

# Reasoning as simulation

Nicholas L. Cassimatis · Arthi Murugesan ·  
Perrin G. Bignoli

Received: 17 September 2008 / Accepted: 29 January 2009  
© Marta Olivetti Belardinelli and Springer-Verlag 2009

**Abstract** The theory that human cognition proceeds through mental simulations, if true, would provide a parsimonious explanation of how the mechanisms of reasoning and problem solving integrate with and develop from mechanisms underlying forms of cognition that occur earlier in evolution and development. However, questions remain about whether simulation mechanisms are powerful enough to exhibit human-level reasoning and inference. In order to investigate this issue, we show that it is possible to characterize some of the most powerful modern artificial intelligence algorithms for logical and probabilistic inference as methods of simulating alternate states of the world. We show that a set of specific human perceptual mechanisms, even if not implemented using mechanisms described in artificial intelligence, can nevertheless perform the same operations as those algorithms. Although this result does not demonstrate that simulation theory is true, it does show that whatever mechanisms underlie perception have at least as much power to explain non-perceptual human reasoning and problem solving as some of the most powerful known algorithms.

**Keywords** Perceptual simulation · Reasoning · Inference

## Introduction

A full explanation of human cognition must describe the mechanisms that enable humans to reason, solve problems and converse in such a wide range of complex scenarios

while explaining how these mechanisms developed from and integrate with mechanisms for perceiving and acting within the narrower range of simpler situations children and our evolutionary ancestors dealt with. An obstacle to achieving this goal is the apparent differences between computational methods for modeling reasoning and inference and those used to model processes such as perception, attention, imagery, memory and motor action. In this paper, we suggest that the difference is only apparent and that there is a deep correspondence between the most powerful artificial intelligence reasoning algorithms and the mechanisms of human perception and imagery.

On some accounts (e.g., Barsalou 1999), human cognition proceeds through simulations performed by the mechanisms of perception. Such “simulation theories” are attractive for several reasons. If true, they would explain how perceptual mechanisms are integrated with the rest of cognition. They would also eliminate the need to explain the evolution of a new set of mechanisms specifically for cognition unique to humans (e.g., complex reasoning and language use). Finally, simulation theories would make the large body of knowledge about perception relevant to our understanding of inference and vice versa.

One potential problem for simulation theories is that perceptual mechanisms do not obviously have the representational and computational power to make inferences<sup>1</sup> in situations as complex as people can. Although there is increasing evidence (e.g., Barsalou et al. 2003; Boroditsky and Ramscar 2002; Richardson et al. 2003) that in *some* situations perceptual mechanisms are involved in reasoning

<sup>1</sup> We refer to inference and reasoning interchangeably as the process by which people come to believe or suspect new facts given existing facts. We intend this broadly to include, for example, subconscious and automatic inference in perception and language understanding in addition to cognition studied in the psychology of reasoning.

---

N. L. Cassimatis (✉) · A. Murugesan · P. G. Bignoli  
Rensselaer Polytechnic Institute, Troy, NY, USA  
e-mail: cassin@rpi.edu

about more abstract relations, this does not prove that perceptual mechanisms are powerful enough to explain *all* human inference. Further, much of the discussion on the computational power of simulation theory pertains mostly to its representational power. There is little work addressing whether simulation mechanisms have the computational power to explain the full range of human intelligence.

While it is common to conceive of the perceptual system's operation as solving inference problems, the computational methods used to address them are very different from those used to study much non-perceptual cognition. For example, the Davis–Putnam–Logemann–Loveland (DPLL) algorithm (Davis et al. 1962; Davis and Putnam 1960) and its modern variants (e.g., Eén and Sorensson 2005; Heras et al. 2008; Hoos and Stützle 2002; Marques-Silva and Sakallah 1996) are some of the most powerful and widely used reasoning algorithms in artificial intelligence. They are often used to perform many kinds of non-perceptual problem solving and inference tasks for which visual processing algorithms seem not to be well suited. Does this mean that human perceptual mechanisms are not powerful enough for non-perceptual reasoning and problem solving?

In this paper, we demonstrate that simulation theories are more closely related to modern AI algorithms than is commonly thought and that their computational power is at least as strong as other mechanisms posited by cognitive modelers. Specifically, we show how the sequence of operations produced by the interaction of some basic mechanisms of human perception and imagery is under some circumstances the same sequence of operations that is generated by some of the most widely used modern AI algorithms. We focus specifically on DPLL and Gibbs sampling (Geman and Geman 1984) because these algorithms and their variants are central to much work on reasoning in artificial intelligence. Key for relating these algorithms to simulation is the fact that a basic operation of each of these algorithms is to simulate alternate states of the world. They differ from each other mostly in how they choose and elaborate these simulations. We claim that this connection with AI algorithms, though it does not establish the truth of simulation theories, eliminates much concern about their computational power. Further, contrary to the views of many modern researchers in both fields, but consistent with precedent, we claim that this work demonstrates that studying modern AI algorithms can advance our understanding of human inference and vice versa.

To be clear, we do not argue that human perception is conducted using reasoning algorithms such as DPLL and Gibbs sampling. Instead, we argue that even if other forms of mechanisms implement perception, these mechanisms can interact to create the same sequence of operations as

many AI reasoning algorithms. This implies that human perceptual mechanisms have at least as much power to explain non-perceptual human reasoning and problem solving as some of the most powerful known algorithms.

## Simulation mechanisms

In this section, we describe some basic mechanisms we will be assuming. These mechanisms are either perceptual or required for perceptual simulations and they are presupposed by most simulation theories. Although their power to make inferences is in question, their existence is well-supported empirically. We will refer to them collectively as “the simulation mechanisms”. Only a subset of known human perceptual mechanisms is discussed. Our goal is not an account of human perception, but only to show that human perceptual mechanisms have a high level of reasoning power. This can be demonstrated using only a few specific mechanisms.

### Activations and mental structures

Several aspect of brain states influence behavior. The activation of neurons, the topology of connections among neurons, the strength of these connections and the mix of hormones in the brain are all examples. We will be agnostic about the exact nature of these factors, but assume that certain patterns of brain state, whatever they are, regularly appear in the presence of certain stimuli. As an example of our notation, we will use the following notation to represent the patterns of brain state that regularly occur in the presence of a red object (COLOR red). When this is part of the current state of the brain, we will say that the “mental structure”<sup>2</sup> (COLOR red) is “active”.

Some (e.g., Spivey 2006) wonder whether such a discrete symbolic notation can capture the full array of human perceptual abilities, which seem to have many continuous and analog characteristics. We do not claim that it can. Instead, we use this notation to characterize only a specific subset of human perceptual abilities. Since our goal is to show that reasoning abilities we describe below are a subset of perceptual abilities, this will be adequate.

There is evidence that, at least while an object is being attended to, the brain keeps track of its properties. Thus, if a red square and a green circle are in a scene, the brain (in some cases, at least) keeps track of the fact that the redness and squareness belong to the same object and the greenness

<sup>2</sup> Following Jackendoff (2007), we avoid the word “representation” in order to make it clear that we are referring to nothing more than aspects of brain state. This is intended to avoid philosophical confusions regarding the use of the term.

and circlehood belong to the other object. Further, we know (Damasio 1989) and simulation theories often assume that there are parts of the brain that associate activations caused by the same object across different modalities. We can use the following notation to represent the associative component of brain state activated by, for example, the red square: (ASSOCIATIVE rs). We use square brackets to group the attributes of a single tracked object. In the case of the square and circle, we have: [(ASSOCIATIVE rs) (COLOR red) (SHAPE square)], [(ASSOCIATIVE gc) (COLOR green) (SHAPE circle)].

We further assume that part of the mental structure that becomes active when an object is perceived corresponds to how likely the object is to have that property. We can add this as another element in our notation. Thus (in poor lighting conditions, for example) uncertain redness can be written as (COLOR red 0.55). The activation corresponding to the object not being red can be written as (COLOR red 0) or (COLOR red false). Similarly, (COLOR red) is just shorthand for (COLOR red 1) or (COLOR red true). Nothing below depends on the exact nature of the mental structures associated with uncertainty.

Just as we assume there are mental structures relating objects to their attributes, we assume that there are mental structures for relating objects to each other. For example, we will indicate the mental structure that becomes active when objects o1 and o2 certainly touch with (TOUCH o1 o2 true).

We will sometimes further abbreviate mental structures relating objects to their attributes or to other objects with capital letters such as P or Q. “P” indicates that the relation is certain and “-P” that it is false.

### Simulations

Perceptual simulation theories hold that people use the same mechanisms to represent unseen objects, relations and events as they do those they are currently perceiving. For example, this would imply that if [(Assoc l) (Shape line) (Color red)] become active when a red line is perceived, then quite similar structures should become active when a red line is imagined. However, the structures cannot be completely identical. The behavior that follows from a particular brain state depends on whether the situation it corresponds to is occluded, past, future, hypothetical or counterfactual. For example, if in one’s desired world state an object is within reach, but in the actual perceived world state it is far away, he will generally not reach for it. Thus, in addition to mental structures for properties of objects, there must be some mental structure corresponding to the status of that property. We use “worlds” to make these distinctions. If a relation is perceived, we say it occurs in the “actual” world. If it is hypothetical, we say the world in

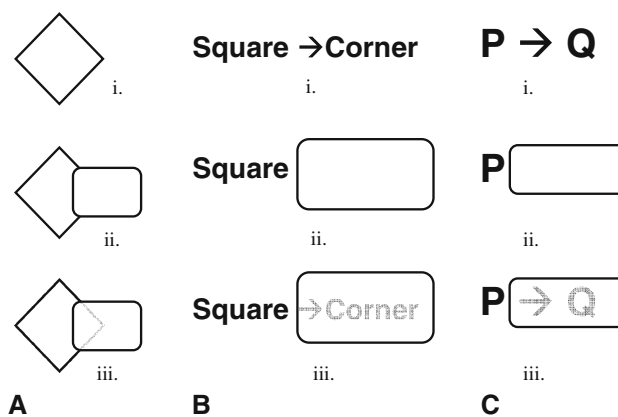
which it occurs is “hypothetical”. We indicate this in our notation by adding world elements to mental structures. Thus, a hypothetical color of an object being brown is indicated thus: [(Color c brown w)] [(Status w hypothetical)]. If a hypothetical world is based on the assumption of a set of relations being true, we say that those relations form its basis. We will refer to the process of representing a not-immediately-perceived world as “simulation”.

### Pattern completion

There is significant evidence (Barsalou et al. 2003) that when people see part of a scene, they form mental structures corresponding to unseen parts of it, often by using past experience about what sorts of entities tend to combine into commonly appearing patterns. For example, when seeing an object that is partially occluded, subjects form activations in their visual cortex corresponding to the appearance of the unseen parts of the object (Lerner et al. 2002). Figure 1a illustrates such a scenario.

In what follows, we assume the existence of at least one pattern completion mechanism that stores mental structures of perceived and inferred situations. We call the sum of mechanisms performing pattern completion, the “pattern completion mechanism” (PCM). Since we are only trying to demonstrate the computational power of the simulation mechanisms, the exact details of how many pattern completion mechanisms there are in the human brain or their exact nature is not relevant for our purposes.

We will assume that the PCM performs two operations when presented with a scene: (1) It stores (perhaps a subset of) the mental structures that are formed; (2) It creates mental structures corresponding to unseen parts of the scene that can be predicted based on the experiences it has stored. For example if an object is moving from a place p1



**Fig. 1** Examples of pattern completion. In **a**, *iii*, the parts of the square (**a**, *i*) occluded in **a**, *ii* are imagined. **b**, **c** illustrate modes ponens as an instance of pattern completion

towards an adjacent place  $p_2$  and nothing is there to block its motion, one can infer that it will go to  $p_2$ . Thus, when the PCM attends to

[(LOCATION  $p_1$  true a) (MOMENTUM right true a)]  
[(Status a actual)]

it completes to the following (where the completed part of the pattern is in bold):

[(LOCATION  $p_1$  true a) (MOMENTUM right true a)]  
[(Status a actual)] **[(LOCATION  $p_2$  true f) (MOMENTUM right f)] [(Status f future)]**.

In general the PCM takes as input a set of mental structures and outputs a set of mental structures that often or always occur with it. When it does so, we will say that the PCM “inferred” those mental structures.

During perception, the human PCM focuses on a definite stimulus. There is no doubt about what that stimulus is (e.g., what the retinal activation is), although there is often doubt about the underlying reality generating the stimulus. Thus, we assume that the PCM operates only on true or false relations and not uncertain relations. When the PCM focuses on a relation  $R$  that is uncertain, we say that it coerces it into relation  $R/w$ , which is the relation  $R$  in the alternate world  $w$  whose basis is the assumption that  $R$  is true. This kind of “uncertainty coercion” will ensure that the PCM operates on only true or false propositions. Of course, since the inferences drawn from an input are often uncertain, the output of the PCM can (and must) include uncertain relations.

Note that the PCM is also a memory mechanism. Since it remembers what it has seen, can focus on what it infers, and can extrapolate from these to make more general inferences, it includes some of the functions of both short-term and long-term memory.

#### Simulation success and failure

The failure of simulations can be informative. For example, suppose the simulation of a hypothetical world where a car has just been involved in an accident involves the car having dents. When the person performing the simulation perceives that the car has no dents, he can infer that the accident did not occur. More generally, when a simulation based on a relation,  $R$ , leads to the inference of false relations, then  $R$  can be inferred to be false. We call this “simulation assumption lifting” and we say a world or simulation that undergoes this process is “contradicted”. Evidence that people keep track of contradicted simulations is provided by Kaup et al. (2007).

We will further assume that the PCM keeps a record of contradicted worlds and incorporates that into its pattern completion thus: when a world based on relations  $R_1, \dots,$

$R_n - 1, R_n$  is contradicted, the PCM will complete  $R_n$ 's probability to being false in a world whose assumptions are  $R_1, \dots, R_n - 1$ . Thus, to return to our example, if the basis of the world,  $W$ , is the assumption,  $A$ , that a car has been in an accident and the simulation of  $W$  includes dents that are later observed not to exist, then the simulation of  $W$  is contradicted and the PCM will assume  $A$  is false (i.e., that the car was not just in an accident).

Finally, we will assume that when a world is contradicted that the PCM will cease outputting relations about that world. We will call this policy “contradiction cessation”. It is a well-known fact in logic that anything can be inferred from a contradiction and thus a pattern completer has limited utility in a contradicted world.

#### Attention control

If there is some limit to the size of the input of the pattern completion mechanism, then at any given moment of time, only a subset of the active mental structures can be elaborated. We will say that the PCM is attending to a set of mental structures if those mental structures are its current input.

These limitations have analogs in human visual attention. There is a limit to the number of objects, relations and events people can attend to at any given time. The precise extent and nature of this limitation is not relevant for the current discussion. We only assume that there is a limit. This limitation creates a problem. For example, if the front and rear of a car are occluded and only its midsection is perceived, the pattern completer will retrieve relations involving both the front and rear of the car. Since the front and rear are separated in space, both of these cannot be attended to simultaneously and the appropriate region upon which to focus attention must be selected.

A similar problem involves uncertainty in pattern completion. There may be two possible next locations for an object even though it is only possible to simulate one per time. We will assume that, as in perception in the real world, it is impossible to attend to both locations simultaneously. In general, the PCM will retrieve objects and relations that cannot be simultaneously attended to and the order in which these are attended must be chosen.

The visual system has several ways of dealing with the problem of limited attention. We will call these “attention selection heuristics”. In accord with simulation theory's claim that perceptual mechanisms are used for non-perceptual cognition, we will assume that these attention control heuristics also influence the control of mental simulations. For our purposes in future sections, we only need to discuss the following heuristics:

*Likelihood bias.* There is considerable evidence (Carpenter and Williams 1995; Kowler 1990; Kowler et al.

1984) that when people search for an object, they are in many cases more likely to attend to its most probable location. We extrapolate this heuristic to what we will call the “likelihood bias”:

When there is more than one relation to attend to, attend to the one marked with the highest certainty.

A special case of this heuristic occurs when there is probability information or when the certainties among potential relations competing for attention are equal. In that case, this heuristic leads to, barring the influence of other heuristics, picking one of these at random.

*Negative priming.* When there is more than one object or location in a scene and people attend to one of them, they are often less likely or slower to focus attention to the competing objects (Tipper 1985). This phenomenon is often called “negative priming”. We extrapolate negative priming in visual attention to the following more general simulation attention heuristic:

When there are competing relations to attend to and one of these is focused on, bias against focusing on the other relation for X seconds.

We parameterize negative priming by X because its duration has been found to vary considerably.

Note that although there is much uncertainty (Tipper 2001) regarding the character of the negative priming effect and the mechanisms that underlie it, there is no doubt that the effect is real. That is, there is no doubt that in many cases people are slower or to respond to a stimulus that was initially ignored in favor of another stimulus. This phenomenon, whatever its underlying explanation, is all that is assumed in this paper.

Heuristics such as these can conflict. For example, if the most likely relation is one that was ignored initially, negative priming will make the relation less likely to be focused on while the likelihood bias will make it more likely. How competition among heuristics is resolved is an important question, but for the purposes of the discussion that follows, we need only consider cases where these heuristics do not conflict.

Figure 2 depicts the interplay of the PCM and the attention heuristics. When the PCM attends to mental structures, it automatically outputs mental structures it can infer. The attention heuristics decide which of these the PCM attends to next.



**Fig. 2** The interplay of perception, the pattern completion mechanism and the attention heuristics

The plausibility of the simulation mechanisms

Pattern completion, imagery, negative priming and the likelihood bias are the only mechanisms we will be discussing in this paper. Let us consider the level of support for each of these mechanisms. First, the existence of brain configuration that correspond to certain stimuli is presupposed in much neuroscience, for example, in research that studies which patterns of stimuli particular neurons respond to. We have already cited evidence that the human perceptual system performs pattern completion, that the likelihood bias and negative priming influence visual attention and that there is some form of scanning of mental images that is similar to visual attention. There is little evidence for the specific details of how simulations operate (e.g., assumption bases for simulations), but it would be difficult for simulations to be of much use without the basic assumptions we have made about it. These assumptions are certainly not inconsistent with simulation theories. Finally, there are two reasons to assume that mental image scanning involves negative priming and the likelihood bias. First, there is extensive evidence that mental image scanning is regulated according to visual selection heuristics in general (Finke 1989). Second, extrapolating heuristics for controlling visual attention to mental simulations is certainly a natural consequence of the simulation theory doctrine that non-perceptual cognition is based on perceptual mechanisms. Thus, there is considerable reason to believe that the mechanisms described thus far exist in the human brain.

Finally, perceptual simulation theories often insist on mental structures that belong to specific perceptual modalities. Some of theories, however, discuss mechanisms that are, if not amodal, at least cross-modal. For example, Barsalou et al. (2003a, b), following Damasio (1989), presupposes a region that correlates activations in different modalities. Nothing that follows depends on how modality specific the simulation mechanisms are.

### Simulation mechanisms behave like inference algorithms

In this section, we demonstrate that the simulation mechanisms produce the same sequence of operations as some of the most powerful modern artificial intelligence algorithms.

Propositions, symbols, variables

It is often falsely assumed or implied that perceptual simulation theories are inconsistent with theories of cognition that involve discrete mental structures, variable bindings and predicates with arguments. This view is not a necessary

component of simulation theory (Barsalou 2005). For example, recall that we indicate that the color region of the brain is in a state of activation that occurs in the presence of red stimuli with [(COLOR red) (ASSOC: a)]. We could easily establish mappings between this notion and a more traditional logical notation, e.g., Color (a, red). Anderson (2007) makes a similar point about variables and rules by showing how they can be represented as feedforward networks. Thus, there is nothing about perceptual simulation theories that rules out mental “symbolic”<sup>3</sup> representations.

### Inference rules and production rules as pattern completion

Many theories of inference (e.g., Braine and O’Brien 1998) rely on inference rules as basic steps of inference. Such rules include modus ponens (“if P is true and P implies Q, then infer Q is true”) and modus tollens (“if Q is not true and P implies Q, then infer that P is not true”). These rules work because of regularities underlying the world. “p implies q” means that q regularly occurs when p. Implication in ordinary Boolean logic (e.g., “ $P \rightarrow Q$ ”) is the case where the regularity has no exceptions.

Storing and making inferences from such regularities is the also the function of a pattern completion mechanism. Rule-like behavior can therefore be produced by a pattern completer.<sup>4</sup> The following are two examples, illustrated in Fig. 3a, b, involve perhaps the most well-known inference rules.

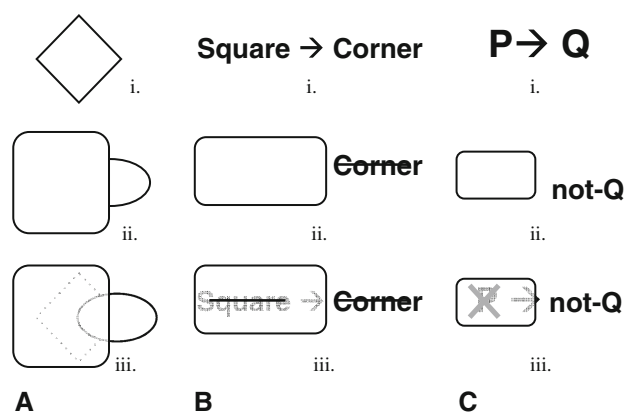
- *Modus ponens*. Suppose in every situation encountered by the PCM, P always occurred with Q. When the mechanism attends to a situation with P and has not yet observed whether Q, it will complete to Q. Specifically, [... P...] will complete to: [... P... Q...]. This mirrors the modus ponens inference rule.
- *Modus tollens*. If, as above, all occurrences of P have included occurrences of Q, then if the PCM attends to a situation where -Q, then the mechanism should complete the pattern to -P. Thus [... Q...] should complete to [... -P... -Q...].

### Hypothetical reasoning

Many theories of reasoning (e.g., Johnson-Laird 2007) hold that people imagine or form models of past, future, hypothetical or counterfactual states of the world. As we will

<sup>3</sup> It is perhaps less obvious (and not relevant to this paper) that logical or symbolic rule-based notions can capture analog (as opposed to discrete) aspect of cognition. However, many logics and rule systems take real numbers as values.

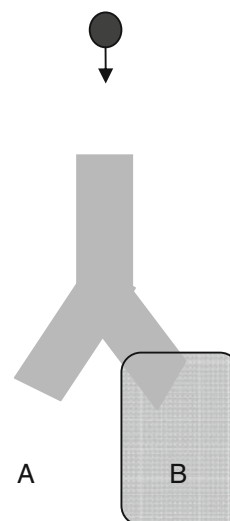
<sup>4</sup> In fact, there is no reason why a production rule system might not be part of a pattern completion mechanism.



**Fig. 3** Modus tollens as a form of pattern completion. When part of an occluded object (a, ii) is not consistent with the existence of a square (a, i), the visual system infers that the object is not a square (a, iii). As b and c illustrate, this is the same pattern of reasoning as in modus tollens

begin to discuss in subsequent subsections, many key AI algorithms are primarily methods for exploring such alternate states of the world.

The simulation mechanisms straightforwardly produce hypothetical reasoning. Consider the following situation, depicted in Fig. 4. A ball falls down a forked tube and has two possible destinations, A or B. A is visible and B is occluded. When viewing the ball fall down the tube, the pattern completion mechanism first takes as input: [(LOCATION start true)]. It can complete to either (LOCATION left-fork .5) or (LOCATION right-fork .5). Assume the first possible relation is attended to. Since that relation is uncertain, uncertainty coercion applies and the pattern completer thus attends to (LOCATION left-fork w1) (STATUS w1 hypothetical) (“the ball went through the left fork in the hypothetical world w”). PCM will then complete this to (LOCATION A w1)



**Fig. 4** A ball falls through a configuration of tubes

**Table 1** Hypothetical reasoning (A) emerging from the simulation mechanism (B). Simulation mechanisms are in bold

	A	B
1.	Be given a relation, A, that can be true or false	If the PCM infers that relation P is uncertain, then the <b>likelihood bias</b> will choose A or -A to focus on. Suppose it is A. <b>Negative priming</b> will prevent attention shifting to -A and thus enable the following steps to occur
2.	Imagine that A is true/false	Since A is uncertain, it is, according to <b>uncertainty coercion</b> , marked as being in a world, w, whose basis is the assumption that A
3.	Infer consequences of this assumption	<b>PCM</b> focuses on P in w and makes inferences
4.	If the consequences lead to a contradiction, infer that A is false if it was assumed to be true or infer it was true if assumed to be false	If <b>PCM</b> infers a contradiction, w is contradicted and <b>PCM simulation assumption lifting</b> will now complete P to false when it attends to it in the future

(STATUS w1 hypothetical) (“the ball lands at A if it went through the left fork”). Since it is perceived that there is no ball at location A, the simulation of w1 is contradicted. Since the basis of w1 was the assumption that the ball went through the left tube, PCM, according to simulation assumption lifting, can now complete the ball’s location to the right tube. In sum, when the ball fell through the tube, two paths were possible. The hypothetical world where the ball went through the left tube and landed in A was simulated and found to contradict the perceived evidence. The ball was thus inferred to not be at A, but to be at B.

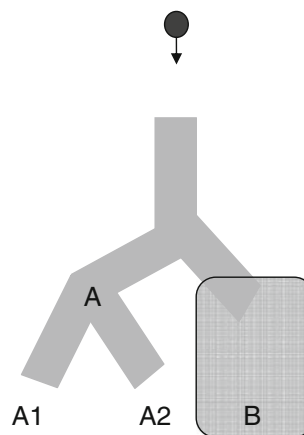
This particular pattern of hypothetical reasoning (Table 1A) is quite general and useful: when uncertain about whether A is true or false, imagining the world where A is true and deriving a contradiction allows one to conclude that A is in fact false. This pattern of reasoning emerges from the interaction of the simulation mechanisms (Table 1B). When a relation A is uncertain, either A or not-A can be attended to by the PCM. If the likelihood bias leads to A being chosen, A will be focused on in a hypothetical world because A is uncertain. Further pattern completion in that simulation might lead to it being contradicted and thus, because of simulation assumption lifting, lead to A being inferred as false in the actual world by the pattern completion mechanism. Thus, the specific operations of hypothetical reasoning are a natural result of the interaction of the simulation mechanisms.

**Backtracking search**

In many cases, one step of hypothetical reasoning is not sufficient to explore all possibilities because there may be more than one event or relation in the world whose status is uncertain. For example, the effect of a single move (M1) in chess often depends on which of the possible countermoves the opponent chooses. Thus, when performing hypothetical simulation of the outcome of M1, it often becomes

necessary to simulate several of the possible countermoves. To evaluate the effect of these countermoves, it might be necessary to simulate possible responses (M2) to countermoves. This process can be repeated for several steps. At any step in this process, if the simulation of a particular move is seen to be disadvantageous, one often “backtracks” and considers an alternative move.

This process is often called “backtracking search” and has been historically an important a component of many theories of human reasoning and problem solving (e.g., Newell et al. 1958). It also forms the basis of many important algorithms in modern artificial intelligence research on inference. One of the most important of these algorithms is the DPLL procedure. DPLL is a kind of depth-first search algorithm. Variations of DPLL (e.g., Moskewicz et al. 2001; Een and Sorensson 2005) are some of the most effective reasoning algorithms in modern artificial intelligence. DPLL solves constraint satisfaction problems that equivalent to many planning (Kautz and Selman 1999) and probabilistic inference (Kautz and Selman 1999; Sang et al. 2005) problems.



**Fig. 5** Backtracking search used to determine the destination of the ball

**Table 2** DPLL (A) behavior emerges from the simulation mechanisms (B). Simulation mechanisms are in bold

	DPLL (world-knowledge, world-state):	When simulating a world:
1.	Elaborate (world-state)	The <b>PCM</b> will find relations that can be inferred with complete confidence. Because their likelihood is the highest, the <b>likelihood bias</b> will cause these to be focused on earlier than other relations
2.	If	If
2.1	World-state is not inconsistent with world knowledge, return assignment	None of these certain relations is a contradiction and no uncertain relations remain, then there will be no relations in this world left to focus on. The world will be remembered by the <b>PCM</b> as possible
2.2	There is a contradiction, return failure	One of these relations leads to a contradiction, then nothing else will be focused on because of <b>PCM contradiction cessation</b>
3.	Choose a proposition from among the uncertain variables in world state	If <b>PCM</b> outputs uncertain relations, <b>likelihood bias</b> chooses one of them to focus on. <b>Infinite negative priming</b> would prevent contradictory relations from intruding on what follows
4.	Return DPLL (world-state with this variable true) or DPLL (world-state with this variable false )	This process will repeat itself until there are no more new variables to focus on

The normal operation of the simulation mechanisms produces the same behavior as DPLL. We can demonstrate this in terms of the hypothetical reasoning apparatus of the last section. In the example of the last section, recall that the simulation of the world  $w_1$  was aborted (via contradiction cessation) because it simulated the ball being at A when it was perceptually evident that there was no ball at A. Consider the modified configuration in Fig. 5. It is identical to the one in Fig. 4 except that A includes a fork leading to tubes ending in positions A1 and A2. In this case, since A is not visible, the simulation of world  $w_1$  (where the ball traveled through A) would not be contradicted by perception. Since there is uncertainty about whether the ball went to A1 and A2, each of those options would be simulated. By the same process of hypothetical reasoning in the last section, simulations of each of those options would be contradicted by perception. Simulation assumption lifting would therefore lead to the inference that the ball did not go to A1 in world  $w$  and that it did not go to A2 in world  $w$ . This causes a contradiction in the PCM because those were the only two possible options. Thus, the simulation of world  $w$  (where the ball went through A) halts because of contradiction cessation and simulation proceeds again as we described in the last section to infer that the ball did not go through A but to B instead. This example illustrates that hypothetical simulations nested within hypothetical simulations follow the same pattern of operation as backtracking search.

Table 1A describes the DPLL algorithm. It is essentially the recursive application of the hypothetical reasoning procedure described in the last subsection. Given a world

state and some knowledge about the world, DPLL elaborates that world state using a generalized version of the modus tollens and modus ponens rules. If there are no contradictions and the constraints are all satisfied, then nothing else happens. So far, this is just like hypothetical reasoning. However, if there is a contradiction over a proposition or if a proposition that is part of an unsatisfied constraint remains uncertain, DPLL will repeat hypothetical reasoning again on that proposition.

Table 1B illustrates how this sequence of steps results from the simulation mechanisms. The process is essentially the same as in hypothetical reasoning, except that more than one proposition is uncertain. The world knowledge input to DPLL can be cast a set of conditional rules (i.e.,  $p \rightarrow q$ ) of that the PCM can, as described in “[Inference rules and production rules as pattern completion](#)”, make inferences over (Table 2).<sup>5</sup>

#### Gibbs sampling

A popular class of methods for performing probabilistic inference involve “stochastic simulation”. To illustrate, consider a version ball-and-tube example from the last section that is identical except for locations A1 and A2

<sup>5</sup> DPLL’s “world knowledge” is expressed in conjunctive normal form (CNF), e.g.,  $(P \text{ or } Q)$  and  $(P \text{ or not-}Q \text{ or not-}R)$  and.... The parenthesized “clauses” in CNF can be written as a rule. For example,  $(P \text{ or not-}Q \text{ or not } R)$  is logically equivalent to  $(R \text{ and } Q) \rightarrow P$ . DPLL’s elaboration step involves performing “unit propagation”. Unit propagation sets propositions to truth values implied by already inferred and/or assumed truth values. Thus, with the CNF above, if P is false, then Q is inferred.

**Table 3** Gibbs sampling (A) behavior results from the simulation mechanism (B). Simulation mechanisms are in bold

	A	B
1.	Begin with a possible state of the world	Possible states of the world are arrived at using algorithms such as DPLL and hence the <b>simulation mechanisms</b>
2.	For each variable	The <b>PCM</b> outputs relations with their likelihoods.
a.	Find a probability distribution for its values given the rest of the state	The <b>likelihood bias</b> will focus on these, tending to choose the most likely values. Negative priming will ensure that all the variables are chosen
b.	Sample a value from that distribution and set create a new world state identical to the current one except for this one changed value	
3.	Do a large number of times	Once all variables have been so set, all the relations will have been inhibited equally by <b>negative priming</b> . Thus, the process will repeat itself again
4.	$P(\text{variable}) = \frac{\text{the number of worlds in which it is true}}{\text{the number in which it is false}}$	The <b>PCM</b> uses these simulations to add probabilities to its pattern completions

being occluded. In this case, the ball could have landed in either of A1 and A2 or B. In order to find out which is more likely, one could repeatedly simulate the ball falling through the apparatus. Each time there is some uncertainty about which way the ball would go, the simulation would make a choice such that it is more likely to follow the more likely direction through the fork. After running several such simulations, one could guess that the most likely landing spot of the ball will be the location it landed on most frequently during the simulations. Such methods form perhaps the most widely used group of methods for probabilistic inference in artificial intelligence. Although AI researchers and cognitive modelers do not generally consider them models of human inference, they are often used to make inferences in cognitive models that are based on Bayesian Networks.

Perhaps the most commonly used form of stochastic simulation is Gibbs sampling. It begins with a possible state of the world and continually generates new possible worlds by changing the value of a variable according to how likely that value is given the rest of the world state. The specific steps are illustrated in Table 3A.

Like hypothetical reasoning, ordinary backtracking search and DPLL, Gibbs sampling is a method that proceeds by exploring alternate states of the world. The simulation mechanisms operating normally produce the same behavior as Gibbs sampling as shown in Table 3B.

Thus, in the case where there are no contradictions, negative priming is absolute and lasts only for one fixation, the simulation mechanisms will produce the same behavior as Gibbs sampling. Of course, contradictions do often occur and the negative priming is often longer and less severe. This section does illustrate, however, that the simulation mechanisms are powerful enough to produce the kinds of inference MCMC and that in certain conditions approximate its behavior.

## Conclusions

We have demonstrated that pattern completion, imagery, the likelihood bias and negative priming can combine to perform the same operations as some powerful AI algorithms. This has several implications for cognitive theory.

First, the computational power of simulation theory is no less than that of some of the key algorithms in artificial intelligence. This is not just a trivial point about Turing equivalence, i.e., that it is possible to use simulations to build a Turing machine and thus it is possible to implement any algorithm on top of that. That merely implies that simulation mechanisms can be used to implement any computer program. Our conclusion is stronger: if people perform mental simulations guided by the aforementioned attention heuristics, then they would therefore be performing the same sequence of operations as the AI algorithms. To be clear, we do not claim that human perceptual mechanisms are implemented using these reasoning algorithms, only that they are powerful enough to produce the same behavior as these algorithms. In fact, though we have not argued this, we believe human perceptual mechanisms have many abilities that these algorithms do not.

Of course, this exposition has assumed a highly idealized set of mechanisms. For example, negative priming in humans is never absolute or infinite, there are many other mechanisms guiding attention and the mechanisms do not always operate as described. Thus, it is unlikely that people ever perform “pure” DPLL or Gibbs sampling simulation. However, the fact that idealized versions of human perceptual mechanisms perform these operations does underscore the inferential power of these mechanisms and it opens the potential of using sophisticated methods from artificial intelligence to understand their characteristics.

Similarly, our assumptions about the capacity of attention are likely to be highly idealized. We have assumed that

attention capacity is quite limited. This is, however, hardly a firmly established property of human cognition. Parallel competitive accounts of attention (Desimone and Duncan 1995; Itti and Koch 2001) devote some degree of processing to larger numbers of objects, locations and features. Scholl and Pylyshyn (1999) provide evidence that can simultaneously track approximately five objects. While understanding the limits on attention is key to a full account of human cognition, the potential for more attention capacity than has been assumed does not affect the ability of the perceptual mechanisms to execute these algorithms. Specifically, the main impact of serialism above is to mirror the fact that algorithms such as DPLL and Gibbs sampling, in their simplest incarnations, focus on one relation at a time. One could achieve this effect using parallel mechanism by artificially imposing limits on it. Although this would not be perfectly consistent with actual human cognition, it would demonstrate that the simulation mechanisms are as powerful as algorithms such as DPLL and Gibbs sampling.

A second implication of the relationship between simulation mechanisms regards the suppositions that (1) accepting simulation theory commits one to modeling methods that operate at the neural level or are somehow neurally-inspired and that (2) algorithms from “symbolic artificial intelligence” are more relevant for “amodal” theories of cognition, e.g., those involving rules. This work instead shows that in fact algorithms from both “symbolic” artificial intelligence are often based on simulations and that therefore are not inconsistent with simulation theory.

It does not follow, however, that the connection between AI algorithms and human cognition depend crucially on simulation theory being purely perceptual. No part of this work depends on mental structures being entirely perceptual or amodal. What makes *perceptual* simulation theory attractive is that it accounts for the origin of mental structures and PCM mechanisms that can operate on them. However, to the extent that there are amodal components to mental structure and the PCM, there is still the question of whether reasoning proceeds through simulations of alternate states of the world (as in Mental Model theory, Johnson-Laird 1983, 2007) and/or through other means alone, e.g., purely through rule matching. This work shows that simulation, modal or not, has more computational power than is commonly thought.

A final implication is that work on inference in artificial intelligence is very relevant to understanding the mechanism of human cognition and vice versa. This paper demonstrates that AI algorithms can give us a new way to understand the operation and limitations of human perceptual and cognitive mechanisms such as pattern completion, imagery and attention. Further, there are many issues in AI than can be addressed by reflecting on human cognition.

For example, some of the most significant recent advances in reasoning algorithms involve “clause learning” methods (Bayardo and Schrag 1997; Eén and Sorensson 2005; Marques-Silva and Sakallah 1996) which are special cases of production compilation in ACT-R (Anderson and Lebiere 1998) and rule chunking in Soar (Laird et al. 1987). Since the approach of Soar<sup>6</sup> and ACT-R to clause learning is much richer, it is conceivable that incorporating insights of those methods into DPLL-like algorithms can lead to further advances. As another example, DPLL does not specify the order in which variables are chosen. But different “variable orderings” have an important impact on the performance of DPLL (e.g., Aloul et al. 2001). The literature on human attention is of course full of insights into how people order their attention and thus a place to look for potential variable ordering heuristics. In preliminary work (Cassimatis et al. 2007), we have used the Polyscheme cognitive architecture to implement these algorithms as simulations. The work suggests that such an integration between reasoning algorithms and “lower-level” mechanisms can lead these algorithms to exhibit abilities that had not previously been possible. The correspondence between simulation mechanisms and AI algorithms thus not only suggests how to improve the performance of those algorithms, but also potentially motivates a method for exploring and quantifying the functional role of memory and attention mechanisms.

Of course, none of this demonstrates that human reasoning is exclusively based on perceptual simulations. It does, however, show that simulation theory is worth exploring, not despite its computational power, but because of it and because it enables cross-fertilization between AI algorithms and theories of human inference.

## References

- Aloul FA, Markov IL, Sakallah KA (2001) MINCE: a static global variable-ordering for SAT and BDD. Paper presented at the IEEE 10th international workshop on logic and synthesis
- Anderson JR (2007) How can the human mind occur in the physical universe?. Oxford University Press, New York
- Anderson JR, Lebiere C (1998) The atomic components of thought. Lawrence Erlbaum Associates, Hillsdale
- Barsalou LW (1999) Perceptual symbol systems. *Behav Brain Sci* 22:577–609

<sup>6</sup> Although not developed explicitly as a simulation theory, several aspects of Soar are consistent with this work. For example, Soar’s use of “problem spaces” to consider hypothetical actions is a kind of simulation. Further, Soar uses “impasses” to control its use of problem spaces. These impasses often involve conflicting options. Negative priming and the likelihood bias are both heuristics for dealing with conflicting options.

- Barsalou LW (2005) Abstraction as dynamic interpretation in perceptual symbol systems. In: Gershkoff-Stowe L, Rakison D (eds) Building object categories. Erlbaum, Mahwah, pp 389–431
- Barsalou LW, Niedenthal PM, Barbey A, Ruppert J (2003a) Social embodiment. In: Ross B (ed) The psychology of learning and motivation, vol 43. Academic Press, San Diego, pp 43–92
- Barsalou LW, Simmons WK, Barbey A, Wilson CD (2003b) Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci* 7:84–91. doi:[10.1016/S1364-6613\(02\)00029-3](https://doi.org/10.1016/S1364-6613(02)00029-3)
- Bayardo RJ, Schrag RC (1997) Using CSP look-back techniques to solve real world SAT instances, (pdf document). Paper presented at the 14th national conference on artificial intelligence
- Boroditsky L, Ramscar M (2002) The roles of body and mind in abstract thought. *Psychol Sci* 13(2):185–188. doi:[10.1111/1467-9280.00434](https://doi.org/10.1111/1467-9280.00434)
- Braine MDS, O'Brien DP (1998) Mental logic. Lawrence Erlbaum Associates, Mahwah
- Carpenter RH, Williams ML (1995) Neural computation of log likelihood in control of saccadic eye movements. *Nature* 377(6544):59–62. doi:[10.1038/377059a0](https://doi.org/10.1038/377059a0)
- Cassimatis NL, Bugjaska M, Dugas S, Murugesan A, Bello P (2007) An architecture for adaptive algorithmic hybrids. Paper presented at the AAAI-07, Vancouver, BC
- Damasio AR (1989) Time-locked multiregional retroactivation: a systems level proposal for the neural substrates of recall and recognition. *Cognition* 33:25–62. doi:[10.1016/0010-0277\(89\)90005-X](https://doi.org/10.1016/0010-0277(89)90005-X)
- Davis M, Putnam H (1960) A computing procedure for quantification theory. *J ACM* 7(1):201–215. doi:[10.1145/321033.321034](https://doi.org/10.1145/321033.321034)
- Davis M, Logemann G, Loveland D (1962) A machine program for theorem proving. *Commun ACM* 5(7):394–397. doi:[10.1145/368273.368557](https://doi.org/10.1145/368273.368557)
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222. doi:[10.1146/annurev.ne.18.030195.001205](https://doi.org/10.1146/annurev.ne.18.030195.001205)
- Een N, Sorensson N (2005) MiniSat-A SAT solver with conflict-clause minimization. In: SAT 2005 Competition
- Finke RA (1989) Principles of mental imagery. MIT Press, Cambridge
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 7:721–741
- Heras F, Larrosa J, Oliveras A (2008) MiniMaxSAT: an efficient weighted max-SAT solve. *J Artif Intell Res* 31:1–32
- Hoos HH, Stützle T (2002) SATLIB: an online resource for research on SAT. In: Gent IP, Maaren HV, Walsh T (eds) SAT 2000. IOS Press, Amsterdam, pp 283–292
- Itti L, Koch C (2001) Computational modeling of visual attention. *Nat Rev Neurosci* 2(3):194–203. doi:[10.1038/35058500](https://doi.org/10.1038/35058500)
- Jackendoff R (2007) Language, consciousness, culture: essays on mental structure. MIT Press, Cambridge
- Johnson-Laird P (1983) Mental models. Harvard University Press, Cambridge
- Johnson-Laird PN (2007) How we reason. Oxford University Press, New York
- Kaup B, Yaxley RH, Madden CJ, Zwaan RA, Lüdtke J (2007) Experiential simulations of negated text information. *Q J Exp Psychol* 60:976–990. doi:[10.1080/17470210600823512](https://doi.org/10.1080/17470210600823512)
- Kautz H, Selman B (1999) Unifying SAT-based and graph-based planning. Paper presented at the IJCAI-99
- Kowler E (1990) The role of visual and cognitive processes in the control of eye movement. In: Kowler E (ed) The role of visual and cognitive processes in the control of eye movement. Elsevier, Amsterdam, pp 1–63
- Kowler E, Martins AJ, Pavel M (1984) The effect of expectations on slow oculomotor control: IV anticipatory smooth eye movements depend on prior target motions. *Vis Res* 24(3):197–210. doi:[10.1016/0042-6989\(84\)90122-6](https://doi.org/10.1016/0042-6989(84)90122-6)
- Laird JE, Newell A, Rosenbloom PS (1987) Soar: an architecture for general intelligence. *Artif Intell* 33:1–64. doi:[10.1016/0004-3702\(87\)90050-6](https://doi.org/10.1016/0004-3702(87)90050-6)
- Lerner Y, Hendler T, Malach R (2002) Object-completion effects in the human lateral occipital complex. *Cereb Cortex* 12(2). doi:[10.1093/cercor/12.2.163](https://doi.org/10.1093/cercor/12.2.163)
- Marques-Silva JP, Sakallah KA (1996) GRASP: a new search algorithm for satisfiability. Paper presented at the international conference on computer-aided design
- Moskewicz M, Madigan C, Zhao Y, Zhang L, Malik S (2001) Chaff: engineering an efficient SAT solver. Paper presented at the 39th design automation conference, Las Vegas
- Newell A, Shaw JC, Simon HA (1958) Elements of a theory of human problem solving. *Psychol Rev* 65:151–166. doi:[10.1037/h0048495](https://doi.org/10.1037/h0048495)
- Richardson DC, Spivey MJ, Barsalou LW, McRae K (2003) Spatial representations activated during real-time comprehension of verbs. *Cogn Sci* 27:767–780
- Sang T, Beame P, Kautz H (2005) Solving Bayes networks by weighted model counting. Paper presented at the AAAI-05
- Scholl BJ, Pylyshyn ZW (1999) Tracking multiple items through occlusion: clues to visual objecthood. *Cogn Psychol* 38:259–290. doi:[10.1006/cogp.1998.0698](https://doi.org/10.1006/cogp.1998.0698)
- Spivey M (2006) The continuity of mind. Oxford University Press, New York
- Tipper SP (1985) The negative priming effect: inhibitory priming with to be ignored objects. *Q J Exp Psychol* 37A:571–590
- Tipper SP (2001) Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Q J Exp Psychol* 54:321–343. doi:[10.1080/02724980042000183](https://doi.org/10.1080/02724980042000183)