovered and specified. Alternatively, infants may not have knowledge of a theory at all; their imitation may instead simply reflect the operation of a neurocognitive mechanism.*

- ✚ *The 'mirror neurons' discovered by Rizzolatti in his investigations of cells in the ventral premotor cortex (area F5) in monkeys are suggestive in this regard. Neurons with similar properties are apparently involved in human imitation (Iacoboni et al. 1999). Such mechanisms with their sensitivity to the 'hidden' goals of an actor's actions may form the innate basis for the concept of desire.*

- ✚ Early pretending provides another example of an early social sensitivity where it is hard to imagine what the content of the infant's theory, in this case 'theory of pretending,' might be. Leslie (1987) drew attention to a key and long ignored aspect of early pretending, namely, that when the ability for solitary pretending first emerges, between 18 and 24 months, so does the ability to recognize pretending in other people.

- ✚ Leslie went on to argue that pretending should be understood as part of our theory of mind ability. Instead of proposing that the infant possessed innate knowledge of a theory, he postulated the existence of a specialized neurocognitive mechanism, ToMM, that forms the innate basis for theory of mind concepts, including the concept of pretending. A number of empirical predictions followed from the link between pretending and ToMM, including the prediction that autistic children, who do not develop spontaneous pretending, would be specifically impaired in their understanding of belief. As we saw, this prediction has been confirmed.

- ✚ However, this confirmation in itself does not prove the link between autism and an impaired neurocognitive mechanism like ToMM. One alternative possibility is that children with autism have some kind of general processing difficulty, for example, with language, or with abstract reasoning, or with working memory, or with some other general factor that is demanded by false belief tasks.

- ✚ Any of these possibilities would provide an alternative explanation to the idea of a specific impairment caused by an impaired ToMM. To test the alternative

general processing explanation for autistic performance on false belief, Leslie and Thaiss (1992) adapted tasks devised by Zaitchik (1990).

- ✚ Zaitchik had developed a task in which a Polaroid photograph is taken of an object in a certain location. The object is then moved to a new location. Following control questions about where the object was when the photograph was taken and where the object is now, the child is asked to say where the object is in the photograph.
- ✚ The child does not get to see the photograph. After all, the child in the Sally and Ann task does not get to see Sally's belief! Zaitchik found that preschool children tend to fail or pass both the photographs task and the Sally and Ann task about the same age. This is not surprising given how similar both tasks are from the point of view of general problem solving.
- ✚ When Leslie and Thaiss tested a group of high-functioning autistic adolescents and a group of normally developing four-year-old children on similar pairs of photographs and false belief tasks, they found that the four year olds performed well on both sets of tasks.
- ✚ The autistic subjects, as expected, performed poorly on the false belief tasks, but their performance on the photograph tasks was nearly perfect. This result too has been replicated and extended. The failure of autistic children on false belief tasks cannot be attributed to the general problemsolving demands of the tasks. The photograph tasks make the same or similar demands and the autistic subjects passed these. So far, we have considered only comparisons between autistic subjects who fail false belief tasks and normally developing four year olds or nonautistic developmentally disordered subjects who pass. But we should also compare autistic subjects with normally developing three-year-olds who also fail. Do the latter two groups fail false belief tasks for the same reason?
- ✚ The ToMM model predicts that they fail for different reasons. Only in the case of autism is ToMM neurologically impaired; in the typically developing threeyear-old, ToMM should be intact, limited only by immaturity. One way to test this prediction is to examine whether ways that help three year olds to pass false belief tasks also help autistic subjects. Siegal and Beattie (1991) found a very simple way to help three year olds achieve better performance on the Sally and Ann task. Instead of asking, 'Where will Sally look for her marble?,' they asked 'Where will Sally look first for her marble?' Surian and Leslie (1999) found that while this small change did indeed help three-year-olds to calculate Sally's false belief, it had no effect on autistic performance, which remained just as poor.
- ✚ Other manipulations which help typically developing threeyear-olds but which do not help older children with autism have been reported by Roth and Leslie (1991, 1998).
- ✚ Finally, although the existence of the neurocognitive mechanism ToMM was originally postulated on the basis of purely cognitive findings, recent advances

in brain imaging techniques are beginning to reveal the neural systems basis of this instinct (Frith and Frith 1999).

4. Summary

The ability to detect the hidden mental states of other people is a remarkable cognitive power. It is tempting to assume that the idea of mental states must rest upon the same sort of foundation as many other highly abstract ideas, like electrons or inflation—that is, upon the capacity for scientific reasoning and theorizing. However, the speed and precocity with which ideas about mental states emerge in normal development, together with the patterns of normal and abnormal development reviewed above, suggest a different picture. The basic theory of mind that typically emerges during the preschool period is based upon an in-built mechanism that provides an intuitive and instinctual sensitivity to the mental states of other people.

D. SELF-AWARE: HUMAN LIKE INTELLIGENCE SUCH AS SUPER ROBOTS IN SPACE ETC.

What is Artificial Intelligence (AI)?

- ✚ There are probably as many definitions for AI as there are areas within AI and the numbers keep growing. However, I am including a few below.
- ✚ AI refers to “robots, computers, and other machines with a humanlike ability to reason and solve problems” (McPherson, 2018, p. 4).
- ✚ Artificial intelligence is the theory and development of computer systems that are able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.
- ✚ Artificial intelligence makes it possible for machines to learn from experience, adjust to new inputs, and perform human-like tasks. Most AI examples that you hear about today—from chess-playing computers to self-driving cars—rely heavily on deep learning and natural language processing
- ✚ AI or machine intelligence (MI) is intelligence displayed by machines, in contrast with the natural intelligence (NI) displayed by humans and other animals.
- ✚ AI is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with human beings. It is technology with the ability to reason and solve problems.
- ✚ AI mainly focuses on understanding and performing intelligent tasks such as reasoning, learning new skills, and adopting to new situations and problems. It is a combination of computer science, psychology, and philosophy (Mogali, Artificial Intelligence and its applications in Libraries).
- ✚ AI refers to science and engineering that explores how to simulate various issues and functions in the field of human intelligence. AI technology fields cover perception, recognition, reasoning, the learning process, natural language, machine translation, games, chess, and so on.

Self-Aware: Human like intelligence such as Super Robots in Space etc.

- ✚ From Apple’s SIRI to self-driving cars, AI is progressing rapidly. While science fiction often portrays AI as robots with human-like characteristics, AI can encompass anything from Google’s search algorithms to IBM’s Watson to autonomous weapons (Benefits & Risks of Artificial Intelligence).
- ✚ Future of Life Institute. Max Tegmark). AI is a major part of many cutting-edge technologies, including robotics, driverless cars, web searches, and video games. AI technologies use sophisticated algorithms, or sets of instructions, to solve very difficult tasks (Hulick, 2016, p. 12).
- ✚ Some AI technologies work behind the scenes to figure out who and what people like while they are using social media or shopping online (Hulick, 2016, p. 12).
- ✚ *The Space Robotics Technical Committee has two main areas of interest: Microgravity and Planetary Robotics. Microgravity Robotics includes manipulation*

and mobility for scenarios such as International Space Station (ISS) operations and satellite servicing. Planetary Robot systems address scenarios such as Mars and lunar exploration using manipulation or mobility on or near the surface. Some scenarios, such as asteroid and comet exploration, have environments with low gravity which may blur the distinctions between these categories. For Microgravity Robotics the space environment (radiation, contamination sensitivity, thermal extremes, etc.) poses unique challenges to robot and robot algorithms. Despite this, it is expected that the robotics discipline will find increasing importance in coming years, particularly as the opportunities for human-robot and robot-robot cooperation arise in space exploration. Priority areas for this technical committee include:

- *Electromechanical design and control.*
- *Microgravity locomotion.*
- *Machine vision for inspection and assembly, including compensation for stark lighting, glare, glint, and deep shadows.*
- *Command and control interfaces, including teleoperated modes.*
- *Power sources and consumable recharging techniques.*
- *Radiation hardening and effects on processing throughput.*
- *Thermal considerations in space robot design.*

For Planetary Robotics, the surface environment also poses unique challenges. These include Microgravity Robotics' issues during cruise phase, or if an atmosphere is not present. Further, there is usually the greater uncertainty of interacting with an unexplored natural terrain instead of man-made structures. Planetary Robotics technical topics include:

- *Sensing and perception for planetary exploration, including terrain-relative precision position estimation.*
- *Above-surface, surface, and sub-surface planetary mobility, possibly from novel vehicle design concepts.*
- *Command and control with limited bandwidth, often precluding teleoperation and requiring autonomous surface operations, with natural terrain navigation and manipulation.*
- *Planetary rovers systems engineering.*
- *Testing and qualification, including field tests on Earth and Mars.*
- *Human-Robot system design and development.*

<https://www.ieee-ras.org/space-robotics>