Pierre Marquis
Odile Papini
Henri Prade
Editors

A Guided Tour of Artificial Intelligence Research

3 Interfaces and Applications of AI

Springer

Artificial Intelligence and High-Level Cognition

Marco Ragni

Abstract Artificial intelligence (AI) and cognitive science (CS) both investigate information processing, but with a different focus: AI aims to build *problem solving machines*, i.e., systems capable of solving diverse problems in an efficient and effective way while CS analyzes human cognition. Both approaches increase an understanding of the foundations, methods, and strategies that can be employed to perform in a natural or artificial environment. This chapter focuses on *high-level cognition*, i.e., cognitive processes that are related to reasoning, decision making, and problem solving. After an introduction to the core principles, intersections, and differences between both fields, some central psychological findings are presented. In a next step cognitive theories for high-level cognition are introduced. While the architecture of cognition has an impact too, main approaches for cognitive modeling from cognitive architectures to multinomial processing trees are analyzed. Current challenges conclude the chapter.

1 Introduction

Artificial intelligence (AI) and cognitive science (CS) deal with information processing including analyzing and understanding how to store, manipulate, and derive new information. Both fields differ in their respective goals: AI aims to built efficient problem solving systems and CS aims to understand and to model human behavior. But both have something to offer to each other: AI provides a precise language and methods to describe information processes while CS investigates a working intelligent system. A connection between the disciplines is built upon the mapping from *brain* and *mind* to *hardware* and *program* (Searle 2004). An AI that focuses on excelling on a clear defined bounded domain is called *weak AI*. Strong AI's ultimate goal is to develop a system that does not only "simulate having a mind; it literally has a mind" (Searle 2004). This requires an understanding about the limits and powers

M. Ragni (✉)
Cognitive Computation Lab, Technical Faculty, University of Freiburg,
Freiburg im Breisgau, Germany
e-mail: ragni@cs.uni-freiburg.de

P. Marquis et al. (eds.), *A Guided Tour of Artificial Intelligence Research*,
https://doi.org/10.1007/978-3-030-06170-8_14

of the human mind and to identify its setpoints, i.e., the specifics or properties of the system. If the properties are unknown, then it is not possible to evaluate if a system demonstrates capabilities we typically ascribe *a mind*. Arguments put forward (e.g., in the Chinese room argument Searle 1980) show that it is not easy to distinguish a mind from a mindless simulation. What makes the human mind as a system interesting from an AI perspective, is that it is not limited to a specific domain (like navigation), but that it is a general system that goes beyond any domain limitations, e.g., the mind that navigates is the same that performs any other cognitive operation. Still, this strong AI approach is not what many AI researchers focus on nowadays.

One way to develop strong AI systems is to take humans as a cognitive prototype that demonstrates intelligent processes. Humans are functioning instances of *intelligent systems* and despite great advances in AI that demonstrates the power of the systems over human performance, there are still specific types of problems that humans solve better. If, for instance, only imprecise information is given or *insight* is necessary, humans can often generate a solution that is *satisficing* and, they can adapt and generalize results to other domains. As humans *do* show features we expect from intelligent systems, it is worthwhile to learn *how* humans process information, to gain insights in order to model high-level cognitive processes computationally, and to make cognitive processes available for technical systems. An artificial system does not necessarily need to mimic human behavior, but it can be built on cognitive principles. Furthermore, as AI systems enter more and more our everyday life the ways humans interact with AI systems increases. For a better human/AI interaction this requires to equip interacting systems with an understanding of human information processing. Examples are any kind of technical systems that need to provide information in a comprehensible way. For this endeavor methods from AI are interesting as the possibility to express cognition as algorithms, to analyze the algorithmic features of cognition, and having a general and formalized set of tools available is methodologically sound. However, an interaction between AI and CS would not be possible if there is no overlap between the fields. Consider for example the following definition (Foundation 1978) based on computational processes:

> What the subdisciplines of cognitive science share, indeed, what has brought the field into existence, is a common research objective: to discover the representational and computational capacities of the mind and their structural and functional representation in the brain.

Relevant are the notions of *representational* and *computational capacities of the mind*. These notions can be described using concepts from AI, but the core paradigm of CS is to *"equate mental processes with information processing"* (Strube et al. 2013) and that "[t]he overall accepted notion in cognitive science is that symbols lie at the root of intelligent action" (Newell and Simon 1976) and hence a "structural requirement for intelligence [...] is the ability to store and manipulate symbols" (Newell and Simon 1976). This is termed the *physical symbol system* hypothesis (Newell and Simon 1976): "A physical symbol system has the necessary and sufficient means for general intelligent action". Hence, AI and CS use similar methods such as symbolic descriptions. *The* characteristic method of CS—*cognitive modeling*—is simply not possible without techniques from AI (Strube et al. 2013). Despite the fact that human intelligence is the current creator and main foundation in creating any AI system, a new movement to create cognitive systems, i.e., systems that incorporate principles of human cognition, has started (Hollnagel and Woods 2005). Once psychological theories are formulated algorithmically, they cannot only be tested—they can be made available for AI systems.

A distinction is made between *low-level* and *high-level* cognitive processes: Low-level processes are typically associated with sensation or simple memory processes; they do not require effort, are often unconscious, and most humans perform them easily and intuitively. In contrast, high-level cognitive processes are demanding and often require the simultaneous execution of several mental processes of memory and imagination. Examples in the literature list problem solving, decision making, learning, and language comprehension among high-level processes, e.g., Just et al. (1999), Sternberg (1999), lle Lépine et al. (2005), O'Reilly (2006), Dubois et al. (2008). As it is hardly the case that intelligence is considered by researchers from CS without relating it to the capability to reason and to solve problems we will focus on both aspects in the following sections.

2 Core Empirical Results on High-Level Cognition

In this section we will get a flavor about specifics of human high-level cognition and how cognition is actually not captured by classical normative AI approaches. A core aspect of intelligence, be it natural or artificial intelligence, is the ability to *reason* about given information and to *solve problems*. *Reasoning* is the ability to gain new information from existing knowledge. Processes of classical reasoning and problem solving are similar to each other. They are often not treated differently—neither in AI nor in CS (Sternberg 1980; Newell 1979; Wason 1971; Greeno and Simon 1988; Baral 2003). Can we define a common basis for problem solving and reasoning? Typically, *information* and a *task* is given. The information can comprise premises in the case of reasoning or a description of an initial state for a problem. It is often so-called *declarative knowledge* (see below). The task description is often more *procedural*, i.e., derive new information or generate a different scenario by the application of some operators. In logical reasoning, however, humans do not expect to be informed about applicable operators. Humans take implicitly for granted how they have to process given information (cf. Table 1). Despite the great similarities between the processes of logical reasoning and problem solving, a difference lies in the (under-) specified operations. While applying known operations in a search problem is simple; it is still not possible, despite great advances in the field of AI, to construct machines that solve arbitrary *insight* problems (see below), while humans have demonstrated this ability.

Table 1 A comparison between reasoning and problem solving

	Logical reasoning	Problem solving
Given information	Premises	Initial scenario; goal scenario; (partially) operators
Task/Question	What follows? Or: Does conclusion X follow?	Is there a transformation from the initial to the goal scenario (using the operators)?
Operators	Often not given	Sometimes explicitly given

2.1 Psychological Findings

Research on human reasoning can be roughly divided into the categories of deductive, inductive, and abductive reasoning (cf. Strube et al. 2013).

Deductive reasoning can be defined as the method of drawing a conclusion from a given set of statements. As a normative framework psychologists often evaluate the answers from a normative perspective, e.g., a statement is only accepted as a correct conclusion if it can be derived by applying predicate logic to the premises.

Given: A set of premises $p_1, \ldots, p_n$.
Question: What follows? (Which conclusion(s) can be drawn?)

In contrast to logic, conclusions humans draw, need to be *meaningful* and *different* from the premises though there is no crisp definition frame. This *meaningfulness* often orients itself by Gricean communication principles (see below).

Inductive reasoning often require to formulate and test a hypothesis, i.e., a statement that describes data. A typical example is Halbmayer and Salat (2012): Given the observations that doves, eagles, hawks are birds and can fly, an observer could form the hypothesis that all birds can fly. Hence such an inference is based on a number of observations and a summarizing expression (often formulated as a conditional, e.g., if something is of type B then they have property F).

Given: Two sets M', M with $M' \subset M$ and relation R holds for all m of M'.
Question: Holds relation R for all m of M?

Humans do accept conclusions as long as there is no identified counterexample. If we learn that a penguin is a bird that cannot fly, the previously drawn conclusion that birds can fly does not hold anymore. In other words, inductive inferences cannot be inferred with certainty. *Reasoning about analogies* is often classified as a special form of inductive thinking (Beller and Spada 2001). It transfers knowledge of a *source* domain to a *target* domain. A possible formal definition of an analogical problems is (cp. Strube et al. 2013):

Given: Two domains D_1 and D_2; in D_1 the relation R holds between elements E_1 and E_2.
Question: Is there a function f, such that for elements $f(E_1), f(E_2)$ from D_2: $R(E_1, E_2) \in D_1 \Leftrightarrow R(f(E_1), f(E_2)) \in D_2$ holds?

Abductive reasoning is a mode of reasoning (Douven 2011). It has been characterized of as finding a best explanation $E_1, \ldots, E_n$ for a fact F. It has been so far mostly neglected in the psychology of reasoning.

In contrast to classical methods from AI, human reasoning mechanisms are not always sound with respect to a given normative theory. Though some errors are due to lack of focus or misunderstanding the majority of errors by most reasoners are systematic deviations from classical normative theories like predicate logic or probability theory. Researchers have focused on such aspects as they provide insights about the underlying mental representations and mechanisms.

2.1.1 Relational Reasoning

The way how humans represent and reason about relational information and what can cause reasoning difficulty depends on many factors on different cognitive levels. Before we analyze them, consider the following problems (Ragni and Knauff 2013):

(1a) Flight UA is north of flight LH.
Flight LH is north of flight AA.
Flight AA is north of flight DL.
____
What follows for flight UA and flight DL?

(1b) Flight UA is north of flight LH.
Flight UA is north of flight AA.
Flight AA is north of flight DL.
____
What follows for flight UA and flight DL?

The left-hand problem (1a) is called a *determinate problem*, i.e., there is only one qualitative arrangement of the aircrafts possible. Instead the right problem (1b) is called an *indeterminate problem*, i.e., different qualitative arrangements of the flights are possible. Nonetheless, the conclusion is for both cases identical: flight UA is north of flight DL.

We now analyze the processing of such (spatial) relational information and its relation to the internal mental representation and will refer to the following three levels later on.

Level 1: Processing of information. Some core findings are: The *symbol distance effect*, i.e., information that is presented in such a way that it is in relation to an information presented immediately before (*continuous order*) is easier to process than information that is at first unrelated and only later related and integrated (*discontinuous order*) (Potts 1974). Another factor is the *relational complexity effect*, i.e., information that contains relations with higher arity (e.g., the ternary relation in-between) is more difficult to reason with than information formulated with relations of smaller arity (Halford et al. 2010).

Level 2: Internal representation of information. The *ambiguity or indeterminacy effect of information*, i.e., if a relational description is ambiguous and, hence, allows for several possible models, a conclusion is harder to draw then if the description allows only for one model, e.g., the indeterminate vs. determinate problems above. The *preference effect*: Reasoners tend to build a preferred mental model, based on a direct integration of information, for ambiguous descriptions and form preferred conclusions based on that. The generation of the preferred model can be due to working memory limitations. The *visual impedance effect*, i.e., relations that are easier to visualize can *impede* the reasoning process in contrast to relations that are rather visually abstract but spatial in their nature (Knauff and Johnson-Laird 2002). This potentially requires additional effort in brain regions connected with visual information (Knauff 2013). But additional visual presentations can support reasoning processes too, e.g., Bauer and Johnson-Laird (1993) presented problems with diagrams that enhance the idea of alternative interpretations resulting in better performance.

Level 3: Manipulation of the internal representation. The *transformation distance effect*: Reasoners tend to neglect alternative models, especially if the operational distance to transform the preferred mental model into the alternative mental model is high (Ragni and Knauff 2013). This explains a source of reasoning errors on an operational level, especially if an alternative mental model is a counterexample to a putative conclusion a reasoner forms based on their preferred mental model.

2.1.2 Syllogistic Reasoning

The above introduced three levels are often part in any kind of reasoning processes. One particular domain deals with reasoning about quantities and often only with syllogisms. In CS and psychology, syllogisms use the classical quantifiers such as *All, Some, Some .. not*, and *None*. Recently, generalized quantifiers such *Most* or *Few* and *Normally* have been investigated. Consider the following example (Klauer et al. 2000):

(2a) No cigarettes are inexpensive.
Some addictive things are inexpensive.
―――――――――――――――
Some addictive things are not cigarettes.

(2b) No addictive things are inexpensive.
Some cigarettes are inexpensive.
―――――――――――――――
Some cigarettes are not addictive.

Most participants accept the conclusion (below the line) in (2a) which is a valid answer but fewer (46%) accept the conclusion in (2b) despite being valid. This indicates that humans tend to evaluate the truth of a putative conclusion not necessarily on the given premises but how convincing a conclusion is. This is called the *belief-bias effect*. Another result is that just the internal problem representation can influence the responses. Consider:

(3a) All astronauts are computer specialists.
Some computer specialists are nerds.
―――――――――――――――
What, if anything, follows?

(3b) Most As are Bs.
Most Bs are Cs.
―――――――――――――――
What, if anything, follows?

Many participants (70%) conclude for similar problems like (3a) that "Some astronauts are nerds" (Khemlani and Johnson-Laird 2012). Although this is a possibility, it cannot be logically concluded, but this suggests that humans deviate from classical logical inferences. Problems using the generalized quantifier *Most* (see (3b) above), where about 60% select *Most As are Cs*, give hints at the internal mental representation that influences the answers but the counterexample require to think about different set sizes. On the level of *processing the information* (see above) the quantifiers are interpreted differently than in formal logics: The Gricean implicature claims that communication principles have an influence on the interpretation of quantifiers (Newstead 1995), e.g., *Some* is interpreted as *Some, but not all*. Similar to relational reasoning the order in which information is presented has an influence on accuracy (Khemlani and Johnson-Laird 2012).

On the level of *the internal representation of information* the adequate representation for the average reasoner is not yet identified: A meta-analysis (Khemlani and Johnson-Laird 2012) demonstrates that any cognitive theory, be it model-based, rule-based, probabilistic or heuristic, deviates significantly from the empirical data of six studies; the current best model is mReasoner (Khemlani and Johnson-Laird 2013), a system based on mental models and heuristics. A potential explanation for the deviations could be the large inter-individual differences between reasoners (see Challenge 4 below). Recently, an analysis of subgroups has been undertaken (Khemlani and Johnson-Laird 2016). As the internal representation of the second level is not yet fully understood, only some theories (e.g., the mental model or the PHM theory) formulate processes on Level 3 where internal representations need to be manipulated; but there is not enough psychological data to analyze this. Recently, an analysis of the cognitive difficulty of language processing and cognitive resources including formal approaches like parametrized complexity measures has been conducted (Szymanik 2016).

2.1.3 Reasoning about Conditionals and Propositions

A conditional statement can be used to describe observations or explain facts, e.g., "if it rains then the street gets wet", it allows one to formulate scientific predictions, e.g., "if the air pollution continues, the ozone hole increases" or to reason about counterfactuals, e.g.,"if Oswald had not shot Kennedy, then someone else would have" (Byrne 2007). Conditionals can describe causal or temporal dependencies, definitions, and compressions of observations, aggregating different aspects in a short description:

(4) If the system passes the Turing test, then the system is intelligent.

A conditional consists of an *antecedent*, e.g., in the conditional above, "the system passes the Turing test", and a consequence, "the system is intelligent". Four inference rules have been investigated in the context of conditional reasoning. Let us consider the case where participants have the conditional above and additionally a fact be given, then four rules are possible: The *modus ponens* rule (for a, and, if a then b, conclude b), hence, for "The system passes the Turing test" then it can be inferred that "the system is intelligent". The three other rules are *denial of antecedence* (from $\neg a$, and, if a then b, conclude $\neg b$), *affirmation of the consequence* (from b, and, if a then b, conclude a), *modus tollens* (from $\neg b$, and, if a then b, conclude $\neg a$). While only modus ponens and modus tollens are logically correct, humans do, depending on the information and the background context accept or derive other conclusions, too. If participants receive first a negative consequence $\neg c$ and only then the conditional *if a then c*, the participants apply the modus tollens more often (Legrenzi et al. 1993). Participants without training in formal logic suppressed previously drawn conclusions when additional information became available (Byrne 1989). Interestingly, in some instances the previously drawn conclusions were valid whereas in other instances the conclusions were invalid with respect to classical two-valued logic. Consider the following *suppression effect* (Byrne 1989):

(5a) If she has an essay to write,
then she will study late in the library.
If she has a textbook to read,
then she will study late in the library.
She has an essay to write.
What, if anything, follows?

(5b) If she has an essay to write
then she will study late in the library.
If the library stays open,
then she will study late in the library.
She has an essay to write.
What, if anything, follows?

Most participants (98 %) concluded for problem (5a) *"She will study late in the library"*. If participants instead receive problem (5b), only 38% of the participants concluded *"She will study late in the library"*. This shows that although the conclusion is logically still correct, it is suppressed by an additional conditional which is an supports the assumption that human reasoning demonstrates features of *non-monotonic* reasoning. A characteristic of non-monotonic reasoning is, that new information can reduce knowledge. The famous Wason Selection Task tests how humans evaluate a hypothesis formulated as a conditional (Ragni et al. 2018):

(6) The experimenter explains to the participants that each card in a pack has a letter on one side and a number on the other side. Four cards chosen at random from the pack are placed on the table, e.g., E K 2 3. The experimenter presents the following general hypothesis: If there is a vowel on one side of a card, then there is an even number on the other side. The participants task is to select all those cards, and only those cards, which would have to be turned over in order to discover whether the hypothesis is true or false about the four cards on the table. Participants make their selection; and the task is over.

Many participants check only the card with the vowel or the card with the even number, albeit the logical correct answer is to select the card with the vowel and the card with the odd number. If the abstract task is replaced by an isomorphic representation with drinking beer and being over 18—a deontic formulation of the task is obtained for which more participants chose the classical logically correct answers. For conditionals so-called *enablers* or *defeaters* can exist, i.e., facts that support the conditional or not and they can have an effect on the construction of a mental representation. As a consequence they can increase or decrease the likelihood to draw an inference (Verschueren et al. 2004).

Reasoning about counterfactuals is a "mental undo" of a fact or observation, e.g., "if Oswald had not shot Kennedy, then someone else would have", these are often generated after goal failures and are related to causal thoughts (Byrne 2002). The role of counterfactuals is to test the relation between different antecedents and the cause hereby identifying the strength of the (causal) connection between antecedent and consequence in a conditional. They lead to an increase of the application of the modus tollens in reasoning (Byrne 2002) and there is a preference to think about exceptional alternative events and actions (Dixon and Byrne 2011). There exists illusions in reasoning about propositional assertions (Khemlani and Johnson-Laird 2009). Illusions are inferences where participants are convinced that the drawn conclusions are true, while they are wrong:

(7) Consider, for instance, the following assertion: "You have the bread, or else you have the soup or else the salad. Given the further premise: You have the bread." What follows? Can you have the soup too? What about the soup and the salad?

Please note that participants were instructed to interpret the or else as an exclusive or (XOR). Hence the problem can be reformulated as bread XOR soup XOR salad. But, only 17% of the participants gave the correct answer that you can have all three. This answer may depend on the underlying mental representation as we will see later.

2.1.4 Common-Sense and Heuristic Reasoning

Psychological findings indicate that human reasoners do deviate from classical logical approaches. But do human reasoners adhere to the laws of probability? Consider the following Linda problem (Tversky and Kahneman 1983):

(8) Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

A Linda is a bank teller.
B Linda is a bank teller and is active in the feminist movement.

Statement B is a conjunction of the statement A: "Linda is a bank teller", and another statement "is active in the feminist movement" (which we abbreviate by C). Hence the probability of the joint event $P(B)$ is the probability $P(A \land C)$ and this can be at most as high as the single probability $P(A)$ (short: $P(A \land C) \leq P(A)$). But, most participants (85%) select response B and decide that the answer $P(A \land C)$ is more likely then $P(A)$. Hence, they deviate from the predictions of the probabilistic calculus.

An example of a connection between heuristic reasoning and the involvement of memory is the so-called *availability heuristic* (Tversky and Kahneman 1973):

(9) Suppose you sample a word at random from an English text. Is it more likely that the word starts with a K, or that K is its third letter?

About 66% of the participants stated incorrectly that there are more words that start with a K with the rest stating the correct answer that the third letter is more likely. People do consistently commit such an error because the initial letter has a more relevant role in our memory. As a lexical arrangement, it seems to play a large role, so ultimately, a memory management principle is the cause of this misconception.

In decision making theory another deviation from probability theory is the *sure thing* principle (Busemeyer and Bruza 2012):

(10) Imagine that you have just played a game of chance that gave you a 50% chance to win $200 and a 50% chance to lose $100.
(a) The coin was tossed and you have [won $200/lost $100] and you are now offered a second identical gamble.
(b) Imagine that the coin has already been tossed but that you will not know whether you have won $200 or lost $100 until you make your decision concerning a second, identical gamble.
Would you accept or reject the second gamble?

While in condition (10a) most participants accept a second gamble (69% if they have won and 59% if they have lost) only 36% of the participants would do so in condition (10b).

2.1.5 Problem Solving

Research in psychology on problem solving started with the work of Gestalt psychologists with many problems that are called *insight problems* (Duncker 1945; Wertheimer 1923). Much due to the influence of the General Problem Solver research into *permutation problems* started shortly afterwards (Newell and Simon 1972). And, with computer analysis and strategy games the domain of *complex problems* followed Funke (2006). They are characterized depending on additional features (see, Table 2). The probably most famous AI method, namely a *search of the problem space* often fails to describe the human reasoning process due to limitations of the human working memory. Instead, content is accessed from long-term working memory or the method of case-based reasoning is used Strube et al. (2013). All three problem classes are now introduced.

Table 2 An overview of different problem classes. *Static*: the problem does not change while the agent deliberates; *Observable*: all relevant information of the problem is given; *Search Space*: the set of all possible states is given with the problem description; *Operators*: the set of all possible operations is given with the problem description. *redefine**: requires to adapt given operators or the search space. [a]Introduced by Russell and Norvig (2003); [b]Introduced by Duncker (1945); [c]Introduced by Frensch and Funke (1995)

Problem type	Environment	Observable	Search space	Operators	Examples
Permutation[a]	Static	Fully	Defined	Known	Tower of Hanoi
Insight[b]	Static	Fully	Identify and redefine*	Identify and redefine*	Raven's IQ test
Complex[c]	Dynamic	Partially	Partially	Partially	Lohhausen

Permutation problems. An instance of this class is the classical *cannibals and missionaries* problem (Simon and Newell 1962):

(11) There are three missionaries and three cannibals on the bank of a wide river, wanting to cross. There is a boat on the bank, which will hold no more than two persons, and all six members of the party know how to paddle it. The only real difficulty is that the cannibals are partial to a diet of missionaries. If, even for a moment, one or more missionaries are left alone with a larger number of cannibals, the missionaries will be eaten. The problem is to find a sequence of boat trips that will get the entire party safely across the river-without the loss of any missionaries.

Other examples include the Tower of Hanoi/London (Anderson 2007; Kaller et al. 2004) or the PSPACE-complete Rush Hour problem (Flake and Baum 2002). Psychologists often call the applied method *means-end analysis*: Identify the *ends*, possible operations (the *means*), and then select the operators that, applied to the current state, reduce the distance to the goal (*difference reduction*), which is a form of Greedy strategy. During problem solving applying an operation that requires to increase the distance from the current state to the goal is more difficult (as in the cannibals and missionaries example above). Recent results indicate that humans apply a specific kind of chunking of objects in the search space (Bennati et al. 2014). Hence, not only the working memory capacity but as well the specific internal representation influences the planning process. As a result it indicates that humans are not searching the entire problem space, but rather prune the search tree by systematically preferring specific operations or heuristics.

Insight Problems

Another important class of problems are so-called *insight problems*. These problems, originally inspired by Gestalt psychologists, e.g., Duncker (1945), cannot be solved by an exhaustive search (Chu and MacGregor 2011). They almost always require a spontaneous insight, called the *Eureka effect*, and the identification of operators (Gilhooly and Murphy 2005). From a computational perspective, such problems are difficult to conceptualize and to algorithmize. The candle problem, originally termed the box problem (Duncker 1945), is the following:

(12) On the door, at the height of the eyes, three small candles are to be put side by side (“for visual experiments”). On the table lie, among many other objects, a few tacks and the crucial objects: three little pasteboard boxes (about the size of an ordinary matchbox, differing somewhat in form and color and put in different places). Solution: with a tack apiece, the three boxes are fastened to the door, each to serve as platform for a candle. In the setting a.p., the three boxes were filled with experimental material: in one there were several thin little candles, tacks in another, and matches in the third. In w.p., the three boxes were empty. Thus F_1: “container”; F_2: “platform” (on which to set things).

An explanation of this solution is the so-called *functional fixation*: human reasoner consider the matchbox as a container for objects, but not to be a fixture for the candle. Other findings (Fauconnier and Turner 2008) support that human reasoners construct *mental spaces*. To solve some insight problems, these mental spaces must become superimposed to so-called blended spaces where operations are possible that were not in any of the original mental spaces. Many IQ-test problems require insight as the operators of the problem are not specified. A specific class of IQ-tests are *geometrical analogy problems*. An example is Raven's *Standard Progressive Matrices* (SPM), to test average intelligent adults, and *Advanced Progressive Matrices* (APM) to test above average intelligent adults (Raven et al. 2000; Raven 1962). Participants are presented with a 3×3 matrix where in 8 of the 9 cells geometrical objects are present and the underlying function needs to be inferred. Especially the latter aspect, identifying patterns, is an important aspect of any intelligent system, and can provide a fruitful benchmark for general approaches.

Complex Problems

In contrast to the static problems above, complex problems change over time and are intransparent (Frensch and Funke 1995). The lack of transparency is due to the fact that the exact properties of the given state, the target state and the operations (or at least some) are unknown. Solving complex problems requires an efficient interaction between the user and the situation-related constraints of the task. It may require cognitive, emotional, and other skills and knowledge. Examples include a strategic game scenario "Lohhausen" (Dörner et al. 1983). In Lohhausen participants had to govern a small city by successfully managing the almost 2000 system variables. Findings from Lohhausen indicate that classical intelligence tests have only a small predictive power for failure or success, in contrast to self-confidence (Funke 2006). The internal representation followed a linearization effect, namely that participants rather thought in causal chains instead of causal networks, i.e., a tendency to linearize cause-effect and a reduction to often only one cause were observed. Another example are dynamic stock and flow problems (DSF) that represent stocks, accumulations of a certain amount of a quantifiable unit, and flows, that increase or decrease the stock amount over time. The task is to predict and react to the underlying changes in order to keep the stock in balance. Humans are not good in predicting the changes (Cronin et al. 2009). An overview about the involved cognitive complexity can be found in Schmid et al. (2011).

2.2 Features of Human Reasoning and Problem Solving

High-level human cognition is *context-dependent*, sometimes *heuristic*, *representation dependent*, and it does not obey pure *normative approaches*, i.e., human inferences deviate from predictions of classical logic and probability theory. Particularities arise from the structure and the limitation of the working memory, the mental representation, and the bounded-rationality. Some features of human reasoning that can be identified are:

- *Inferences can be nonmonotonic*: The suppression effect (cp. Example 5) above, among other findings indicate that inferences humans draw are nonmonotonic (Oaksford and Chater 2007; Johnson-Laird 2006; Stenning and Lambalgen 2008). Small number of defeaters are neglected by participants (Verschueren et al. 2004).

- *Internal representations of propositions and connectives imply partial orders*: Humans generate specific mental representations and neglect others (cp. Example 1, the indeterminacy effect). Partial orders of mental elements are responsible for illusions (cp. Example 7). Open are the number of truth values and how uncertainty is mentally represented, e.g., as possibilities or probabilities as discussed from an AI perspective in Chap. 3 of Volume V1.
- *The role of objects and form of representations can restrict the search space*: Gestalt principles, e.g., the principle of good continuation (Wertheimer 1923) or functional fixation can impact the cognitive processes (cp. Example 12). Humans tend to employ rather qualitative than quantitative representations (Knauff 2013).
- *The drawn inferences depend on different reasoning systems*: In a first step reasoners employ a fast and heuristic reasoning process and then in a second step they reason analytically (Kahneman 2003). This is sometimes termed dual system reasoning.
- *Knowledge influences inferences*: Inferences are not drawn if the given information already seems plausible (e.g., the belief-bias effect, Example 2), or different inferences are drawn between the abstract and deontic version of the Wason Selection Task (cp. Example 6).
- *Inferences deviate from classical normative frames of rationality*: Neither classical logic nor probability theory adequately reflect the inferences (cp. Example 2, 7, 8, 10) and few to none experiments precise the underlying concept of rationality. Concepts like bounded rationality provide better frameworks (Gigerenzer and Selten 2002).
- *Actions can be partially explained by expectations of information gain*: Experiments based on a repeated Wason Selection Task show that participants might select those cards that allow for a higher information gain (Oaksford and Chater 2007).
- *Context, plausibility, and common sense are evaluation criteria of human reasoners*: The Linda problem (cp. Example 7) among others indicates that instead of a pure deductive inference process humans do prefer to take plausibility related reasoning and the interpretation of context into account.

3 The Cognition of Reasoning and Problem Solving

A distinguishing feature of any cognitive theory from a purely logical or AI theory is that the aspect of *cognitive adequacy* is relevant (Strube 1992). The term cognitive adequacy of a theory can be subdivided into (i) *representational adequacy*, i.e., do humans use a similar mental representation as the theory predicts and (ii) *inferential adequacy*, that is, do humans draw the same conclusions as the theory predicts. While most cognitive theories of reasoning focus on the latter, the first is relevant too, as the internal representation influences the kind of inferences that can be drawn.

3.1 Mental Models

The theory of mental models (MMT for short) assumes that for given assertions we construct *iconic models* representing relevant parts of the events or objects based on background knowledge, semantics, and level of expertise (Johnson-Laird 2006). These models represent possibilities and have recently been modeled with modal operators (Hinterecker et al. 2016). Consider the assertion *if it rains then the street is wet*. In classical mathematical logic all possible valuations of *rain* and *wet street* make the conditional true, except for the assignment *rain* to true and *wet street* as false. The mental model theory assumes an order on the interpretations. One factor is the *principle of truth* claiming that human reasoners do not represent what is false, but only what is considered to be true. As Goodwin and Johnson-Laird (2005) explicates, the nature of mental models is *iconic*, meaning that mental models do not represent truth values but humans instead prefer to represent *abstract tokens* in these models. These tokens function as place holders for any kind of events, terms, or objects such as houses, fruits, or mathematical objects (Bara et al. 2001). So the conditional above is represented by the model where both events *rain* and *wet street* co-occur:

$$\begin{array}{cc} \textit{rain} & \textit{wet street} \\ & \ldots \end{array}$$

The ellipsis ("...") represents potential alternative models. These models can be generated in a cognitive "flesh out process" (Johnson-Laird 2006), if necessary. Hence, the mental model theory can also be reckoned among cognitive dual process theories. The ordering of information is a principle that can explain many deviations from logic such as illusions for instance (cp. Example 7). The mental models of assertion represent three possibilities as the initial models for the example above (with empty cells in each line for not fleshed out values of the objects in the column):

$$\begin{array}{ccc} \textit{bread} & & \\ & \textit{soup} & \\ & & \textit{salad} \end{array}$$

These models imply that one cannot have the soup, the salad, or both of them. This is a possible explanation why only 17% of the participants gave the correct answer (Khemlani and Johnson-Laird 2009). Hence, the mental ordering of information is a predictor of the kind of answers participants give. The preferred mental model theory can explain the *indeterminacy effect* and the *preference effect* for spatial reasoning. An analysis of the inference mechanism demonstrates that the mental model theory is "somewhere in between" a credulous and sceptical inference process (Bonnefon 2004), demonstrating nonmonotonic aspects for nonmotonic reasoning problems (Johnson-Laird 2006).

3.2 Models of Cognition Inspired by Nonmonotonic Logics

The specifics of human inferences exclude classical logic as a potential model of human reasoning. This does, however, not necessarily hold for other logics as discussed in chapter "Knowledge Representation: Modalities, Conditionals and Nonmonotonic Reasoning" of Volume 1. Even though logics focus on formalizing correct inferences, cognitive scientists were always inspired by these approaches and aimed to adapt such accounts. In the last years some cognitive scientists have proposed System P (Neves et al. 2002; Benferhat et al. 2005; Pfeifer and Kleiter 2005; Kuhnmünch and Ragni 2014; Ragni et al. 2016), or the Weak Completion Semantics (WCS) based on the the three valued Łukasiewicz logic (Dietz et al. 2012; Hölldobler and Ramli 2009). In addition to the classical truth values of *true* and *false*, the latter two use also the value *unknown*. The WCS guarantees the existence of least models. The general idea is based on introducing an abnormality predicate *ab* in a conditional: Consider the example (3) from above: "If the system *p*asses the Turing test then the system is *i*ntelligent" can be represented as clauses in a logic program as a license for an inference $P = \{i \leftarrow p \wedge \neg ab\}$, with i called the *head* and the $p \wedge \neg ab$ called the *body*. Then, for a given program P, two kinds of transformations are considered: (1) If i is the head of more then one clause then replace all these clauses by $i \leftarrow body_1, \vee \cdots \vee body_m$), then (2) replace all occurrences of $\leftarrow$ by $\leftrightarrow$. The obtained set of formulas is called weak completion of the program. The WCS has been so far applied to the Wason Selection Task, the suppression effect, syllogistic reasoning, and on relational reasoning (Dietz et al. 2015). Other nonmonotonic logics could possibly explain human reasoning, too. Recently, an analysis of many de-facto standards of nonmonotonic logics in explaining human behavior in the suppression effect have been conducted (Ragni et al. 2016). Apart from Łukasiewicz logic none of the nonmonotonic logics like System P, logic programming, Reiter's Default logic, and ranking models could explain the results. Only by a manipulation of background knowledge this changes (Ragni et al. 2017).

3.3 Syntactic Approaches

Syntactic logic approaches, sometimes termed *rule-based* or mental logic theories have been applied to explain human reasoning. The underlying idea is that humans do apply syntactic procedures to derive inferences from given premises. These theories have been applied to reasoning about conditionals, syllogistic reasoning, and relational reasoning (Braine and O'Brien 1998; Rips 1994; Van der Henst 2002). Deviations of human reasoning from classical logical inferences for conditional reasoning can be explained by the derivation length, the working memory capacity, and the kind of inference processes necessary. For example the modus tollens ($A \rightarrow B$ and $\neg B$, then $\neg A$) is more difficult than the modus ponens as in the first case a mental derivation is necessary (comparable to a mathematical proof) and this makes the modus tollens from the perspective of syntactic approaches more difficult than modus ponens. Syntactic rules are extended with additional principles such as Gricean implicature about the quantifiers in syllogistic reasoning. While nowadays the number of proponents of such theories is smaller, combinations of probabilistic and rule based theories have been recently proposed, e.g., Zhai (2015).

3.4 Probabilistic and Heuristic Approaches

Probabilistic theories assume that human reasoning is of probabilistic nature and can best be modeled by theories developed and based on Bayesian inference. Probabilistic theories and, to some extend, heuristic approaches claim that humans do represent uncertainty by assigning probabilities to knowledge. This allows not only to model absolute statements such as *As* are *Bs* (with probability 1), but as well statements including uncertainty such as *probably, As are Bs*. For instance, a conditional *if a then c* is represented by the conditional probability $P(c \mid a)$. For syllogisms, human reasoning is described by the probabilistic heuristic model (PHM; Oaksford and Chater 2007), i.e., the quantifiers can be ordered according to their informativeness from *All*, *Most*, *Few*, *Some*, *None*, to *Some ... not* identified by a computational analysis (Oaksford and Chater 2001). Three heuristics are employed: min-heuristic, p-entailment, and attachment heuristic to derive an answer based on the ordering. Other applications are in modeling inductive learning (Tenenbaum et al. 2006), causal inference (Steyvers et al 2003; Griffiths and Tenenbaum 2005, 2007), language acquisition and processing (Chater and Manning 2006; Xu and Tenenbaum 2007), and semantic memory (Steyvers et al. 2006). The theories assume that humans may conduct a statistical sampling to derive the required knowledge of basic probabilities for events (in particular for singular events). Probabilistic models can explain the aggregated behavior of groups of participants; it has not yet been applied for modeling individual decisions. Probabilistic models have been implemented as Bayesian nets and artificial neural network models (Neal 2012). On the other hand, results demonstrate that a possibilistic approach instead of a probabilistic approach can explain results in some cases more adequate (Raufaste et al. 2003).

Heuristic approaches are often domain-dependent, e.g., the *matching heuristic* for syllogistic reasoning: The most conservative quantifier is preferred in the conclusion and it is given by the following ordering of the quantifiers:

$$\text{No} > \text{Some} \ldots \text{not} = \text{Some} >> \text{All}$$

i.e., if one premise is “No” and the other is “Some” then the conclusion contains the quantifier “Some”. There are two lines of research: One that focuses more on fallacies and limitations due to reasoning about heuristics (Kahneman 2011) and another one illustrating the potential of heuristics for those considered fast-and-frugal heuristics from a computational perspective (Gigerenzer and Selten 2002). Progress has been made by a combined approach of AI and Psychology towards a more general and formally founded theory of decision making based on a rankings (Dubois et al. 2008).

Of course, combinations of different approaches leading to hybrid theories would also be conceivable. Here, as elsewhere in the sense of Occam's razor, a simpler theory should be preferred. These methods have hitherto been demonstrated in the sense of a *proof of concept* that the different theories are able to predict at least some human responses.

3.5 The Cognition of Analogical Reasoning

A prominent model of analogical reasoning (as discussed in chapter "Case-Based Reasoning, Analogical Reasoning, Interpolation" of Volume 1) is the Structure Mapping Engine (SME), which proposes three steps in human analogy making (Gentner 1983; Falkenhainer et al. 1986; Gentner et al. 2001): The first is to *access a target domain*, i.e., identify a source domain similar to the target domain from long-term memory. The second step is to *identify a mapping*, i.e., to identify the relation for individual elements in both domains and generalize them to a general mapping between the domains. The third is to *evaluate and apply* the generalized mapping. The mapping that best fits is the analogy sought.

Core elements of analogies are *objects* and *relations* and the *structural consistency* requires that the relational relationship between elements of the domain must be preserved. Representation is thus related to at most one element of the other domain in the sense of an *injective* mapping. A thorough analysis of the IQ-test Raven's Progressive Matrices yielded two procedural cognitive models (FAIRAVEN and BETTERAVEN; Carpenter et al. 1990) simulating the solution process of human adults with average and above average intelligence, respectively. The models implement six rules that are able to solve the Raven problems. Differences between both models depend on working memory limitations. A cognitive model (Lovett et al. 2009, 2010) based on the computational implementation of the *Process of Structure Mapping* combined with *CogSketch* a sketch understanding tool (Forbus et al. 2008). It uses automatically generated semantic and relational knowledge to successfully solve Raven's *Standard Progressive Matrices* and successfully simulate the answers of human adults (Lovett et al. 2009). Problems that could not be solved were considered difficult for humans (Lovett et al. 2010). A model excluding aspects of human problem solving (Strannegård et al. 2013) is able to solve 28 of the 36 SPM problems (Cirillo and Ström 2010). The program computed the solutions without considering the given possible solutions (Cirillo and Ström 2010). However, the program does not solve arbitrary geometrical problems. A different approach considers a logical view of analogical proportions on the pixel level and is so able to solve 32 out of 36 problems (Correa et al. 2012).

Taken together, there are programs that try to solve the Progressive Matrices in a non-cognitive approach (Evans 1968; Cirillo and Ström 2010) and cognitive models that solve them similarly to humans (Lovett et al. 2010; Carpenter et al. 1990). None of these approaches have been applied to APM and SPM problems at the same time and none of these approaches uses working memory assumptions, makes response time predictions or is generalizable to arbitrary analogical problems (probably besides Lovett et al. 2010).

4 The Architecture of Cognition and Cognitive Models

It should be borne in mind that cognitive theories of reasoning often incorporate only few assumptions about the underlying human working memory and the specific processing of diverse and modality-specific information. We now turn to models of the data structure underlying human cognition. The contrast between the well-defined concepts of Turing machine or λ-calculus in theoretical computer science (Papadimitriou 1994) and the mystery of the human mind makes CS a fascinating discipline. Four core objectives for cognitive theories (Foundation 1978) are: the *abstraction*, i.e., "to formulate abstract descriptions of the mental capacities manifested by the structure, content, and function of various cognitive systems", the *instantiation*, i.e., "the systematic exploration of alternatives as and their realizations in different physical systems", the *plausibility*, i.e., "to characterize the mental processes underlying cognitive function in living organisms", and the *realization*, i.e., "the study of the neurological mechanisms involved in cognition". But it is not only the flow of information, it is the *architecture of information processing* that is vital; neurophysiological findings show that brain damage for example, can drastically alter cognitive abilities (Shallice 1988). This leads to the ultimate goal of CS—to develop a unified theory of cognition on all three of Marr's levels (see below). Many architectures assumed a modality specific information processing, i.e., that certain modules are responsible for the processing of different types of information, e.g., visual information is processed in a different memory location than auditory information and so on. This is supported by findings from neuroscience (Anderson 2007).

4.1 Evaluation Criteria for Cognitive Models

While AI systems aim at a general measure of efficiency or absolute performance as a normative factor, cognitive theories aim at a theory that is both explanatory and predictive for human behavior performance: this often includes accuracy (for a given normative framework), response time, and process steps. However, cognitive models are never just *simulation models* that can reproduce only existing experimental data. A good cognitive model is largely independent of experimental data and has general strategies from which the experimental data (response times and given responses) can be generated. In summary, requirements of cognitive models can be determined by several criteria (Ragni 2008):

- *Transparency of the underlying model assumptions and the role of parameters.* If the assumptions of the model are opaque the predictions cannot be related back to the model processes.
- *Independence of the modeling principles of experimental data*. A model is never just a post-hoc explanation of experimental data, but relates responses to general cognitive processes.
- *Coverage of relevant phenomena*. The degree of coverage may be further specified in accuracy, response time, and intermediate step correspondence (Simon and Wallach 1999). If applicable, models can be additionally compared by information criteria (see below).
- *Capability to represent and explain inter-individual differences*. Inter-individual differences appear in reasoning and can be traced back, e.g., to different working memory sizes, knowledge, and concepts.
- *Generalizability and predictability of the cognitive model*. This includes a possible domain-independence and whether new and so far untested predictions about cognitive phenomena can be predicted.

While current approaches in AI are benchmarked against some current test problems (e.g., in planning or in theorem proving), the benchmarks of cognitive theories and models can differ across modeling approaches as we will see.

4.2 The Architecture of High-Level Human Cognition

Cognitive modeling can be understood as an algorithmization of psychological theories in a cognitive architecture. The ultimate goal is to move beyond the reproduction of empirical results (reaction times and error rates) and to obtain an equivalency on the process level on activations on the brain level. The modeling process is iterative and takes place in at least three steps: First, certain psychological phenomena or effects are identified. These are then explained and reproduced by a cognitive model in a second step. In a third step, new, not yet empirically tested model predictions are tested experimentally. Information processing in the human mind can be described on at least three levels: "the first level, known as the *computation level* abstractly represents the characteristics and objectives specified in the problem (Marr 1982). The second level, the *algorithmic level*, indicates how this calculation is implemented using algorithms. The third level, the *implementation level*, reflects the biological realization, i.e., the neuronal implementation. These three levels are also called semantic level, syntactic level, and physical level" (Marr 1982; McClamrock 1991). Cognitive theories have been developed for all levels and recently new approaches aim at hybrid approaches (e.g., ACT-R Anderson 2007) incorporating symbolic and subsymbolic processes. A variety of approaches were inspired by the architecture of computers with a Central Process Unit (*CPU*) and a short term memory. Other possibilites are *hierarchical architectures*, the control of the entire processing being ensured by a special module (*supervisor*) or by *production control systems* (Anderson 2007; Sun 2001; Laird 2012), the direct control performed by an interpreter by a central data structure comprising various modules for cognition-specific tasks (*working memory*). Two different methods in modeling can be distinguished: For *top-down* processes knowledge, abilities, or reflection drives the behavior. In contrast, *bottom-up* processes immediately start at the level of perception and stimuli from the environment. Most actions are based on the interaction of both types of cognitive processes.

4.2.1 Cognitive Architectures

Most cognitive architectures are inspired by the General Problem Solver (GPS, Newell and Simon 1972), a model that uses means-end analysis as a search heuristic. It has been reimplemented as a production rule system. Production rule systems realize the physical symbol system hypothesis. These systems are composed of (i) *production rules*; they consist of a *condition* part and an *action* part, an (ii) *interpreter* that checks, if conditions of existing production rules are satisfied in a given model's state (they can *fire*). In the event that several rules can fire a conflict resolution process starts. Architectures often specify additional data structures. We focus on two architectures with most published (cognitive) models: an AI oriented approach SOAR and the hybrid cognitive architecture ACT-R.

SOAR (States, Operators, And Reasoning) is a production rule system with reinforcement learning built upon the GPS, hence aiming at a general problem solving agent (Newell 1990). It uses a problem space representation by differentiating between different forms of knowledge, e.g., procedural and semantic knowledge, and a distinctive working and long-term memory. Its emphasis lies on applying learning on all levels, hence implementing all AI and cognitive learning principles. The current version SOAR 9 integrates non-symbolic representations and other learning mechanisms (Laird 2008, 2012). It is a responsive system, i.e., each decision depends on the sensory input, the state of the working memory and encoded knowledge in the long-term memory. It performs a variety of problems from planning, robotic systems, interactions with virtual humans, and an air combat simulation for pilot training at the USAF (Tambe et al. 1995).

The cognitive architecture ACT-R 7.0 (Anderson 2007) aims at a unified human cognition approach. It is a hybrid theory, consisting of symbolic and subsymbolic parts. Its data structure is oriented on modality specific knowledge modules for perception (e.g., *visual, aural*), goal and sub-goal representations (*goal, imaginal*) and interfaces (so-called *buffers*) which can be accessed by production rules. ACT-R uses chunks as the atomic knowledge representation format with procedural knowledge encoded in production rules and declarative knowledge that uses the concept of activation. Cognitive models have been developed for learning and memory, problem solving, deductive reasoning, perception, attention control, and human-computer interaction (HCI). Recently, ACT-R allows for the prediction of task-specific brain activations allowing modeling of findings from fMRI research.

A limitation and point of criticism is that many *cognitive architectures* are Turing complete (Anderson 1983) and, hence, do not provide cognitive bottlenecks or other architectured based constraints on computation processes of cognition.

4.2.2 Models Based on Artificial Neurons

A different modeling approach does not focus on symbols as the atomic components but on artificial neural models which are a simplification of brain neurons with a focus on electrical excitation while neglecting neurotransmitters or hormonal activity. Logical and arithmetic functions can be calculated by such artificial neural networks (ANNs) (McCulloch and Pitts 1943). The Hebbian learning rule (Hebb 1949) realizes the strengthening of the connection between two neurons when both neurons are active at the same time. First ANNs had two layers of nodes (*perceptron*), but the limitation to not represent the boolean operator XOR lead to the development of multilayer models using *backpropagation* (McClelland and Rogers 2003) or recurrent networks (Hölldobler and Ramli 2009). Most connectionists, proponents of ANNs, regard ANNs as “calculation models and not as models of biological reality” (Smolensky 1988). A recent approach, NEF (neural engineering framework), is based on biologically inspired spiking neural networks. It is build upon three principles that cover the nonlinear encoding and linear decoding for representations and transformations as dynamic systems (Eliasmith 2013). An artificial “brain” called SPAUN, is built based on NEF, and consists of about 2 million simulated neurons that cover different brain regions like the posterior parietal cortex, the prefrontal cortex, and occipital cortex, and the basal ganglia that have a similar role as the production rule mapping in ACT-R for distributing tasks to the specific brain areas. SPAUN is capable of solving tasks in high-level cognition from Raven’s Progressive Matrices to serial working memory among others (Stewart et al. 2012).

4.2.3 Bayesian Modeling and Quantum Models

The starting point and the basic idea of Bayesian cognitive models is the question of how a cognitive agent revises its current assumptions in the light of observed data. In principle these models assume that an agent has degrees of beliefs, hypotheses that can be represented by probability distributions, and that the agent updates its belief distributions on new evidence according to Bayes’ rule. However, some Bayesian modelers claim that “the human mind learns and concludes according to Bayesian principles is not the assertion that the human mind implements Bayesian inferences" (Tenenbaum et al. 2011). In contrast to the symbolic and hybrid models presented above, Bayesian models do not aim at being cognitive process models. For example, cognitive restrictions of the *cognitive bottlenecks* are not relevant. Instead these models are realizations of a method called *rational analysis* (Anderson 2007). A "large number of known connectivist algorithms have a Bayesian interpretation" (Griffiths and Tenenbaum 2011), and these could serve as a first approach to a neural modeling of Bayesian approaches. Causal Bayesian networks allow the representation of structural aspects between random variables (Pearl 2000) and have been used for cognitive models. The causal structure, that is, the dependencies of the random variables, is represented by an acyclic directed graph. The directional graphs represent the relationship between cause and effects. Such causal Bayesian networks have not only a purely probabilistic relationship between the variables, but also a causal structure with implications on the statistical data, which can then be checked by empirical data (Hagmayer and Waldmann 2006). The strength is, in particular, to present a modeling approach on the computational level which is robust enough against noisy data. Application areas cover all domains of reasoning and such models are dominant in the area of decision making. Quantum probability models (Busemeyer and Bruza 2012) are a recent modeling approach in the field of decision making that extends the classical Bayesian approaches to model some paradoxes such as the sure thing principle (cp. Example 10) (Pothos and Busemeyer 2009).

4.2.4 Multinomial Processing Tree Models

Multinomial processing tree (MPT) are a class of models described by directed acyclic graphs, where each inner node represents a cognitive state, the leaves represent possible responses of participant(s), and the edges represent transition probabilities for each processing step. Hence, MPTs aim at explaining the generated output via the underlying latent cognitive processes. MPTs are a tool to compare theories (Oberauer 2006). Typical statistical measures for the goodness-of-fit are information criteria (e.g., Bayesian Information Criteria) that punish overly complex models. A limitation in contrast to process models is that the nodes are not algorithmically specified but that the sequences of the processes are predominant. As a result, models for reasoning do not necessarily have implications on the required working memory capacity and do not pose modelling constraints like cognitive bottlenecks. Applications of MPT-models are in models of recognition memory, decision making, and conditional and syllogistic reasoning, e.g., the belief-bias effect (cp. Example 2) (Klauer et al. 2000).

Cognitive modeling can be pure *symbolic*, *connectionistic*, or *hybrid*, and may include one or more Marrs' levels. It can integrate specific assumptions about the structure of the working memory, individual behavior, or likelihood-based prediction of the behavior of a group. The various approaches reflect different areas of human cognition, and offer the possibility of the empirical falsifiability of cognitive theories as opposed to pure descriptive theories.

5 Challenges in High-Level Cognition Research

Despite great progress in the field of AI, the construction of machines that fully implement high-level cognition has not yet been achieved. Human thinking is not simply the realization of the laws of classical logic, but clearly differs and is more multifaceted. Humans, in contrast to current cognitive and artificial systems, have an impressive ability to deal with underspecified data and imprecise knowledge and to solve problems by insight. The purpose of this line of research is that by understanding and modeling human thinking and reasoning, we can learn about feasible techniques that can be implemented in systems to deal with imprecise and complex problem descriptions.

1. **What are relevant benchmark problems?** The fields of action planning and automatic theorem proving in AI have greatly benefited from well defined benchmark problems and annual competitions. This made a fair comparison between different approaches and systems possible and triggered a competitive spirit to improve the state-of-the-art of the fields and to incorporate new concepts. We see the necessity to have competitions in the field of human reasoning as well, as the number of cognitive theories that argue to explain parts of human reasoning is continuously increasing, but few comparisons on common data sets exist. While psychological experiments can provide such benchmark problems, some findings are more important than others. So far there are no criteria identified that can be applied to identify *relevant* problems but this is a necessary condition to develop a generally accepted benchmark.
2. **How do humans represent and process information in high-level cognition?** A recent study in syllogistic reasoning demonstrated that any of the main cognitive theories deviates significantly from the empirical data. This demonstrates that even for reasoning about quantified assertions the underlying representation is still open. A current limiting factor is that there is no common language to formalize different representations and no precise well-defined benchmark. Additionally, many cognitive theories are underspecified. Hence, the question is how can these descriptive theories be turned into appropriately implemented models?
3. **How to model insight processes and meaning?** Central to many processes in high-level cognition is the ability to gain insight and assign meaning. At the same time these processes are hard to formalize or algorithmize. How is meaning generated and how can it be implemented?
4. **What are necessary features of "good" cognitive models?** While there are several definitions of cognitive models none is formally specified. Without a clear definition what is accepted as a cognitive model limits the search for better fitting models in the space of all cognitive models. Two steps can improve cognitive modeling:

   - The turn to *predictive cognitive models*: Current models are often post-dictive in contrast to predictive models, but only the capability to predict new phenomena extends current limitations and aims at a general and unified cognitive modeling approach.

- The turn to *cognitive models for individual reasoners*: Many cognitive models aim at modeling an average human reasoner by an aggregation of the data of individual participants. However, aggregation inserts noise and blurs the cognitive processes of each individual reasoner. Moreover, any cognitive model that adequately models individual reasoner can model groups.

5. **How to develop cognitive models that are domain-independent?** Most AI and cognitive systems are specialized for respective domains with some recent exceptions in the field of modeling like SOAR or NEF. In contrast human reasoning is not limited to one domain. What are necessary features of such general and unified cognitive models?
6. **What are necessary properties of cognitive architectures?** There already exists a broad variety of cognitive architectures (Kotseruba et al. 2016), of which many even perform comparably. But the general foundations of such architectures are not specified, formalized, or compared.

These challenges are on a foundational level and demonstrate that many questions are open, even after decades of research. The field of cognitive modeling can strongly benefit from the rigorous formal approaches from AI.

6 Conclusion

AI aims at improving systems efficiently to find optimal solutions or at least good approximations. In contrast, CS concentrates more on *modeling* the mental processes underlying human behavior. The level of modeling comprises understanding the information theoretic processes on an algorithmic level and aligning it with neural activity. Yet the differences between AI and CS should not be mistaken for a disadvantage. From the separate viewpoints of AI and CS emerges a fertilization process. Humans can *adapt* themselves to new domains and solve problems by *insights*, two high-level cognition phenomena that could improve current AI systems. To improve AI systems we require cognitively adequate frameworks that are suitable for representing information and have good computational properties at the same time, i.e., that solutions can be computed in a reasonable time. High-level cognition is not a black-box, the performance of human reasoning and problem solving can be analyzed, reproduced, and predicted. This requires, however, a multi-disciplinary approach covering psychological experiments, formal and cognitive modeling, and logics. Understanding cognition is often a *reverse engineering problem*, i.e., it is necessary to reconstruct the underlying functions from behavioral findings such as error rates and reaction times. At the same time cognitive models provide an important and interesting bridge between formal methods and empirical psychological results. They offer the possibility to formalize psychological theories, even to produce human-like error rates and reaction times and finally to compare these results with predictions of psychological theories. Recently, this has lead to a stronger interest in formal methods in the psychology of reasoning (Bonnefon 2013). Recent approaches do focus more on human reasoning processes about preferences and behavior of other agents (Bonnefon et al. 2012).

By analyzing the specific features of cognitive architectures it is possible to integrate all models into a general system based on Newell's idea of a "unified theory of cognition" (Newell 1994). Such a unified theory of cognition should offer a small or even single set of mechanisms that can account for human performance on cognitive tasks from perception to problem solving. Most research is performed on a normative scale without reflecting the underlying premises. This leads to the impression that human reasoning is "weaker" or "erroneous" in contrast to formal methods from AI or logic. But, classical logic cannot always be applied, it requires specific properties and has its limitations if applied in the wrong context, e.g., in a nonmonotonic world. In this sense human reasoning that is nonmonotonic, inductive, plausible, context-dependent, integrating different reasoning systems has adapted itself to reasoning efficiently and satisficing with respect to bounded rationality and it is adaptable to different domains. These properties still make human reasoning interesting for developing better AI systems.

Acknowledgements The author is grateful to Bernhard Nebel (University of Freiburg), Emmanuelle-Anna Dietz Saldanha (TU Dresden), and Nicolas Riesterer (University of Freiburg) for their substantial feedback.

References

Anderson JR (1983) The architecture of cognition. Hillsdale, NJ, US

Anderson JR (2007) How can the human mind occur in the physical universe? Oxford University Press, New York

Bara BG, Bucciarelli M, Lombardo V (2001) Model theory of deduction: a unified computational approach. Cogn Sci 25:839–901

Baral C (2003) Knowledge representation, reasoning and declarative problem solving. University Press, Cambridge

Bauer MI, Johnson-Laird PN (1993) How diagrams can improve reasoning. Psychol Sci 4:372–378

Beller S, Spada H (2001) Denken. In: Strube G (ed) Wörterbuch der Kognitionswissenschaft. Klett-Cotta, Stuttgart

Benferhat S, Bonnefon J, Neves RDS (2005) An overview of possibilistic handling of default reasoning, with experimental studies. Synthese 146(1–2):53–70

Bennati S, Brüssow S, Ragni M, Konieczny L (2014) Gestalt effects in planning: rush-hour as an example. In: Bello P, Guarini M, McShane M, Scassellati B (eds) Proceedings of the 36th annual conference of the cognitive science society. Cognitive Science Society, Austin, TX, pp 1234–1240

Bonnefon J (2004) Reinstatement, floating conclusions, and the credulity of mental model reasoning. Cogn Sci 28(4):621–631

Bonnefon J (2013) Formal models of reasoning in cognitive psychology. Argum Comput 4(1):1–3. https://doi.org/10.1080/19462166.2013.767559

Bonnefon J, Girotto V, Legrenzi P (2012) The psychology of reasoning about preferences and unconsequential decisions. Synthese 185(Supplement-1):27–41

Braine MDS, O'Brien DP (1998) Mental logic. Erlbaum, Mahwah
Busemeyer JR, Bruza PD (2012) Quantum models of cognition and decision. Cambridge University Press, Cambridge
Byrne RMJ (1989) Suppressing valid inferences with conditionals. Cognition 31:61–83
Byrne RMJ (2002) Mental models and counterfactual thoughts about what might have been. Trends Cogn Sci 6(10):426–431
Byrne RMJ (2007) The rational imagination: how people create alternatives to reality. MIT Press, Cambridge
Carpenter PA, Just MA, Shell P (1990) What one intelligence test measures: a theoretical account of the processing in the raven progressive matrices test. Psychol Rev 97(3):404–431
Chater N, Manning CD (2006) Probabilistic models of language processing and acquisition. Trends Cogn Sci 10:335–344
Chu Y, MacGregor JN (2011) Human Performance on insight problem solving: a review. J Probl Solving 3(2):119–150
Cirillo S, Ström V (2010) An anthropomorphic solver for raven's progressive matrices. Master's thesis, Chalmers University of Technology, Department of Applied Information Technology, SE-41296 Goeteborg, Sweden
Correa W, Prade H, Richard G (2012) When intelligence is just a matter of copying. In: Proceedings of the 20th European conference on artificial intelligence. IOS Press, Amsterdam, The Netherlands, ECAI'12, pp 276–281
Cronin M, Gonzalez C, Sterman JD (2009) Why don't well-educated adults understand accumulation? a challenge to researchers, educators and citizens. Organ Behav Hum Decis Process 23(1):108
Dietz EA, Hölldobler S, Ragni M (2012) A computational approach to the suppression task. In: Miyake N, Peebles D, Cooper RP (eds) Proceedings of the 34th annual conference of the cognitive science society. Cognitive science society, Austin, TX, pp 1500–1505
Dietz EA, Hölldobler S, Höps R (2015) A computational logic approach to human spatial reasoning. In: 2015 IEEE symposium series on computational intelligence. IEEE, pp 1627–1634
Dixon JE, Byrne RMJ (2011) "if only" counterfactual thoughts about exceptional actions. Mem Cogn 39(7):1317–1331
Dörner D, Kreuzig HW, Reither F, Stäudel T (1983) Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität, Huber, Bern
Douven I (2011) Abduction. In: Zalta EN (ed) The stanford encyclopedia of philosophy
Dubois D, Fargier H, Bonnefon J (2008) On the qualitative comparison of decisions having positive and negative features. J Artif Intell Res 32:385–417
Duncker K (1945) On problem-solving. Psychological Monographs ix(58):113
Eliasmith C (2013) How to build a brain: a neural architecture for biological cognition. Oxford University Press, Oxford
Evans TG (1968) A program for the solution of a class of geometric-analogy intelligence-test questions. In: Minsky ML (ed) Semantic information processing. MIT Press, Cambridge, MA, chap 5, pp 271–351
Falkenhainer B, Forbus KD, Gentner D (1986) The structure-mapping engine. Report Department of Computer Science, University of Illinois at Urbana-Champaign
Fauconnier G, Turner M (2008) The way we think: conceptual blending and the mind's hidden complexities. Basic Books, New York
Flake GW, Baum EB (2002) Rush Hour is PSPACE-complete, or why you should generously tip parking lot attendant. Theor Comput Sci 270:895–911
Forbus K, Usher J, Lovett A, Lockwood K, Wetzel J (2008) CogSketch: open-domain sketch understanding for cognitive science research and for education. In: Alvarado C, Cani MP (eds) Proceedings of the fifth eurographics workshop on sketch-based interfaces and modeling
Foundation S (1978) Cognitive science 1978: Report of the state of the art committee. http://csjarchive.cogsci.rpi.edu/misc/CognitiveScience1978_OCR.pdf

Frensch PA, Funke J (eds) (1995) Complex problem solving: the European perspective. Lawrence Erlbaum, Hillsdale, NJ
Funke J (2006) Lösen komplexer Probleme. In: Funke J, Frensch P (eds) Handbuch der Allgemeinen Psychologie - Kognition. Handbuch der Psychologie, Hogrefe, Göttingen, pp 439–445
Gentner D (1983) Structure-mapping: a theoretical framework for analogy. Cogn Psychol 7:155–170
Gentner D, Holyoak KJ, Kokinov BN (2001) The analogical mind: perspectives from cognitive science. Bradford Books, MIT Press, Cambridge, MA
Gigerenzer G, Selten R (2002) Bounded rationality: the adaptive toolbox. MIT Press, Cambridge
Gilhooly KJ, Murphy P (2005) Differentiating insight from non-insight problems. Think Reason 11(3):279–302
Goodwin GP, Johnson-Laird PN (2005) Reasoning about relations. Psychol Rev 112:468–493
Greeno JG, Simon HA (1988) Problem solving and reasoning. Technical report, University of Pittsburgh
Griffiths TL, Tenenbaum JB (2005) Structure and strength in causal induction. Cogn Psychol 51:354–384
Griffiths TL, Tenenbaum JB (2007) From mere coincidences to meaningful discoveries. Cognition 103(2):180–226
Griffiths TL, Tenenbaum JB (2011) Predicting the future as Bayesian inference: people combine prior knowledge with observations when estimating duration and extent. J Exp Psychol: Gen 140(4):725
Hagmayer Y, Waldmann MR (2006) Kausales Denken. In: Funke J (ed) Enzyklopädie der Psychologie 'Denken und Problemlösen', no. 8 in C, Hogrefe, Göttingen, chap II, pp 87–166
Halbmayer E , Salat J (2012) Qualitative Methoden der Kultur- und Sozialanthropologie. http://www.univie.ac.at/ksa/elearning/cp/qualitative/qualitative-5.html
Halford GS, Wilson WH, Phillips S (2010) Relational knowledge: the foundation of higher cognition. Trends Cogn Sci 14:497–505
Hebb DO (1949) The organization of behavior. Wiley, New York
Hinterecker T, Knauff M, Johnson-Laird PN (2016) Modality, probability, and mental models. J Exp Psychol: Learn Mem Cogn 42(10):1606–1620
Hölldobler S, Ramli CDPK (2009) Logic programs under three-valued Łukasiewicz semantics. In: Logic programming. Springer, Berlin, pp 464–478
Hollnagel E, Woods DD (2005) Joint cognitive systems: foundations of cognitive systems engineering. CRC Press, Boca Raton
Johnson-Laird PN (2006) How we reason. Oxford University Press, New York
Just MA, Carpenter PA, Varma S (1999) Computational modeling of high-level cognition and brain function. Hum Brain Mapp 8:128–136
Kahneman D (2011) Thinking, fast and slow. Macmillan, New York
Kahneman D (2003) A perspective on judgement and choice. Am Psychol 58:697–720
Kaller CP, Unterrainer JM, Rahm B, Halsband U (2004) The impact of problem structure on planning: insights from the Tower of London task. Cogn Brain Res 20:462–72
Khemlani S, Johnson-Laird PN (2009) Disjunctive illusory inferences and how to eliminate them. Mem Cogn 37:615–623
Khemlani S, Johnson-Laird PN (2012) Theories of the syllogism: a meta-analysis. Psychol Bull
Khemlani S, Johnson-Laird PN (2013) The processes of inference. Argum Comput 4(1):4–20
Khemlani S, Johnson-Laird PN (2016) How people differ in syllogistic reasoning. In: Proceedings of the 36th annual conference of the cognitive science society. Cognitive Science Society, Austin, TX
Klauer KC, Musch J, Naumer B (2000) On belief bias in syllogistic reasoning. Psychol Rev 107(4):852–884
Knauff M (2013) Space to reason: a spatial theory of human thought. MIT Press, Cambridge
Knauff M, Johnson-Laird PN (2002) Visual imagery can impede reasoning. Mem Cogn 30:363–71

Kotseruba I, Gonzalez OJA, Tsotsos JK (2016) A review of 40 years of cognitive architecture research: focus on perception, attention, learning and applications. arXiv:161008602
Kuhnmünch G, Ragni M (2014) Can formal non-monotonic systems properly describe human reasoning? In: Bello P, Guarini M, McShane M, Scassellati B (eds) Proceedings of the 36th annual conference of the cognitive science society, Austin, TX, pp 1222–1228
Laird JE (2008) Extending the SOAR cognitive architecture. In: AGI, pp 224–235
Laird JE (2012) The SOAR cognitive architecture. The MIT Press, Cambridge
Legrenzi P, Girotto V, Johnson-Laird PN (1993) Focussing in reasoning and decision making. Cognition 49(1):37–66
lle Lépine R, Parrouillet P, Camos V, (2005) What makes working memory spans so predictive of high-level cognition? Psychon Bull Rev 12(1):165–170
Lovett A, Tomai E, Forbus K, Usher J (2009) Solving geometric analogy problems through two-stage analogical mapping. Cogn Sci 33(7):1192–1231
Lovett A, Forbus K, Usher J (2010) A structure-mapping model of raven's progressive matrices. In: Ohlsson S, Catrambone R (eds) Proceedings of the 32nd annual conference of the cognitive science society. Cognitive Science Society, pp 2761–2766
Marr D (1982) Vision. A computational investigation into the human representation and processing of visual information. Freeman, San Francisco
McClamrock R (1991) Marr's three levels: a re-evaluation. Minds Mach 1(2):185–196
McClelland JL, Rogers TT (2003) The parallel distributed processing approach to semantic cognition. Nat Rev Neurosci 4(4):310
McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133
Neal RM (2012) Bayesian learning for neural networks, vol 118. Springer, Berlin, Heidelberg
Neves RDS, Bonnefon J, Raufaste E (2002) An empirical test of patterns for nonmonotonic inference. Ann Math Artif Intell 34(1–3):107–130
Newell A (1979) Reasoning, problem solving and decision processes: the problem space as a fundamental category. Technical report, Computer science department CMU
Newell A (1990) Unified theories of cognition. Harvard University Press, Cambridge
Newell A (1994) Unified theories of cognition. The William James lectures. Harvard University Press, Cambridge
Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs
Newell A, Simon HA (1976) Computer science as empirical enquiry: symbols and search. Commun ACM 19:113–126
Newstead SE (1995) Gricean implicatures and syllogistic reasoning. J Mem Lang 34:644–664
Oaksford M, Chater N (2001) The probabilistic approach to human reasoning. Trends Cogn Sci 5(8):349–357
Oaksford M, Chater N (2007) Bayesian rationality: the probabilistic approach to human reasoning. Oxford University Press, Oxford
Oberauer K (2006) Reasoning with conditionals: a test of formal models of four theories. Cogn Psychol 53:238–283
O'Reilly RC (2006) Biologically based computational models of high-level cognition. Science 314(5796):91–94
Papadimitriou CM (1994) Computational complexity. Addison-Wesley, Reading
Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press, Cambridge
Pfeifer N, Kleiter GD (2005) Coherence and nonmonotonicity in human reasoning. Synthese 146:93–109
Pothos EM, Busemeyer JR (2009) A quantum probability explanation for violations of 'rational' decision theory. Proc R Soc Lond B: Biol Sci
Potts GR (1974) Storing and retrieving information about ordered relationships. J Exp Psychol 103(3):431

Ragni M (2008) Räumliche Repräsentation, Komplexität und Deduktion: Eine kognitive Komplexitätstheorie[Spatial representation, complexity and deduction: A cognitive theory of complexity]. Ph.D. thesis, Albert-Ludwigs-Universität Freiburg

Ragni M, Knauff M (2013) A theory and a computational model of spatial reasoning with preferred mental models. Psychol Rev 120(3):561–588

Ragni M, Kola I, Johnson-Laird P (2018) On selecting evidence to test hypotheses: a theory of selection tasks. Psychol Bull 144(8):779–796

Ragni M, Eichhorn C, Kern-Isberner G (2016) Simulating human inferences in the light of new information: a formal analysis. In: Kambhampati S (ed) Proceedings of the twenty-fifth international joint conference on artificial intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016, IJCAI/AAAI Press, pp 2604–2610. http://www.ijcai.org/Proceedings/2016

Ragni M, Eichhorn C, Bock T, Kern-Isberner G, Tse APP (2017) Formal nonmonotonic theories and properties of human defeasible reasoning. Minds Mach 27(1):79–117

Raufaste E, Neves RDS, Mariné C (2003) Testing the descriptive validity of possibility theory in human judgments of uncertainty. Artif Intell 148(1–2):197–218

Raven JC (1962) Advanced progressive matrices, Set II. H. K. Lewis, London

Raven J, Raven JC, Court JH (2000) Manual for ravens progressive matrices and vocabulary scales. Harcourt Assessment, San Antonio, TX

Rips LJ (1994) The psychology of proof: deductive reasoning in human thinking. The MIT Press, Cambridge

Russell S, Norvig P (2003) Artificial intelligence: a modern approach, 2nd edn. Prentice Hall, Englewood Cliffs

Schmid U, Ragni M, Gonzalez C, Funke J (2011) The challenge of complexity for cognitive systems. Cogn Syst Res 12(3–4):211–218

Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–424

Searle JR (2004) Mind: a brief introduction. Oxford University Press, Oxford

Shallice T (1988) From neuropsychology to mental structure. Cambridge University Press, Cambridge

Simon HA, Newell A (1962) Computer simulation of human thinking and problem solving, vol 27 [Society for Research in Child Development, Wiley], pp 137–150

Simon HA, Wallach D (1999) Editorial: cognitive modeling in perspective. Kognitionswissenschaft 8:1–4

Smolensky P (1988) On the proper treatment of connectionism. Behav Brain Sci 11:1–74

Stenning K, Lambalgen M (2008) Human reasoning and cognitive science. Bradford Books, MIT Press, Cambridge

Sternberg RJ (1980) Reasoning, problem solving, and intelligence. Technical report, DTIC Document

Sternberg RJ (1999) The nature of cognition. A Bradford Book, MIT Press, Cambridge

Stewart T, Choo FX, Eliasmith C (2012) Spaun: a perception-cognition-action model using spiking neurons. In: Proceedings of the cognitive science society, vol 34

Steyvers M, Tenenbaum JB, Wagenmakers EJ, Blum B (2003) Inferring causal networks from observations and interventions. Cogn Sci 27(3):453–489. http://dblp.uni-trier.de/db/journals/cogsci/cogsci27.html#SteyversTWB03

Steyvers M, Griffiths TL, Dennis S (2006) Probabilistic inference in human semantic memory. Trends Cogn Sci 10:327–334

Strannegård C, Cirillo S, Ström V (2013) An anthropomorphic method for progressive matrix problems. Cogn Syst Res 22–23:35–46

Strube G (1992) The role of cognitive science in knowledge engineering. In: Contemporary knowledge engineering and cognition, pp 159–174

Strube G, Ferstl E, Konieczny L, Ragni M (2013) Kognition. In: Görz G, Schneeberger J, Schmid U (eds) Handbuch der Künstlichen Intelligenz. Oldenbourg, München

Sun R (2001) Duality of the mind - a bottom-up approach toward cognition. Lawrence Erlbaum

Szymanik J (2016) Quantifiers and cognition: logical and computational perspectives. Springer, Berlin

Tambe M, Johnson WL, Jones RM, Koss FV, Laird JE, Rosenbloom PS, Schwamb K (1995) Intelligent agents for interactive simulation environments. AI Mag 16(1):15–39

Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based Bayesian models of inductive learning and reasoning. Trends Cogn Sci 10:309–318

Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND (2011) How to grow a mind: statistics, structure, and abstraction. Science 331(6022):1279–1285

Tversky A, Kahneman D (1973) Availability: a heuristic for judging frequency and probability. Cogn Psychol 5(2):207–232

Tversky A, Kahneman D (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol Rev 90(4):293

Van der Henst JB (2002) Mental model theory versus the inference rule approach in relational reasoning. Think Reason 8(3):193–203

Verschueren N, Schaeken W, De Neys W, d'Ydewalle G (2004) The difference between generating counterexamples and using them during reasoning. Q J Exp Psychol Sect A 57(7):1285–1308

Wason PC (1971) Problem solving and reasoning. Br Med Bull 27(3):206–210

Wertheimer M (1923) A brief introduction to Gestalt, identifying key theories and principles. Psychol Forsch 4:301–350

Xu F, Tenenbaum JB (2007) Word learning as bayesian inference. Psychol Rev 114(2):245–72

Zhai F (2015) Toward probabilistic natural logic for syllogistic reasoning. Ph.D. thesis, Universiteit van Amsterdam