Gibson J (1971) The information available in pictures. *Leonardo* **4**: 27–35.
Smolensky P (1988) On the proper treatment of connectionism. *Behavioral and Brain Sciences* **11**: 1–23.
Von Eckardt B (1993) *What is Cognitive Science?* Cambridge, MA: MIT Press.
Watson RA (1995) *Representational Ideas: From Plato to Patricia Churchland*. Dordrecht, Netherlands: Kluwer.
Wittgenstein L (1921) *Tractatus Logico-Philosophicus*. London, UK: Routledge.
Wittgenstein L (1953) *Philosophical Investigations*. Oxford, UK: Blackwell.

Representations Using Formal Logics

Intermediate article

Selmer Bringsjord, Rensselaer Polytechnic Institute, Troy, NY, USA
Yingrui Yang, Rensselaer Polytechnic Institute, Troy, NY, USA

CONTENTS

LOGIC AND THE LANGUAGE OF THOUGHT

Johnson-Laird and Savary (1995) present the following 'illusion' (illusion 1):

1. If there is a king in the hand then there is an ace, or if there isn't a king in the hand then there is an ace (but not both).
2. There is a king in the hand.

Given these premises, what can one infer?

Almost certainly your verdict is this: one can infer that there is an ace in the hand. And you reached this verdict despite the fact that we introduced the problem as an 'illusion', which no doubt, at least to some degree, warned you that something unusual was in the air. Why do we refer to it as an *illusion*? Because your verdict seems correct, even perhaps obviously correct, and yet a little logic suffices to show not only that are you wrong, but that in fact what you can infer is that there *isn't* an ace in the hand.

Of course, not everyone is tricked. How do we explain the fact that Jones is, while Smith is not? The explanation offered by traditional cognitive science is that both Jones' and Smith's cognition involves knowledge: knowledge that is represented in logic (or logic-like systems), and knowledge that is processed by reasoning. So, an explanation of why some subjects 'crack' the illusion and others do not must be couched in terms of such representation and reasoning. According to one view, thinking, at least of the 'high-level' sort, takes place in a logic-like language; in other words, logic is the 'language of thought' for *Homo sapiens* (Braine, 1998; Fodor, 1975). We are not concerned here with whether or not this view is true. (For an argument in its favor, see Bringsjord and Ferrucci (1998).) Our concern is rather with presenting the simplest of those formal languages that are popular candidates for the language of thought. We want to explain logic as a means for both representing knowledge and reasoning with that knowledge.

To begin, we note that modern symbolic logic has three main components: one is purely syntactic, one is semantic, and one is metatheoretical in nature. The syntactic component includes specification of the alphabet of a given logical system, the grammar for building well-formed formulae (WFFs) from this alphabet, and a proof theory that precisely describes how and when one formula can be proved from a set of formulae. The semantic component includes a precise account of the conditions under which a formula in a given system is true or false. The metatheoretical component includes theorems, conjectures, and hypotheses concerning the syntactic component, the semantic component, and connections between them. The two simplest and most used logics for

representation in cognitive science are the propositional calculus (also known as 'propositional logic' or 'sentential logic') and the predicate calculus. The second of these subsumes the first, and is often called 'first-order logic' (FOL). We now proceed to characterize the three components for both the propositional calculus and FOL, starting with the former.

PROPOSITIONAL LOGIC

Grammar

The alphabet for propositional logic is simply an infinite list $(p_1, p_2, \ldots)$ of propositional variables (traditionally p_1 is p, p_2 is q, and p_3 is r), and the five familiar truth-functional connectives $\neg$, $\rightarrow$, $\leftrightarrow$, $\wedge$, and $\vee$. These connectives can, at least provisionally, be read, respectively, as 'not', 'implies' (or 'if... then...'), 'if and only if', 'and', and 'or'. In cognitive science it is often convenient to use propositional variables as mnemonics that help one remember what they are intended to represent. For an example, recall illusion 1. Instead of representing 'there is an ace in the hand' as 'p_i', for some $i \in \{1, 2, \ldots\}$, it would be convenient to represent this proposition as 'A'. Now, the grammar for propositional logic is composed of the following three rules:

1. Every propositional variable p_i is a WFF.
2. If ϕ is a WFF, then so is $\neg\phi$.
3. If ϕ and ψ are WFFs, then so is $(\phi * \psi)$, where $*$ is one of $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. (We allow outermost parentheses to be dropped.)

This implies, for example, that $p \rightarrow (q \wedge r)$ is a WFF, but $\rightarrow q$ isn't. To represent the declarative sentence 'if there is an ace in the hand, then there is a king in the hand' we could use $A \rightarrow K$.

Syntactic Proofs (Proof Theory)

A number of proof theories are possible. One such system is an elegant Fitch-style system of natural deduction, $\mathcal{F}$ (Barwise and Etchemendy, 1999). (Such systems are commonly referred to as 'natural' systems.) In $\mathcal{F}$, each of the truth-functional connectives has a pair of corresponding inference rules, one for introducing the connective, and one for eliminating the connective. Proofs in $\mathcal{F}$ proceed in sequence line by line, with successive line numbers incremented by 1. Each line includes a line number, a formula (the one deduced at this line), and, in the rightmost column, a rule cited in justification for the deduction. We use a vertical ellipsis ⋮ to indicate the presence of 0 or more lines in the proof not explicitly shown.

Here is the rule for eliminating a conjunction:

⋮	⋮	⋮
k	$\phi \wedge \psi$	
⋮	⋮	⋮
m	ϕ	k $\wedge$-Elim
⋮	⋮	⋮

Intuitively, this rule says that if at line k in some derivation you have somehow obtained a conjunction $\phi \wedge \psi$, then at a subsequent line m, one can infer to either of the conjuncts alone.

Now here is the rule that allows a conjunction to be introduced; intuitively, it formalizes the fact that if two propositions are true then the conjunction of these two propositions is also true:

⋮	⋮	⋮
k	ϕ	
⋮	⋮	⋮
l	ψ	
⋮	⋮	⋮
m	$\phi \wedge \psi$	k, l $\wedge$-Intro
⋮	⋮	⋮

An important rule in $\mathcal{F}$ is 'supposition', according to which you are allowed to assume any WFF at any point in a derivation. The catch is that you must signal your use of supposition by setting it off typographically, as follows:

⋮	⋮	
k	ϕ	supposition
	⋮	
⋮	⋮	

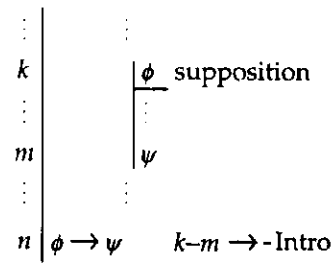
Often a derivation will be used to establish that from some set Φ of propositional formulae a particular formula ϕ can be derived. In such a case, Φ will be given as suppositions (or, as we sometimes say, 'givens'). To say that ϕ can be derived in $\mathcal{F}$ from a set of formulae Φ we write

$$\Phi \vdash_{\mathcal{F}} \phi$$

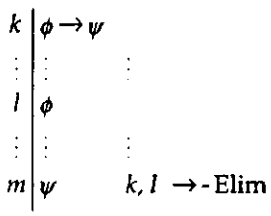
When it is clear from context which system the deduction is to take place in, the subscript on $\vdash$ can be omitted. Here is a proof that puts to use the rules presented above and establishes that $((p \wedge q) \wedge r) \vdash_{\mathcal{F}} q$:

1	$(p \wedge q) \wedge r$	given
2	$(p \wedge q)$	$1 \wedge$-Elim
3	q	$2 \wedge$-Elim

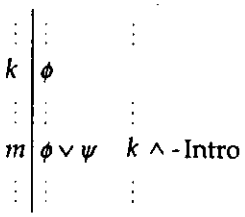
Now here is a slightly more complicated rule, one for introducing a conditional. It basically says that if you can carry out a subderivation in which you suppose ϕ and derive ψ, you are entitled to close this subderivation and infer to the conditional $\phi \rightarrow \psi$:



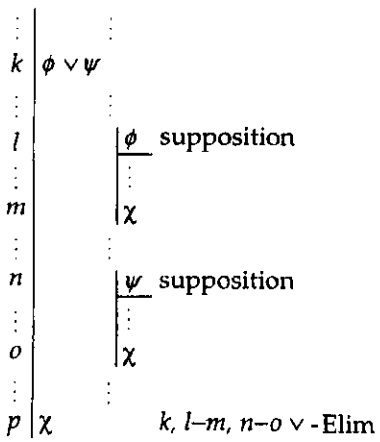
As we said, in a Fitch-style system of natural deduction, the rules come in pairs. Here is the rule in $\mathcal{F}$ for eliminating conditionals:



Here is the rule for introducing $\vee$:

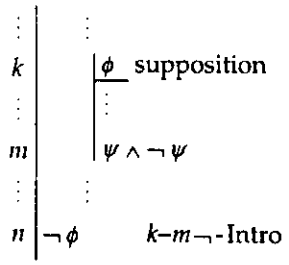


And here is the rather more elaborate rule for eliminating a disjunction:



The rule $\vee$-Elim is also known as 'constructive dilemma'. The intuition behind this rule is that if one knows that either ϕ or ψ is true, and if one can show that χ can be proved from ϕ alone, and from ψ alone, then χ must be true.

Next, here is a very powerful rule corresponding to proof by contradiction (sometimes called 'indirect proof' or 'reductio ad absurdum'). Notice that in $\mathcal{F}$ this rule is $\neg$-Intro:



Sometimes a natural deduction system can be a little awkward, because by insisting that inference rules come exclusively in the form of pairs for each truth-functional connective, it leaves out certain rules that are exceedingly useful. Two examples are *modus tollens* and DeMorgan's laws. The former rule allows one to infer $\neg\phi$ from $\phi \rightarrow \psi$ and $\neg\psi$. This rule can be established through a proof in $\mathcal{F}$, as shown in Figure 1. This proof was constructed in the Hyperproof construction environment (Barwise and Etchemendy, 1994). The core of this proof is *reductio ad absurdum*, or $\neg$-Intro.

DeMorgan's laws for propositional logic sanction moving from a formula of the form $\neg(\phi \wedge \psi)$ to one of the form $\neg\phi \vee \neg\psi$, and vice versa. The laws also allow an inference from a formula of the form $\neg(\phi \vee \psi)$ to one of the form $\neg\phi \wedge \neg\psi$, and vice versa. When, in constructing a proof in $\mathcal{F}$, we want to use modus tollens or DeMorgan's laws, or some other time-saving rule, we can make the inference, using the rule of 'tautological consequence' as a justification. This rule (abbreviated as 'Taut Con' in Hyperproof) is designed to allow the human proof constructor to declare that a given inference is obvious, and could with more work be fully specified using only the rules of $\mathcal{F}$. Hyperproof responds with a check to indicate that an attempted inference is in fact correct. In Figure 2, Hyperproof approves of our use of 'Taut Con', which corresponds in this case not just to DeMorgan's law, but also to the useful inference of $\phi \wedge \neg\psi$ from $\neg(\phi \rightarrow \psi)$.

A formula provable from the null set is called a 'theorem', and where ϕ is such a formula we write $\vdash \phi$ to express this fact. Here are two examples: $\vdash (p \wedge q) \rightarrow q$; $\vdash (p \wedge \neg p) \rightarrow r$. We say that a set Φ of formulae is 'syntactically consistent' if and only if no contradiction can be derived from Φ.

Figure 1. A proof of modus tollens in $\mathcal{F}$, constructed in Hyperproof.

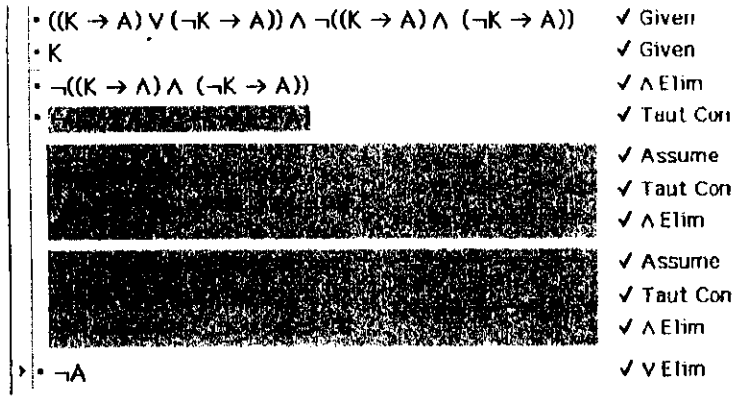


Figure 2. A proof in $\mathcal{F}$ that there is no ace in the hand. The premises are shown in the first two lines.

Semantics (Truth Tables)

The precise meaning of the five truth-functional connectives of the propositional calculus is given via truth tables, which tell us what the truth-value of a statement is, given the truth-values of its components. The simplest truth table is that for negation, which informs us, unsurprisingly, that if ϕ is T then $\neg\phi$ is F and if ϕ is F then $\neg\phi$ is T:

ϕ	$\neg\phi$
T	F
F	T

Here are the remaining truth tables:

ϕ	ψ	$\phi \wedge \psi$
T	T	T
T	F	F
F	T	F
F	F	F

ϕ	ψ	$\phi \vee \psi$
T	T	T
T	F	T
F	T	T
F	F	F

ϕ	ψ	$\phi \rightarrow \psi$
T	T	T
T	F	F
F	T	T
F	F	T

ϕ	ψ	$\phi \leftrightarrow \psi$
T	T	T
T	F	F
F	T	F
F	F	T

Notice that the truth table for disjunction says that when both disjuncts are true, the disjunction is true. This is called 'inclusive' disjunction. In 'exclusive' disjunction, it's one disjunct or another, but not both. This distinction becomes particularly important if one is attempting to symbolize parts of English (or any other natural language). It would not be natural to represent the sentence 'George will either win or lose' as '$W \vee L$', because under the English meaning there is no way both possibilities can be true, whereas by the meaning of $\vee$ it would be possible that W and L are both true. We can use $\vee_x$ to denote exclusive disjunction, which we define through the following truth table:

ϕ	ψ	$\phi \vee_x \psi$
T	T	F
T	F	T
F	T	T
F	F	F

It is worth mentioning another issue involving the meaning of English sentences and their corresponding symbolizations in propositional logic: the issue of the oddity of 'material conditionals' (formulae of the form $\phi \rightarrow \psi$). Consider the following English sentence:

> If the Moon is made of green cheese, then Dan Quayle will be the next President of the United States.

Is this sentence true? If we were to ask 'the man on the street', the answer might well be: 'Of course not!' Or perhaps we would hear: 'This isn't even a meaningful sentence; you're speaking nonsense.' However, when represented in the propositional calculus, the sentence turns out true. Why? The sentence is naturally represented as $G \rightarrow Q$. Since G is false, the truth table for $\rightarrow$ classifies the conditional as true. Results such as these have encouraged some to devise better (but much more complicated) accounts of the conditionals seen in natural languages (e.g. Goble, 2001). These accounts are beyond the scope of this article; we will be content with the conditional as defined by the truth table for $\rightarrow$ presented above.

Given a truth-value assignment υ (i.e., an assignment of T or F to each propositional variable p_i), we can say that υ 'makes true' or 'models' or 'satisfies' a given formula ϕ; this is written $\upsilon \models \phi$.

Some formulae are true on all models. For example, the formula $((p \vee q) \wedge \neg q) \rightarrow p$ is in this category. Such formulae are said to be 'valid' and are sometimes referred to as 'validities'. To indicate that a formula ϕ is valid we write $\models \phi$.

Another important semantic notion is 'consequence'. An individual formula ϕ is said to be a consequence of a set Φ of formulae provided that every truth-value assignment on which all of Φ are true is also one on which ϕ is true; this is written $\Phi \models \phi$.

The final concept in the semantic component of the propositional calculus is the concept of consistency: we say that a set Φ of formulae is 'semantically consistent' if and only if there is a truth-value assignment on which all of Φ are true.

Cracking Two Illusions

We now have at our disposal enough logic to 'crack' illusion 1. In this illusion, 'or' is to be understood as exclusive disjunction, so (using obvious symbolization) the two premises become $((K \rightarrow A) \vee (\neg K \rightarrow A)) \wedge \neg((K \rightarrow A) \wedge (\neg K \rightarrow A))$ and K.

Figure 2 shows a proof in $\mathcal{F}$, constructed in Hyperproof, that demonstrates that from these two givens one can conclude $\neg A$.

Now, consider another illusion (illusion 2):

1. The following three assertions are either all true or all false:
   - If Billy is happy, Doreen is happy.
   - If Doreen is happy, Frank is as well.
   - If Frank is happy, so is Emma.
2. Billy is happy.

Can it be inferred that Emma is happy?

Most people answer 'yes', but for the wrong reasons. They notice that since Billy is happy, if the three conditionals are true, one can 'chain' through them to arrive at the conclusion that Emma is happy. But this is only part of the story, and the other part has been ignored: it could be that all three conditionals are false. Other people realize that there are two cases to consider (conditionals all being true, and conditionals all being false), and because they believe that when the conditionals are all false one cannot prove that Emma is happy, they respond with 'No'. But this response is also wrong. The correct response is 'yes', because in both cases it can be proved that Emma is happy. This can be shown using propositional logic; the proof, again constructed in Hyperproof, is shown in Figure 3. This proof establishes

$$\{\neg(B \rightarrow D), \neg(D \rightarrow F)\} \vdash E$$

Note that the trick is exploiting the inconsistency of the set $\{\neg(B \rightarrow D), \neg(D \rightarrow F)\}$ in order to get a contradiction. Since everything follows from a contradiction, E can then be derived.

Metatheoretical Results

At this point we can give some metatheory for the propositional calculus. In general, metatheory would deploy logical and mathematical techniques in order to answer such questions as whether or not provability implies consequence, and whether or not the converse holds. When provability implies consequence, a logical system is said to be 'sound'. This fact can be expressed as: 'if $\Phi \vdash \phi$ then $\Phi \models \phi$'. Roughly, a logical system is sound if true formulae can only yield (through proofs) true formulae; one cannot pass from the true to the false.

When consequence implies provability, a system is said to be 'complete'. This is expressed by: 'if $\Phi \models \phi$ then $\Phi \vdash \phi$'.

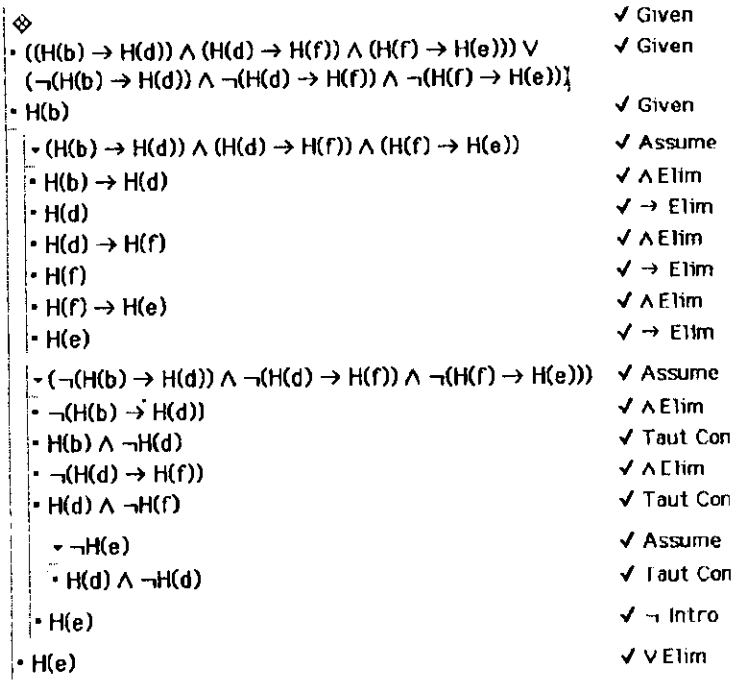


Figure 3. A proof in $\mathcal{F}$ that 'Emma is happy'.

The propositional calculus is both sound and complete. It follows that all theorems in the propositional calculus are valid, and all validities are theorems. This last fact is expressed more formally as: '$\models \phi$ if and only if $\vdash \phi$'.

FIRST-ORDER (PREDICATE) LOGIC

A Logical Illusion in Quantified Reasoning

Consider another illusion (illusion 3), a more complicated one that cannot be adequately represented in propositional logic. Here it is, adapted slightly from Yang and Johnson-Laird (2000):

> Only one of the following statements is true:
> - At least one of the beads is red.
> - None of the beads are red.
>
> Is it possible that none of the red things are beads?

We can remove all the mystery by turning to FOL, which we will introduce after explaining why the propositional calculus isn't expressive enough to represent this puzzle. The propositional calculus can represent propositions, but it cannot represent the internal structure of propositions: for example, propositions to the effect that certain objects have certain properties. In the propositional calculus, 'at least one of the beads is red' would be represented by some propositional variable, say N. We know that from 'at least one of the beads is red' it follows that 'at least one of the beads is red or blue'. The second statement here would be represented by some other propositional variable, say B. Clearly, $\{N\} \nvdash B$ in the propositional calculus. What is needed is some way to represent the fact that N says that at least one of a particular kind of object has a specific property, viz., being red. Only with such machinery can we get to the bottom of illusion 3.

The Syntactic Machinery of FOL

Our alphabet will now be augmented to include the following: the identity or equality symbol $=$; variables $x, y, \ldots$; constants ('proper names' for objects) $c_1, c_2, \ldots$; relational symbols $R, G, \ldots$ (e.g., R for 'being red'); functors (functions) $f_1, f_2, \ldots$; the existential quantifier $\exists$ ('there exists at least one ...'); the universal quantifier $\forall$ ('for all ...'); and the familiar truth-functional connectives $\neg$, $\vee$, $\wedge$, $\rightarrow$, and $\leftrightarrow$.

Predictable 'formation rules' are introduced to allow us to represent propositions like those in illusion 3. With these rules, we can now write such things as $\exists x(Bx \wedge Rx)$, which says that there exists at least one thing x that has property B and property R. As in propositional logic, sets of formulae (say Φ), given certain 'rules of inference', can be

used to prove individual formulae (say ϕ); such a situation is expressed by expressions having exactly the same form as those introduced above (e.g., $\Phi \vdash \phi$). The rules of inference for FOL in such systems as $\mathcal{F}$ include those we saw for the propositional calculus, and also new ones: two corresponding to the existential quantifier $\exists$, and two corresponding to the universal quantifier $\forall$. For example, one of the rules associated with $\forall$ says, intuitively, that if you know that everything has a certain property, then any particular thing a has that property. This rule, known as 'universal elimination' (or 'universal introduction') allows us to move from some formula $\forall x\phi$ to a formula with $\forall x$ dropped, and the variable x in ϕ replaced with the constant of choice. For example, from 'all beads are red', that is, $\forall x(Bx \rightarrow Rx)$, we can infer by $\forall$-Elim that $Ba \rightarrow Ra$, and if we happen to know that in fact Ba we can now infer by familiar propositional reasoning that Ra. The rule $\forall$-Elim in $\mathcal{F}$ is

$$\begin{array}{l|ll} k & \forall x\phi & \\ \vdots & \vdots & \vdots \\ l & \phi(\frac{a}{x}) & k\ \forall\text{-Elim} \end{array}$$

where $\phi(\frac{a}{x})$ denotes the result of replacing occurrences of x in ϕ with a.

Semantics (Interpretations)

FOL includes a semantic side, which systematically provides meaning (i.e., truth or falsity) for formulae. Unfortunately, the formal semantics of FOL are more tricky than the truth tables that are sufficient for the propositional level. In FOL, formulae are said to be true (or false) on 'interpretations' or 'models'; that some formula ϕ is true on an interpretation $\mathcal{I}$ is often written as $\mathcal{I} \models \phi$. (We say that $\mathcal{I}$ satisfies, or models, ϕ.) For example, the formula $\forall x\exists y Gyx$ might mean, on the standard interpretation for arithmetic, that for every natural number n, there is a natural number m such that $m > n$. In this case, the 'domain' is the set $\mathbb{N}$ of natural numbers; and G symbolizes 'is greater than'. Much more could of course be said about the formal semantics (or 'model theory') for FOL; but this is beyond the scope of the present article. For a full discussion using the traditional notation of model theory, see Ebbinghaus *et al.* (1984). There it is shown that FOL, like the propositional calculus, is both sound and complete.

Cracking the Illusion

We now have the tools to crack illusion 3. Yang and Johnson-Laird (2000) found that very few untrained reasoners answered the question in illusion 3 correctly – but these authors tacitly assumed that 'none of the beads are red' should be represented in a manner that entails that there do exist some beads. If we side with this interpretation, then the question becomes whether in both cases it can be proved that it's not possible that none of the red things are beads. That is, can it be proved that in both cases a contradiction arises if one assumes that none of the red things are beads? The first case is shown as an explicit Hyperproof-constructed proof in $\mathcal{F}$ in Figure 4.

This proof can be expressed informally as follows. We begin by assuming in the first line that the first statement of illusion 3 is true and the second statement is false. In the next line we assume that there are some red things, and that none of them are beads. Next, we isolate the proposition that there are some red beads, by deriving it from the first line. In the fourth line, we assume, in keeping with the third line, that some arbitrary object a is a red bead. Now we derive that all red

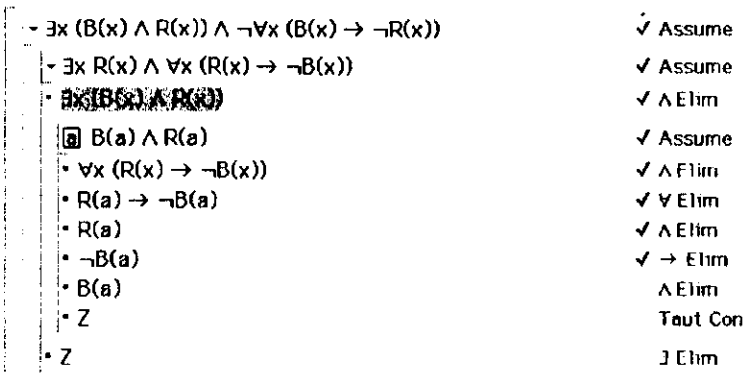


Figure 4. A derivation of a contradiction in $\mathcal{F}$.

things are not beads directly from line 2, by the $\wedge$-Elim rule. From this it follows by universal elimination that if the particular object a is red, it can't be a bead. Since we know that (under our assumptions) a is red, we can infer by modus ponens ($\rightarrow$-Elim, here) that a isn't a bead. But we are operating under the assumption that a is a bead (see line 4), so we have a contradiction. We use the propositional variable Z to represent an explicit contradiction. (Often the symbol $\perp$ is used for this purpose.) Since everything follows from a contradiction, we simply deduce Z from the contradiction of $B(a)$ and $\neg B(a)$. The reason is that we need to obey the rule of existential elimination, which insists that if something ϕ follows from assuming that some arbitrary thing (a in the proof) has some property, then we can infer that ϕ follows from the general existential claim that something x has that property – provided a doesn't occur in ϕ. At this point we have shown that Z, that is, a contradiction, arises if we assume that none of the red things are beads. We leave it to our readers to ascertain whether the second case of illusion 3 can be dealt with in a similar way.

Representations in Logics Beyond FOL

Many declarative sentences cannot be represented in FOL. Consider the sentence: 'If two things x and y are identical, then for every property F that x has, y has it as well, and vice versa.' This is known as Leibniz' law (LL), and it seems self-evident. But LL cannot be represented in FOL. However, LL can be represented in second-order logic (SOL), in which one can quantify not only over individual objects, but over properties as well. In SOL, LL becomes:

$$\forall x \forall y (x = y \leftrightarrow \forall X (Xx \leftrightarrow Xy))$$

Note that this formula contains a part meaning 'for every property X', which cannot be expressed in FOL. FOL permits quantification only over objects, not over the properties they can have. For an introduction to SOL, see Ebbinghaus *et al.* (1984); for a discussion of logics that can represent even more difficult constructions, see Goble (2001).

THE SITUATION CALCULUS

FOL is not only suitable for representing static information. It has long been used with great success to represent dynamic environments. One of the schemes for doing this is the 'situation calculus', which we will briefly describe, in connection with the 'wumpus world' test-bed. An example of a situation in this world is shown in Figure 5. The objective of the agent in this world is to find the gold and bring it back without getting killed. Pits are always surrounded by breezes or by other pits, the wumpus is always surrounded on at least three sides by a stench, and the gold glitters in the square in which it's positioned. The agent dies if it enters a square with a pit or a wumpus in it. (In the figure, the agent has managed to reach the gold.)

We can use the situation calculus to represent change in this world. First, we conceive of the world as consisting of a sequence of 'situations', essentially 'snapshots' of the world. Situations are generated from previous situations by actions. Properties in the world that can change over time are represented by inserting an extra constant to denote situations into the relevant formulae. For example, to describe the location of an agent a in the wumpus world at two situations s_0 and s_1, we could say that a is 'At' the square at row 1 and column 1, i.e., (1, 1), in situation s_0, and 'At' location (1, 2) in s_1. This would be written by:

$$\text{At}(a, (1,1), s_0) \wedge \text{At}(a, (1,2), s_1)$$

In order to use the situation calculus it is necessary to represent how the world changes from one situation to the next, by using the function

$$\text{Result}(\text{action}, \text{situation})$$

to refer to the situation that results from performing an action in some initial situation. With this function we can have sequences like that shown in Figure 6, in which, as a start, we have:

$$\text{Result}(\text{Forward}, s_0) = s_1$$
$$\text{Result}(\text{Backward}, s_1) = s_2$$
$$\text{Result}(\text{Turn}(\text{right}), s_2) = s_3$$
$$\text{Result}(\text{Forward}, s_3) = s_4$$

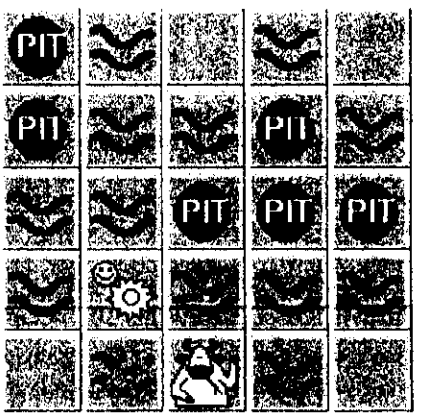


Figure 5. A typical wumpus world.

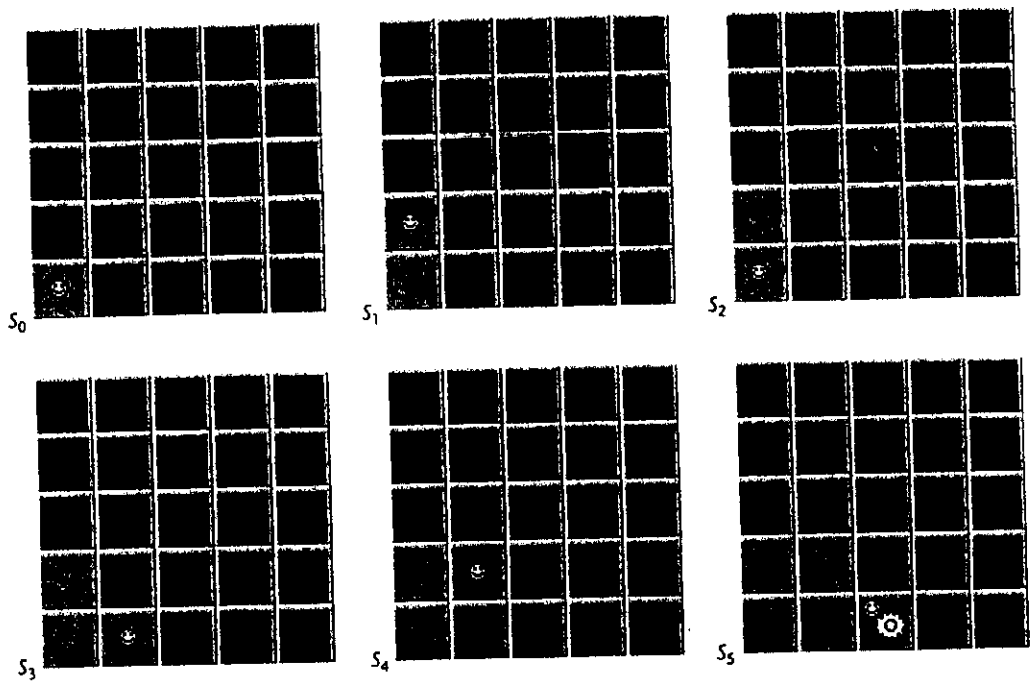


Figure 6. Situations s_0 to s_5 of a sequence in the wumpus world, linked by actions. (Program available at Escobar, 2002.)

The 'Result' function also allows us to use FOL to represent such general rules about the wumpus world as that the agent isn't holding anything after a 'Release' action:

$$\forall x \forall s \neg \text{Holding}(x, \text{Result}(\text{Release}, s))$$

For more on the situation calculus for the wumpus world and beyond, see Russell and Norvig (1994), which includes discussion of the infamous 'frame problem', which plagues such general rules.

THE CHALLENGE TO LOGIC-BASED REPRESENTATION IN COGNITIVE SCIENCE

Illusion 2 shows that cognizers sometimes conceive of 'disproofs'. Specifically, a cognizer fooled by illusion 2 imagines an argument for the view that there is no way to derive 'Emma is happy' from the negated trio of conditionals and 'Bill is happy'. Such arguments cannot be expressed in $\mathcal{F}$. However, a new theory, 'mental metalogic' (Yang and Bringsjord, 2001a,b; Rinella *et al.*, 2001), provides a mechanism in which step-by-step proofs (including disproofs) can at once be syntactic and semantic, because situations can enter directly into line-by-line proofs (see Figure 7). Hyperproof can be viewed as a simple instantiation of part of this theory. In Hyperproof, one can prove such things as that $\Phi \nvdash \psi$, not only such things as $\Phi \vdash \psi$.

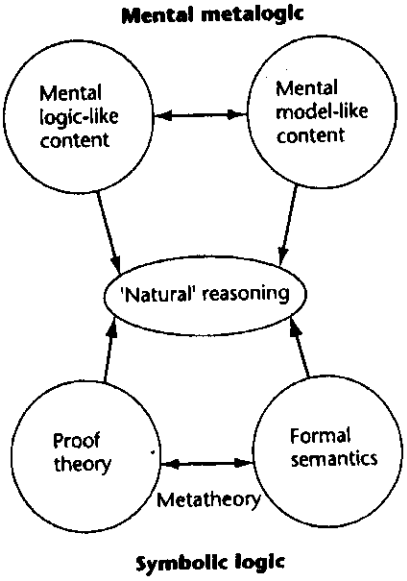


Figure 7. The symmetry of mental metalogic with symbolic logic.

Suppose that in illusion 2, a 'tricked' cognizer moves from a correct representation of the premises when the conditionals are all true to an incorrect representation when the conditionals are false.

Suppose, specifically, that the negated conditionals give rise to a situation, envisaged by the cognizer, in which four people *b*, *d*, *f*, and *e* are present, the sentence 'Billy is happy' is explicitly represented by a corresponding formula, and the question is whether it follows from this given information that Emma is happy. This situation is shown in Figure 8. Notice that the full logical import of the negated conditionals is nowhere to be found in this figure.

Next, given this starting situation, a disproof in Hyperproof is shown in Figure 9. Notice that a new, more detailed situation has been constructed, one which is consistent with the information given (hence the 'CTA' rule which, if correctly used, checks the truth of the assumptions) and in which it is directly observed that Emma isn't happy. This demonstrates that Emma's being happy can't be deduced from the given information. So, though untrained human reasoning may not conform to normative patterns, erroneous thinking can nonetheless be represented by logic, even when it involves pictorial reasoning. (For the reasons why Hyperproof derivations like that in Figure 9 cannot be reduced to purely syntactic reasoning, see Barwise and Etchemendy (1995).)

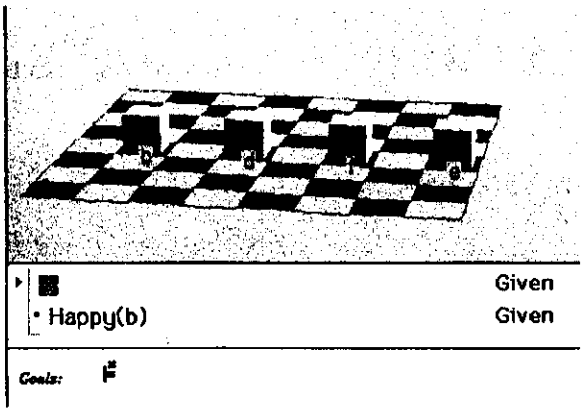


Figure 8. The start of a disproof that may be in the mind of cognizers.

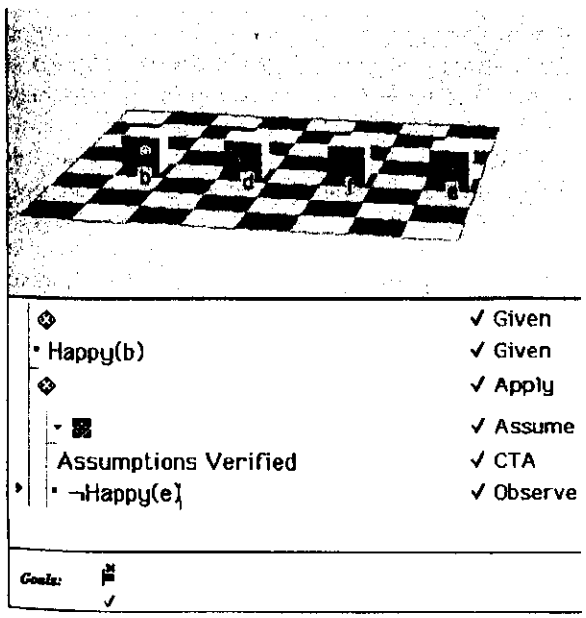


Figure 9. The completed disproof that may be in the mind of cognizers.

References

Barwise J and Etchemendy J (1994) *Hyperproof*. Stanford, CA: CSLI.

Barwise J and Etchemendy J (1995) Heterogeneous logic. In: Glasgow J, Narayanan N and Chandrasekaran B (eds) *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, pp. 211–234. Cambridge, MA: MIT Press.

Barwise J and Etchemendy J (1999) *Language, Proof, and Logic*. New York, NY: Seven Bridges.

Braine M (1998) How to investigate mental logic and the syntax of thought. In: Braine M and O'Brien P (eds) *Mental Logic*, pp. 45–61. Mahwah, NJ: Lawrence Erlbaum.

Bringsjord S and Ferrucci D (1998) Logic and artificial intelligence: divorced, still married, separated...? *Minds and Machines* 8: 273–308.

Ebbinghaus HD, Flum J and Thomas W (1984) *Mathematical Logic*. New York, NY: Springer.

Escobar J and Bringsjord S (2002) *Wumpus World*. http://kryten.mm.rpi.edu/otter/wumpus/Wumpus.html.

Fodor J (1975) *The Language of Thought*. Cambridge, MA: Harvard University Press.

Goble L (ed.) (2001) *The Blackwell Guide to Philosophical Logic*. Oxford, UK: Blackwell.

Johnson-Laird P and Savary F (1995) How to make the impossible seem probable. In: *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pp. 381–384. Hillsdale, NJ: Lawrence Erlbaum.

Rinella K, Bringsjord S and Yang Y (2001) Efficacious logic instruction: people are not irremediably poor deductive reasoners. In: Moore JD and Stenning K (eds) *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pp. 851–856. Mahwah, NJ: Lawrence Erlbaum.

Russell S and Norvig P (1994) *Artificial Intelligence: A Modern Approach*. Saddle River, NJ: Prentice Hall.

Yang Y and Bringsjord S (2001a) Mental metalogic: a new paradigm for psychology of reasoning. In: *Proceedings of the Third International Conference on Cognitive Science (ICCS 2001)*, pp. 199–204. Hefei, China: Press of the University of Science and Technology of China.

Yang Y and Bringsjord S (2001b) The mental possible worlds mechanism: a new method for analyzing logical reasoning problems on the GRE. In: *Proceedings of the Third International Conference on Cognitive Science (ICCS 2001)*, pp. 205–210. Hefei, China: Press of the University of Science and Technology of China.

Further Reading

Bringsjord S, Bringsjord E and Noel R (1998) In defense of logical minds. In: *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, pp. 173–178. Mahwah, NJ: Lawrence Erlbaum.

Johnson-Laird P (1997a) Rules and illusions: a critical study of Rips's *The Psychology of Proof*. *Minds and Machines* 7: 387–407.

Johnson-Laird P (1997b) And end to the controversy? A reply to Rips. *Minds and Machines* 7: 425–432.

Yang Y and Johnson-Laird P (2000) How to eliminate illusions in quantified reasoning. *Memory and Cognition* 28: 1050–1059.

Representations, Abstract and Concrete

Intermediate article

Joan Gay Snodgrass, New York University, New York, USA

CONTENTS

Abstract and concrete representations are symbols that refer to an object, event or idea that has existed, does exist or might exist in the real world. Concrete representations bear a physical resemblance to their referents, whereas abstract representations do not. Both come to be learned through the process of categorization.

CONCRETE VERSUS ABSTRACT REPRESENTATIONS

A representation is a symbol that stands for something else. Usually the 'something else' is an object, event or idea that has existed, does exist or might exist in the real world. A concrete representation is one that bears a physical resemblance to the real-world object it represents. This representation could be visual, as in a picture of an object; auditory, as in the sound the object makes; tactual, as in the shape of an object felt through the skin; or olfactory or gustatory, as in the smell or taste of an object. In contrast, an abstract representation is one that bears no resemblance to the object being represented. The abstract representation *par excellence* is a word or combination of words in a language. Almost everything that can be experienced or thought about can be captured in a word or series of words in a speaker's language.

This article is confined to concrete representations that are pictorial symbols. These are most useful for representing objects: for example, line drawings of common objects and animals, such as those shown in Figure 1, are easily recognized as representing objects in the real world, as evidenced by the fact that they can be easily named by both children and adults, and are given comparable names across languages (e.g. the chair is consistently named 'chair' in English, 'chaise' in French and 'sedia' in Italian). Line drawings are easily recognized as representing objects even by children who have not been previously exposed to drawings.

Other pictorial symbols have been used to represent more abstract ideas. A good example is the international system of road signs, which uses a set of easily understandable symbols for road commands (although such symbols are not universally understandable without some training – for example, does the picture of an upright palm of a hand mean 'stop' or 'no entrance'?) Another example is the icons introduced by Apple Computer and incorporated into the Microsoft Windows operating system and software for lessening the memory load on users.