Modeling real reasoning

Keith Devlin*
CSLI, Stanford University
devlin@csli.stanford.edu.

April, 2007

This paper is dedicated to my former colleague and good friend, the logician Kenneth Jon Barwise (1942–2000). The work presented here is very much in the spirit of his approach to logic, a theme I pick up in my closing remarks.

1 Introduction

In this article we set out to develop a mathematical model of real-life human reasoning. The most successful attempt to do this, classical formal logic, achieved its success by restricting attention on formal reasoning within pure mathematics; more precisely, the process of proving theorems in axiomatic systems. Within the framework of mathematical logic, a *logical proof* consists of a finite sequence $\sigma_1, \sigma_2, \ldots, \sigma_n$ of statements, such that for each $i = 1, \ldots, n$, σ_i is either an assumption for the argument (possibly an axiom), or else follows from one or more of $\sigma_1, \ldots, \sigma_{i-1}$ by a rule of logic.

The importance of formal logic in mathematics is not that mathematicians write proofs in the system. To do so would in general be far too cumbersome. Rather, the theory provides a framework for analyzing the notion of mathematical proof. This has led to several benefits. One is a deeper understanding of mathematical proof. Another is the development of techniques for proving that certain statements are in fact not provable. A third is the development of computer tools to carry out automated proof procedures and to assist the human user construct proofs. Still another benefit is that the study of formal logic has educational value for the apprentice mathematician. Generally speaking, those our our goals in trying to develop a model for what we shall call *real-life* logical reasoning.

Of course, one obvious approach to modeling reasoning is to apply formal logic itself, or simple modifications thereof, and this has been attempted on a number of occasions. The most recent significant attempt was in the early work in artificial intelligence in the second half of the twentieth century. By and large, all such attempts have failed. There are various explanations as to why this failure occurred (we outline our own particular take on this in our book [4]), but for the present purpose we need focus only on two issues.

The first issue is that real-life reasoning is rarely about establishing "the truth" about some state of affairs. Rather it is about marshalling evidence to arrive at a conclusion. If

*Many of the ideas presented in this paper were developed over several years, during which time my research was based at CSLI. Some of the research was supported by an award from the Advanced Research Development Agency, under a subcontract to General Dynamics Advanced Information Systems, as part of ARDA's NIMD Program.

the reasoner wants to attach a reliable degree of confidence to their conclusion, she or he must keep track of the sources of all the evidence used, the nature and reliability of those sources, and the reliability of the reasoning steps used in the process. As such, reasoning is better modeled as a process of gathering and processing information.

When you think about it, however, this observation does not amount to a significant departure from the standard model of formal logic, even insofar as logic is viewed as a model of *mathematical* reasoning. Although proofs, both those expressed in formal logic and the kind you find in professional mathematical journals, are often couched in terms of *truths* established, what any mathematical proof really amounts to is an accumulation of evidence — of information that leads to the stated conclusion. Moreover, that conclusion, by virtue of being shown to be *true*, is of interest precisely because it provides us with information! Talk of truth is, then, just a manner of description — one that is often appropriate when discussing proofs of theorems in mathematics (but on few other occasions, courts of law being the most obvious exception where talk of truth is pertinent).

Our second issue (actually a whole list of issues) is considerably more significant, however, and comes not from philosophical reflections on the nature of proof, but on empirical studies of people reasoning in real-life situations. The following set of features are characteristic of much everyday "logical reasoning," yet *formal logic embodies none of them*:

1. Reasoning is often context dependent. A deduction that is justifiable under one set of circumstances may be flat wrong in a different situation.

2. Reasoning is not always linear.

3. Reasoning is often holistic.

4. The information on which the reasoning is based is often not known to be true. The reasoner must, as far as possible, ascertain and remember the source of the evidentiary information used and maintain an estimation of its likelihood of being reliable.

5. Reasoning often involves searching for information to support a particular step. This may involve looking deeper at an existing source or searching for an alternative source.

6. Reasoners often have to make decisions based on incomplete information.

7. Reasoners sometimes encounter and must decide between conflicting information.

8. Reasoning often involves the formulation of a hypothesis followed by a search for information that either confirms or denies that hypothesis.

9. Reasoning often requires backtracking and examining your assumptions.

10. Reasoners often make unconscious use of tacit knowledge, which they may be unable to articulate.

The above list is taken from Richards J. Heuer, Jr.'s classic book *Psychology of Intelligence Analysis* [5][1], popularly known as the "intelligence analyst's bible."

Because of the nature of intelligence analysis, in particular the need to reach concrete conclusions, to document reasoning, and to supply adequate supporting evidence, this activity provides one of the best examples of "real life" logical reasoning outside of mathematics

[1]Re-published by the United States Central Intelligence Agency in 1999, this book is currently available only in download form from the CIA's website.

and science. Moreover, in order to improve its intelligence analysis capabilities, the United States intelligence communities have, over the years, carried out several in-depth studies of the way professional analysts work.[2] Heuer was involved in such a study. An intelligence analyst for many years, he returned to university to work on the doctoral dissertation that became his book. It provides an excellent summary of formal reasoning processes outside of mathematics as conducted by a body of professionals trained to do just that. We shall base our model on Heuer's findings.

2 How does information arise?

Since we are approaching reasoning as a specific form of purposeful information gathering and processing, a fundamental question to start with is, how is it possible for something in the world, say a book or a magnetic disk, to store, or represent, information? This question immediately generalizes. For, although we generally think of information as being stored (by way of representations) in things such as books and computer databases, any physical object may store information. In fact, during the course of a normal day, we acquire information from a variety of physical objects, and from the environment. For example, if we see dark clouds in the sky, we may take an umbrella as we leave for work, the state of the sky having provided us with the information that it might rain.

Staying for a moment with that example, how exactly does it come about that dark clouds provide information that it is likely to rain? The answer is that there is a systematic regularity between dark clouds in the sky and rain. Human beings (and other creatures) that are able to recognize that systematic regularity can use it in order to extract information.

In general, then, information can arise by virtue of systematic regularities in the world. People (and certain animals) learn to recognize those regularities, either consciously or subconsciously, possibly as a result of repeated exposure to them. They may then utilize those regularities in order to obtain information from aspects of their environment.

What about the acquisition of information from books, newspapers, radio, etc., or from being spoken to by fellow humans? This too depends on systematic regularities. In this case, however, those regularities are not natural in origin like dark clouds and rain. Rather they depend on regularities created by people, the regularities of human language.

In order to acquire information from the words and sentences of English, you have to understand English — you need to know the meanings of the English words and you need a working knowledge of the rules of English grammar. In addition, in the case of written English, you need to know how to read — you need to know the conventions whereby certain sequences of symbols denote certain words. Those conventions of word meaning, grammar, and symbol representation are just that: conventions. Different countries have different conventions: different rules of grammar, different words for the same thing, different alphabets, even different directions of reading — left to right, right to left, top to bottom, or bottom to top.

[2] Incidentally, it would be unwise to judge the quality of US intelligence analysis by what appear to be some spectacular and costly — in terms of money, human life, and global stability — failures of intelligence decisions by the United States government in the last few years. In all those cases, the problem was not the intelligence analysis, which was in fact highly accurate; rather that, for reasons of political ideology, the government of the day chose to ignore or distort the analysts' recommendations, just as they did with many reports on other issues from the scientific community. That's what can happen when the inmates are put in charge of the most powerful asylum in the world.

At an even more local level, there are the conventional information encoding devices that communities establish on an ad hoc basis. For example, a school may designate a bell ring as providing the information that the class should end, or a factory may use a whistle to signal that the shift is over.

The fact is, anything can be used to store information. All it takes to store information by means of some object — or more generally a configuration of objects — is a convention that such a configuration represents that information. In the case of information stored by people, the conventions range from ones adopted by an entire nation (such as languages) to those adopted by a single person (such as a knotted handkerchief). For a non-human example, DNA encodes the information required to create a lifeform (in an appropriate environment).

To make precise these general observations about information, we need to provide a precise, representation-free[3] definition of information, and, second, to examine the regularities, conventions, etc. whereby things in the world represent information. This is what two Stanford University researchers, Jon Barwise and John Perry, set out to do in the late 1970s and early 1980s. The mathematical framework they developed to do this they named Situation Theory, initially described in their book *Situations and Attitudes* [2], with a more developed version of the theory subsequently presented by Devlin in [3]. We shall provide an extremely brief summary of part of situation theory in the following section.

3 Situation theory

In situation theory, recognition is made of the partiality of information due to the finite, *situated* nature of the agent (human, animal, or machine) with limited cognitive resources. Any agent must employ necessarily limited information extracted from the environment in order to reason and communicate effectively.

The theory takes its name from the mathematical device introduced in order to take account of that partiality. A *situation* can be thought of as a limited part of reality. Such parts may have spatio-temporal extent, or they may be more abstract, such as fictional worlds, contexts of utterance, problem domains, mathematical structures, databases, or Unix directories. The distinction between situations and individuals is that situations have a *structure* that plays a significant role in the theory whereas individuals do not. Examples of situations of particular relevance to the subject matter of this paper will arise as our development proceeds.

The basic ontology of situation theory consists of entities that a finite, cognitive agent individuates and/or discriminates as it makes its way in the world: spatial locations, temporal locations, individuals, finitary relations, situations, types, and a number of other, higher-order entities.

The objects (known as *uniformities*) in this ontology include the following:

- *individuals* — objects such as tables, chairs, tetrahedra, people, hands, fingers, etc. that the agent either individuates or at least discriminates (by its behavior) as single, essentially unitary items; usually denoted in situation theory by $a, b, c, \ldots$
- *relations* — uniformities individuated or discriminated by the agent that hold of, or link together specific numbers of, certain other uniformities; denoted by $P, Q, R, \ldots$

[3] Of course, our theoretical framework will have to have its own representations. The theory we will use adopts the standard application-domain-neutral representation used in science, namely mathematics.

- spatial *locations*, denoted by $l, l', l'', l_0, l_1, l_2$, etc. These are not necessarily like the points of mathematical spaces (though they may be so), but can have spatial extension.
- *temporal locations*, denoted by $t, t', t_0, \ldots$. As with spatial locations, temporal locations may be either points in time or regions of time.
- *situations* — structured parts of the world (concrete or abstract) discriminated by (or perhaps individuated by) the agent; denoted by $s, s', s'', s_0, \ldots$
- *types* — higher order uniformities discriminated (and possibly individuated) by the agent; denoted by $S, T, U, V, \ldots$
- *parameters* — indeterminates that range over objects of the various types; denoted by $\dot{a}$, $\dot{s}$, $\dot{t}$, $\dot{l}$, etc.

The intuition behind this ontology is that in a study of the activity (both physical and cognitive) of a particular agent or species of agent, we notice that there are certain regularities or *uniformities* that the agent either individuates or else discriminates in its behavior.[4]

For instance, people individuate certain parts of reality as *objects* ('individuals' in our theory), and their behavior can vary in a systematic way according to spatial location, time, and the nature of the immediate environment ('situation types' in our theory).

We note that the ontology of situation theory allows for the fact that different people may discriminate differently. For instance, Russians discriminate as two different colors what Americans classify as merely different shades of blue.

Information is always taken to be information *about* some situation, and is taken to be in the form of discrete items known as *infons*. These are of the form

$$\ll R, a_1, \ldots, a_n, 1 \gg \,,\ \ll R, a_1, \ldots, a_n, 0 \gg$$

where R is an n-place relation and $a_1, \ldots, a_n$ are objects appropriate for R (often including spatial and/or temporal locations). These may be thought of as the informational item that objects $a_1, \ldots, a_n$ do, respectively, do not, stand in the relation R.

Infons are items of information. They are not things that in themselves are true or false. Rather a particular item of information may be true or false *about a certain part of the world* (a situation).[5]

Given a situation, s, and an infon σ, we write

$$s \models \sigma$$

to indicate that the infon σ is made factual by the situation s, or, to put it another way, that σ is an item of information that is true of s. The official name for this relation is that s *supports* σ.

---

[4]This is true not only of individuals but also of groups, teams, communities. If A and B are engaged in a dialogue or a conversation, or indeed any other form of joint action, they recognize uniformities as individuals in a similar ways. Socially, they negotiate the precise meanings of these, so that they can agree the exact shape of the uniformities that apply in the situation they are in.

[5]One of the advantages of the framework and notation provided by situation theory is that it allows us to express partial information about complex relations. For example, the relation *eat* presupposes agent, object, instrument, place, time, but much of this information can remain implicit, as in "I'm eating." This makes it possible to choose which aspect of the structure to emphasize in a given instance of interaction. And this choice of emphasis also carries information in its own right, since it is recognized and interpreted as attitude or intent.

It should be noted that this approach treats information as a *commodity*. Moreover a commodity that does not have to be true. Indeed, for every positive infon there is a dual negative infon that can be thought of as the opposite informational item, and both of these cannot be true (in the same situation).

A fundamental assumption underlying the situation-theoretic approach to information is that information is not intrinsic to any signal or to any object or configuration of objects in the world; rather information arises from interactions of agents with their environment (including interactions with other agents). The individuals, relations, types, etc. of the situation-theoretic ontology are (third-party) theorist's inventions. For an agent to carry out purposeful, rational activities, however, and even more so for two or more agents to communicate effectively, there must be a substantial agreement first between the way an agent carves up the world from one moment to another, and second between the uniformities of two communicating agents. For instance, if Alice says to Bob, "My car is dirty," and if this communicative act is successful, then the words Alice utters must mean effectively the same to both individuals. In order for a successful information flow to take place, it is not necessary that Alice and Bob share exactly the same concept of "car" or of "dirty," whatever it might mean (if anything) to have or to share an exact concept. Rather, what is required is that their two concepts of "car" and of "dirty" overlap sufficiently. The objects in the ontology of situation theory are intended to be theorist's idealized representatives — prototypes — of the common part of the extensions of individual agent's ontologies. In consequence, the infons are theoretical constructs that enable the theorist to analyze information flow.

Situation theory provides various mechanisms for defining types. The two most basic methods are type-abstraction procedures for the construction of two kinds of types: situation-types and object-types.

Situation-types. Given a *SIT*-parameter, $\dot{s}$, and a compound infon σ, there is a corresponding *situation-type*

$$[\dot{s} \mid \dot{s} \models \sigma],$$

the *type* of situation in which σ obtains.

This process of obtaining a type from a parameter, $\dot{s}$, and a compound infon, σ, is known as *(situation-) type abstraction*.

For example,

$$[SIT_1 \mid SIT_1 \models \langle\langle \text{running}, \dot{p}, LOC_1, TIM_1, 1 \rangle\rangle]$$

Object-types. These include the basic types *TIM, LOC, IND,* REL^n, *SIT, INF, TYP, PAR,* and *POL*, as well as the more fine-grained uniformities described below.

Object-types are determined over some initial situation.

Let s be a given situation. If $\dot{x}$ is a parameter and σ is some compound infon (in general involving $\dot{x}$), then there is a type

$$[\dot{x} \mid s \models \sigma],$$

the *type* of all those objects x to which $\dot{x}$ may be anchored in the situation s, for which the conditions imposed by σ obtain.

This process of obtaining a type $[\dot{x} \mid s \models \sigma]$ from a parameter, $\dot{x}$, a situation, s, and a compound infon, σ, is called *(object-) type abstraction*.

The situation s is known as the *grounding* situation for the type. In many instances, the grounding situation, s, is the world or the environment we live in (generally denoted by w).

For example, the *type* of all people could be denoted by

$$[IND_1 \mid w \models \langle\langle \text{person}, IND_1, \dot{l}_w, \dot{t}_{now}, 1 \rangle\rangle]$$

Again, if s denotes Jon's environment (over a suitable time span), then

$$[\dot{e} \mid s \models \langle\langle \text{sees, Jon}, \dot{e}, LOC_1, TIM_1, 1 \rangle\rangle]$$

denotes the type of all those situations Jon sees (within s).

This is a case of an object-type that is a type of situation.

This example is not the same as a *situation-type*. Situation-types classify situations according to their internal structure, whereas in the type

$$[\dot{e} \mid s \models \langle\langle \text{sees, Jon}, \dot{e}, LOC_1, TIM_1, 1 \rangle\rangle]$$

the situation is typed from the outside.

Types and the type abstraction procedures provide a mechanism for capturing the fundamental process whereby a cognitive agent classifies the world. Applying the distinction between situation types and object types to interaction phenomena, we may say that we all recognize that the relationship between situation-type *fire* and the situation-type *smoke* obtains only if both are in the same place at the same time. This is then a part of the shared knowledge among members of the same group or community that is often assumed and therefore rarely articulated. Situation theory offers a mechanism for articulating these assumptions by means of defined constraints. *Constraints* provide the situation theoretic mechanism that captures the way that agents make inferences and act in a rational fashion. Constraints are linkages between situation types. They may be natural laws, conventions, logical (i.e., analytic) rules, linguistic rules, empirical, law-like correspondences, etc.

For example, humans and other agents are familiar with the constraint:

Smoke means fire.

If S is the type of situations where there is smoke present, and S' is the type of situations where there is a fire, then an agent (e.g. a person) can pick up the information that there is a fire by observing that there is smoke (a type S situation) and being aware of, or *attuned to*, the constraint that links the two types of situation.

This constraint is denoted by

$$S \Rightarrow S'$$

(This is read as "S *involves* S'.")

Another example is provided by the constraint

FIRE *means fire.*

This constraint is written

$$S'' \Rightarrow S'$$

It links situations (of type S'') where someone yells the word FIRE to situations (of type S') where there is a fire.

Awareness of the constraint

$$\textsc{fire} \textit{ means fire}$$

involves knowing the meaning of the word FIRE and being familiar with the rules that govern the use of language.

The three types that occur in the above examples may be defined as follows:

$$\begin{aligned} S &= [\dot{s} \mid \dot{s} \models \langle\langle \text{smokey}, \dot{t}, 1 \rangle\rangle] \\ S' &= [\dot{s} \mid \dot{s} \models \langle\langle \text{firey}, \dot{t}, 1 \rangle\rangle] \\ S'' &= [\dot{u} \mid \dot{u} \models \langle\langle \text{speaking}, \dot{a}, \dot{t}, 1 \rangle\rangle \wedge \langle\langle \text{utters}, \dot{a}, \textit{fire}, \dot{t}, 1 \rangle\rangle] \end{aligned}$$

Notice that constraints link types, not situations. However, any particular instance where a constraint is utilized to make an inference or to govern/influence behavior will involve specific situations (of the relevant types). Constraints function by capturing various regularities across actual situations.

A constraint

$$C = [S \Rightarrow S']$$

allows an agent to make a logical inference, and hence facilitates information flow, as follows. First the agent must be able to discriminate the two types S and S'. Second, the agent must be aware of, or behaviorally attuned to, the constraint. Then, when the agent finds itself in a situation s of type S, it knows that there must be a situation s' of type S'. We may depict this diagrammatically as follows:

$$\begin{array}{rcl} S & \stackrel{C}{\Longrightarrow} & S' \\ s : S \uparrow & & \uparrow s' : S' \\ s & \stackrel{\exists}{\longrightarrow} & s' \end{array}$$

For example, suppose $S \Rightarrow S'$ represents the constraint *smoke means fire*. Agent $\mathcal{A}$ sees a situation s of type S. The constraint then enables $\mathcal{A}$ to conclude correctly that there must in fact be a fire, that is, there must be a situation s' of type S'. (For this example, the constraint $S \Rightarrow S'$ is most likely reflexive, in that the situation s' will be the same as the encountered situation s.)

A particularly important feature of this analysis is that it separates clearly the two very different kinds of entity that are crucial to the creation and transmission of information: one the one hand the abstract types and the constraints that link them, and on the other hand the actual situations in the world that the agent either encounters or whose existence it infers.

For further details of situation theory, the reader should consult [3], upon which the above account was based.

4 A situation-theoretic model of human reasoning

Our framework views reasoning as a temporal cognitive process that acts not on statements σ (as in the model of a mathematical proof) but on entities of the form

$$s \models_{\tau_1, \tau_2, \ldots} \sigma$$

where:

1. σ is a statement (or fact);
2. s is a *situation* which provides support or context of origin for σ; and
3. τ_1, τ_2, ... are the *indicators*[6] of σ, i.e., the specific items of information in s that the reasoner takes as justification of σ.

We call an entity of the form $s \models_{\tau_1,\tau_2,\ldots} \sigma$ a *basic reasoning element.*

Within our framework, a process of reasoning to decide an issue $\mathcal{I}$ can be represented like this:

$$
\begin{array}{ccc}
 & \mathcal{I} & \\
\hline
s_1 & \models_{\tau_1,\ldots} & \sigma_1 \\
s_2 & \models_{\tau_2,\ldots} & \sigma_2 \\
s_3 & \models_{\tau_3,\ldots} & \sigma_3 \\
 & \vdots & \\
\hline
s & \models_{\tau_1,\ldots,\tau_2,\ldots,\tau_3,\ldots} & \sigma
\end{array}
$$

where each basic reasoning element either supplies evidence for the reasoning or else follows from one or more previous elements by a logical deduction rule.

Analogous to the concept of a mathematical proof (sequence), we define (subject to some technical modifications) an *evidential reasoning process* as a finite sequence ρ_1, ρ_2, ..., ρ_n of entities of the above form such that each ρ_i is either *evidential* (i.e., an input to the reasoning process) or else the result of applying some logical rule of reasoning to one or more of ρ_1, ..., ρ_{i-1}. Here is the formal development of this notion.

By an *evidential reasoning element* we mean a 1×3 matrix of the form

FACT	SUPPORT	INDIC(1), INDIC(2), ...

such that

$$\text{SUPPORT} \models_{\text{INDIC}(1),\,\text{INDIC}(2),\,\ldots} \text{FACT}$$

By an *evidential reasoning step* we shall mean a finitary array of the form

OPERATOR	FACT_1	SUPPORT_1	$\text{INDIC}_1(1), \text{INDIC}_1(2), \ldots$
	FACT_2	SUPPORT_2	$\text{INDIC}_2(1), \text{INDIC}_2(2), \ldots$
	...		
	FACT_k	SUPPORT_k	$\text{INDIC}_k(1), \text{INDIC}_k(2), \ldots$
OUTPUT	FACT_{k+1}	SUPPORT_{k+1}	$\text{INDIC}_{k+1}(1), \text{INDIC}_{k+1}(2), \ldots$

where each row

FACT_i	SUPPORT_i	$\text{INDIC}_i(1), \text{INDIC}_i(2), \ldots$

is an evidential reasoning element. The index k depends on the operator OPERATOR, and is called the *arity* of the operator.

[6]Our use of the term "indicators" with this meaning comes from social science.

The idea is that a basic evidential reasoning step consists of the application of the logical operator to one or more constituents of the evidential reasoning elements in its *scope* (the first k elements listed) to produce the *output element* in the final row.

An *evidential reasoning process* is a finite sequence $\rho_1, \ldots \rho_n$ of basic reasoning steps such that each element is either *evidential* (i.e., an input to the reasoning process) or else the output of some previous (in the sequence) evidential reasoning step, or else is the special element STOP, which is the final element in the process. (STOP is a failure condition; we describe it later.)

The sequence of elements in an evidential reasoning process are not intended to provide a temporal model of the actual steps carried out by a reasoner. Rather, an evidential reasoning process models the logical flow of the reasoning as it leads to the conclusion. As we mentioned earlier, much real-life reasoning is not linear. However, our model is such that any linear progression of steps in the actual reasoning a human carries out will be mapped to a linear ordering of the corresponding basic reasoning elements in the model.

The actual operators that arise in any particular instance of reasoning will depend on the specific circumstances that pertain in that application. In this document we simply indicate the general form of some of the more generic operations that are likely to be used in any instance.

For example, among the operators are some that correspond to classical logic. Since classical logic ignores context, we have to exercise care in porting classical logic operators into our calculus. This means that our rules all have restrictions on when they may be applied. We start with the following two rules, each of which involves a binary reasoning operator:

Evidential Conjunction Rule

CONJOIN	σ	s	$\tau_1, \tau_2, \ldots$
	θ	t	$\gamma_1, \gamma_2, \ldots$
OUTPUT	$\sigma \wedge \theta$	$s \cup t \cup \{\delta\}$	$\delta, \tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots$

where $\delta = \text{Con}\{\tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots\}$, the assertion that the set $\{\tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots\}$ is logically consistent (i.e., has no internal contradictions), and where the rule may be applied only if δ is valid. The restriction that δ is called the *indicator consistency condition* for the rule. If this condition is not satisfied, the rule produces the output STOP. (We consider later what happens when the STOP element is generated.)

Evidential Modus Ponens Rule

MP	σ	s	$\tau_1, \tau_2, \ldots$
	$\sigma \rightarrow \theta$	t	$\gamma_1, \gamma_2, \ldots$
OUTPUT	θ	$s \cup t \cup \{\delta\}$	$\delta, \tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots$

where $\delta = \text{Con}\{\tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots\}$, and where the rule may be applied only if δ. If this condition is not satisfied, the rule produces the output STOP.

We need to exercise care in using these two rules. If the supports s and t are identical, there is in general no problem, nor if one support is contained within the other. In either of these cases, the indicator consistency condition can generally be assumed to be automatically satisfied, since reasoning generally proceeds under the tacit assumption that each individual source is internally consistent. (If, however, the reasoner suspects — or comes to suspect — that one of the supports used in the reasoning is internally inconsistent, then resolving that inconsistency becomes part of the reasoning process. This is a particular case of the following general observation concerning reasoning.)

The idea behind our approach is this. Coupling a fact σ with its support s in our framework does two things: (i) it acknowledges that σ does come from a particular source, and (ii) it provides a record of that source. Explicitly listing the indicators $\tau_1, \tau_2, \ldots$ with σ and s puts on record the particular items of information in s that the reasoner believes are salient in supporting σ, and uses to justify making use of σ in the reasoning. When an unexpected or troublesome conclusion is reached, or when the reasoning fails to yield a conclusion, it may be necessary to re-examine the veracity of some of the facts used in the reasoning, and that may involve reconsideration of the indicator already identified, or a search for indicators hitherto ignored. In an extreme case, the reasoner may have to question an entire source, perhaps rejecting it and looking for evidence elsewhere.

There are two unary reasoning operators associated with the indicators in a reasoning element: EVAL-INDIC, which checks the indicators already identified for veracity, and FACTORIZE, which identifies new items of information in the support that are salient to the use of the fact in the reasoning process. The rules associated with these operators are:

Indicator Evaluation Rule

EVAL-INDIC	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT::	σ	s	$\tau_1, \tau_2, \ldots$
	STOP		

where the notation here (note the double-colon after OUTPUT) indicates that the output of the rule is exactly one of the two elements

σ	s	$\tau_1, \tau_2, \ldots$

and

STOP

The former output is obtained if the evaluation of τ_1, τ_2, ... affirms their veracity; the output is STOP if the evaluation determines that one of these indicators is in fact not valid, or at least is in doubt.

Thus, the evidential reasoning step generated by an application of the Indicator Evaluation Rule is of one of the two forms:

EVAL-INDIC	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	σ	s	$\tau_1, \tau_2, \ldots$

EVAL-INDIC	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	STOP		

Indicators Extension Rule

EXTEND-INDICS	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	σ	s	$\tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots$

where $\gamma_1, \gamma_2, \ldots \in s$.

This rule implies that

$$s \models_{\tau_1,\tau_2,\ldots,\gamma_1,\gamma_2,\ldots} \sigma$$

The intuition is that the reasoner identifies additional information (additional indicators) that she or he judges to contribute to the acceptance of the fact σ under consideration.

Use of the following rule, which involves the unary operator EVAL-SUPPORT, indicates a suspicion that the reasoning process has a serious flaw.

Support Evaluation Rule

EVAL-SUPPORT	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT::	σ	s	$\tau_1, \tau_2, \ldots$
	STOP		

The former output is obtained if the evaluation of s affirms its internal consistency and reliability; the output is STOP if the evaluation determines that s is inconsistent or unreliable, or at least that the consistency or reliability of s is in serious doubt.

Thus, the evidential reasoning step generated by an application of the Support Evaluation Rule is of one of the two forms:

EVAL-SUPPORT	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	σ	s	$\tau_1, \tau_2, \ldots$

EVAL-SUPPORT	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	STOP		

When a reasoning step produces the output STOP, the reasoner has to backtrack and examine the process so far. If it is not possible to make any changes to any previous steps, then the reasoning process breaks down. In such a case, the available information is either contradictory or else simply not adequate to resolve the target issue.

A common step in reasoning is to decide between two or more different possibilities, which may or may not be mutually exclusive. The exact mechanism by which the comparison is made will vary from case to case, but functionally such an operation produces the following basic reasoning step:

Selection Rule

SELECT	σ_1	s_1	$\tau_1(1), \tau_1(2), \ldots$
	σ_2	s_2	$\tau_2(1), \tau_2(2), \ldots$
	$\ldots$		
	σ_n	s_n	$\tau_n(1), \tau_n(2), \ldots$
OUTPUT	σ_i	$s_i \cup s$	$\gamma, \delta, \tau_i(1), \tau_i(2), \ldots$

for some i, $1 \leq i \leq n$, where s is the very reasoning process the agent is carrying out (and which we are capturing with our calculus), $\gamma \in s$ is the fact that this particular selection has been made, and $\delta \in s$ is the criterion for making the selection.

Note that the output of a selection step carries a record of the selection having been made and of how it was made.

In practice, making a selection may involve examination of the supports and the indicators associated with the facts being compared, possibly leading to additional factorization for some facts or other operations. Such factorizations, or other steps, will be captured in our model by being represented explicitly as earlier steps in the process sequence.

Sometimes during the course of reasoning, the reasoner believes it is necessary to expand the scope of the domain from which particular facts were obtained, perhaps with a view to finding additional indicators to strengthen confidence in the fact or to replace the fact with a stronger version. This is captured by the following rules, often used in successively in conjunction, together with the indicators extension rule.

Support Expansion Rule

EXPAND-SUPPORT	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	σ	s'	$\tau_1, \tau_2, \ldots$

where $s \subseteq s'$.

Strengthen Fact Rule

STRENGTHEN-FACT	σ	s	$\tau_1, \tau_2, \ldots$
OUTPUT	σ'	s	$\tau_1, \tau_2, \ldots$

where $s \models_{\tau_1,\tau_2,\ldots} \sigma' \rightarrow \sigma$.

Multiple Views Uniformization Rule

Reasoners sometimes view more than one data source in order to use their experience and tacit knowledge to synthesize a conclusion that may not follow directly from the different sources by logical reasoning. To capture such actions, we could add an operator that provides some form of merge or unification for simultaneous views of information from different sources. However, the evidential conjunction rule that we already have will handle many cases of multiple views of data.

In circumstances where two views of a data item σ can be regarded as providing two indicator sets for the same fact within the same context:

$$s \models_{\tau_1,\tau_2,\ldots} \sigma \quad \text{and} \quad s \models_{\gamma_1,\gamma_2,\ldots} \sigma$$

we can apply the following operator:

MV UNIF	σ	s	$\tau_1, \tau_2, \ldots$
	σ	s	$\gamma_1, \gamma_2, \ldots$
OUTPUT	σ	s	$\delta, \tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots$

where δ is the fact that this unification has taken place.

Subtasking

Reasoners often need to break a particular task into subtasks. Typically, this entails defining a set of subtasks that together will complete the given task, and then working on each subtask in turn. Alternatively, the reasoner may decide to abandon (perhaps just for the time being) the current goal and concentrate solely on some subtask, which then becomes the new goal.

The framework as described so far can handle the individual steps in each subtask analysis, and can track choices of subtasks as localized reasoning contexts. But we have not introduced an operator for subtask selection or for breaking a task into a sufficient group of subtasks. Instead, we have left this as a meta-level operation. We did so in order to avoid making our technical machinery more complicated than it already is. Since our primary aim is to provide a framework to aid human reasoners, not a blueprint for an automated reasoning system, we feel this is a reasonable choice. But before moving on let's take a brief look at what would be required to modify our framework to incorporate subtasking.

Within our current framework, a process reasoning to decide an issue $\mathcal{I}$ is represented like this:

$$\begin{array}{ccc} & \mathcal{I} & \\ \hline s_1 & \models_{\tau_1,\ldots} & \sigma_1 \\ s_2 & \models_{\tau_2,\ldots} & \sigma_2 \\ s_3 & \models_{\tau_3,\ldots} & \sigma_3 \\ & \vdots & \\ \hline s & \models_{\tau_1,\ldots,\tau_2,\ldots,\tau_3,\ldots} & \sigma \end{array}$$

The issue $\mathcal{I}$ is kept constant throughout our development. In order to incorporate subtask selection, we could introduce a mechanism to represent the selection of a subtask $\mathcal{J}$ of $\mathcal{I}$ or else the division of $\mathcal{I}$ into a collection of subtasks $\mathcal{J}_1$, $\mathcal{J}_2$, $\ldots$, $\mathcal{J}_n$. The framework would need to keep track of the supports and the indicators, both when the subtask(s) is (are) selected and when the completion of all the tasks in a subdivision results in the completion of the original task. This is all very straightforward.

5 Some special cases

To get a sense of how our framework operates, we show how it applies to some familiar special cases or models for reasoning

Mathematical reasoning. First of all, let's take the case of mathematics, where $\sigma_1, \ldots, \sigma_n$ are statements about some mathematical structure $\mathcal{M}$, say a group or a field. We may assume that $\sigma_1, \ldots, \sigma_n$ are written in the first-order language for $\mathcal{M}$. In that case, each of the expressions $s_i \models \sigma_i$ denotes a standard proposition of classical Tarski-based model theory. In this case, by the Completeness Theorem of first-order predicate logic, $s = \mathcal{M}$ and the deduction takes the form

$$\begin{array}{c} \mathcal{I} \\ \hline \mathcal{M} \models \sigma_1 \\ \mathcal{M} \models \sigma_2 \\ \vdots \\ \mathcal{M} \models \sigma_n \\ \hline \mathcal{M} \models \sigma \end{array}$$

If this reasoning is valid, then we must have

$$\mathcal{I} = ![\mathcal{M} \models \sigma]?$$

where an expression of the form $!P?$ for some proposition P denotes the goal "Determine whether P true or false." That is, the goal is to determine whether or not σ is true of $\mathcal{M}$.

The completeness theorem also tells us that (if the deduction is valid), σ follows from $\sigma_1, \ldots, \sigma_n$ by the rules of logic alone.

Reasoning from a common source. Another special case is where all of the information $\sigma_1, \ldots, \sigma_n$ comes from the same source, $\mathcal{S}$. In this case the conclusion support s is also $\mathcal{S}$, and the deduction takes the form:

$$\frac{\dfrac{\mathcal{I}}{\begin{array}{c}\mathcal{S} \models \sigma_1 \\ \mathcal{S} \models \sigma_2 \\ \vdots \\ \mathcal{S} \models \sigma_n\end{array}}}{\mathcal{S} \models \sigma}$$

For a valid process, we must have

$$\mathcal{I} = !\,\sigma?$$

(Determine whether to do σ, or else determine whether σ is true.)

Bayesian inference. In some cases, knowledge of the source of each data item σ_i may be converted into a numerical probability of the reliability of σ_i, i.e. the probability that σ_i is true. In such a situation, we may be able to apply Bayes' Theorem repeatedly in order to obtain a conclusion σ and assign a probability to σ. In this case, the function F is a numerical function based upon Bayes' Theorem and the function H is an instance of Bayesian inference. This kind of reasoning is quite common, particularly in intelligence gathering.

We may represent a Bayesian reasoning process using the original notation

$$\frac{\dfrac{\mathcal{I}}{\begin{array}{c}s_1 \models \sigma_1 \\ s_2 \models \sigma_2 \\ \vdots \\ s_n \models \sigma_n\end{array}}}{s \models \sigma}$$

with the understanding that each of s_1,, $\ldots$, s_n, s is a number between 0 and 1 inclusive, and each expression $s_i \models \sigma_i$ should be interpreted as a probability statement $p(\sigma_i) = s_i$, and similarly for $s \models \sigma$.

6 Summary and discussion

The basis for our method is to view reasoning as a temporal cognitive process that acts on entities of the form

$$s \models_{\tau_1,\tau_2,\ldots} \sigma$$

where σ is a statement (or fact), s is its support or context of origin, and τ_1, τ_2, $\ldots$ are its indicators, the specific items of information in s that the reasoner takes as justification of σ.

We analyze reasoning so described in terms of a number of basic reasoning steps, an illustrative example being the Evidential Modus Ponens Rule:

MP	σ	s	$\tau_1, \tau_2, \ldots$
	$\sigma \to \theta$	t	$\gamma_1, \gamma_2, \ldots$
OUTPUT	θ	$s \cup t \cup \{\delta\}$	$\delta, \tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots$

where $\delta = \text{Con}\{\tau_1, \tau_2, \ldots, \gamma_1, \gamma_2, \ldots\}$, and where the rule may be applied only if δ.

We list a number of such rules, but acknowledge that many applications will involve rules not listed here. Our framework is designed to allow for such additional rules to be incorporated.

Readers familiar with situation theory will have observed that our present framework amounts to making explicit in the model the features of the context situation — our *indicators* — that provide direct support for the items of information considered in the reasoning — what we call the *facts*. Moreover, we model (aspects of) the process of reasoning, not just the sequence of facts and their situational supports. By making this additional salient information explicit in the model, we can obtain a finer grained analysis than is possible in situation theory, that requires much less ad hocing when we carry out an analysis of a specific reasoning process. In our framework, the Evidential Modus Ponens rule performs the task that was handled by constraints in situation theory. Our decision to ignore much of the machinery for handling situation-theoretic constraints was based on pragmatic grounds, with a view to the kinds of reasoning we are attempting to model.

Although our primary goal is to develop a framework that aids understanding, we are aware that any enterprise such as ours has the potential of forming the basis for the specification of reasoning protocols or the design of reasoning support tools. The model we have developed would result in protocols or support tools that:

1. Force explicit identification and tracking of sources.
2. Force explicit identification and tracking of supporting information (the indicators).
3. Force regular reconsideration of the reasoning process itself.
4. Allow for backtracking when a problem is encountered, without the necessity of starting over afresh.

Above all, our framework makes it clear that reasoning involves three components: facts, sources, and indicators. Real-life reasoning typically involves all three. Any protocol or tool developed in line with our model should provide the user with regular prompts to check all three components. Many examples of failures in human reasoning and analysis have resulted from a neglect of one or more of the three basic components.

Jon Barwise

I think it is appropriate to end with a quotation from my former friend and colleague Jon Barwise, whose untimely death in 2000 deprived the world of one of the most innovative logicians of the twentieth century. In his collected work *The Situation in Logic* [1], Barwise wrote [pp.xv–xvi]:

> Back in the days before I became interested in the situated aspects of logic, I sometimes used to wonder how logicians felt in the first quarter of this century. Did they *feel* confused. Reading the literature of that period, one senses the extent to which they were groping toward the view of logic that eventually emerged, but also the extent to which they were still in the dark about what was central and what was peripheral. One also realizes that they were just missing certain key distinctions. In other words, they were confused. It was only with the pioneering work of Gödel, Church, Turing, Tarski, and Kleene in the 1930's that the modern conception of logic really took hold.
>
> I now feel I have some idea of how logicians must have felt in that period before the really seminal work, since I feel we are in an analogous stage now ... As we try to let go of some of the simplifying idealizations made in standard logic, we too are groping for the key notions, and probably missing some key distinctions. In giving up these simplifying assumptions, there are many things to be rethought, many choices to be made, and many things to be tried. It is an exciting time, if you have the patience for that sort of thing, and a taste for the basic task of conceptual clarification. But it is also frustrating ...
>
> ... There is only one point about which I am really certain. That is that the view of language and logic as situated activities is an important one, and that situating logic is a task that must be carried out if we are to come to grips with some of the problems that currently vex the field.

I say Amen to that.

References

[1] Barwise, J. *The Situation in Logic*, CSLI Lecture Notes 17 (1989).

[2] Barwise, J. and Perry, J. *Situations and Attitudes*, Bradford Books, MIT Press (1983).

[3] Devlin, K. *Logic and Information*, Cambridge University Press (1991).

[4] Devlin, K. *Goodbye Descartes: The End of Logic and the Search for a New Cosmology of the Mind*, John Wiley (1997).

[5] Heuer, R. J. Jr., *Psychology of Intelligence Analysis*, Central Intelligence Agency (1999). Available on the Web at http://www.odci.gov/csi/books/19104/ .