How nonmonotonic is Aboutness?

P.D. Bruza
School of Information Systems
Queensland University of Technology
Australia
bruza@qut.edu.au

T.W.C. Huibers
Department of Computer Science
Utrecht University
The Netherlands
theo@cs.ruu.nl

Abstract

The notion of aboutness is fundamental to information retrieval. Assume there is a document d which is about query q. Now, if information is added to d yielding $\tilde{d}$, the question arises whether document $\tilde{d}$ is about q? In other words, is aboutness monotonic with respect to information composition? This article shows that aboutness does have nonmonotonic character with respect to composition.

1 Introduction

An often used premise in information retrieval is the following: if a given document d is *about* the request q, then there is a high likelihood that d will be relevant with respect to the associated information need. Thus, the information retrieval problem is reduced to determining the aboutness relation between documents and requests. Many information retrieval mechanisms have been developed, and there is a wide variation in how they determine aboutness. Recent investigations have centred around formalizing the notion of aboutness by axiomatizing its properties in terms of a neutral, underlying theory of information [BH94, HB94]. This work is incomplete, however, as the apparent nonmonotonic behaviour of aboutness has only briefly been touched upon. By nonmonotonicity we mean the following: Assume that document d is about information carrier q. If information is added to d, the resultant document *may not be* about q. Stated simply, aboutness is nonmonotonic with respect to information composition. This article will investigate this behaviour and propose two rules which permit a conservative form of information composition. These rules are essentially information retrieval equivalents of the rational and cautious monotonicity properties found in the Kraus, Lehmann and Magidor meta-theory of nonmonotonic reasoning [KLM90].

2 Background and Preliminaries

In this article we will investigate certain logical properties of aboutness within the framework of a language the words of which represent so called term phrases. These phrases are natural language expressions used to characterize information objects. The language, denoted $\mathcal{L}_T$, is based on a set T of terms. The words of this language will be referred to as *information carriers*.

A fundamental property of information carriers is that they can be composed [Lan86]. For example, consider the information carriers $\mathsf{information}$ and $\mathsf{retrieval}$. These can be composed to form the carrier $(\mathsf{information} \oplus \mathsf{retrieval})$. This information carrier corresponds to the term phrase $\mathsf{information\ retrieval}$. The symbol $\oplus$ is used to denote information composition. Even though it is a binary operator, higher order term phrases can be expressed in $\mathcal{L}_T$ using brackets to identify structure. For example,

$$(\mathsf{hydrogen} \oplus \mathsf{ion}) \oplus \mathsf{exchange}$$

renders the left compound noun phrase hydrogen ion exchange. More formally,

Definition 2.1

Let T be a set of terms. Then,

- $T \subset \mathcal{L}_T$
- *if* $i, j \in \mathcal{L}_T$ *then* $(i \oplus j) \in \mathcal{L}_T$

⊡

The composition operator $\oplus$ is assumed to be idempotent and commutative. As it is not associative, $\oplus$ cannot be distibuted in the normal way. The following distributions can be motivated from a linguistic standpoint [LD94]:

$$\begin{aligned} (i \oplus j) \oplus k &= (i \oplus j) \oplus (j \oplus k) \\ i \oplus (j \oplus k) &= (i \oplus k) \oplus (j \oplus k) \end{aligned}$$

For example, the information carrier $(\mathsf{hydrogen} \oplus \mathsf{ion}) \oplus \mathsf{exchange}$ can be viewed as the composition of the information carriers $\mathsf{hydrogen} \oplus \mathsf{ion}$ and $\mathsf{ion} \oplus \mathsf{exchange}$.

Not all information carriers can be meaningfully composed. The reason for this is that the information they share clashes, contradicts or is meaningless. In other words, carriers i and j are said to preclude each other, denoted $i \perp j$. Under the assumption that martians are either blue or green, but not both, it can be argued that $\mathsf{green} \oplus \mathsf{martians}$ precludes $\mathsf{blue} \oplus \mathsf{martians}$. This intuition behind this phenomenon can be explained in terms of situations: After characterizing a situation as being a "green martian" situation, it cannot be re-characterized as a "blue martian" situation.

Furthermore, an aboutness relation $\models_a$ is assumed over the information carriers. The notation $i \models_a j$ denotes that information carrier i is about carrier j. The intuition behind this is that the "information borne by j holds in i" [BH94]. This definition can be found explicitly or implicitly in several papers on logic-based information retrieval [Coo71, Rij86, Bru93]. Although this definition is intuitive, it expresses that aboutness is some form of entailment. Recent work has attempted to establish the properties of this relation via an axiomatic approach [BH94, HB94]. This research has shown that the relation is not a classical form of entailment. For example, the And rule does not hold. Ensuing sections will show that aboutness can be considered as preferential entailment.

3 Information Carriers and Preferential Models

An information carrier can be viewed as describing a set of situations. Take, for example, the information carrier migration. This carrier can be seen as discriminating a set of migration situations, for example, the migration of sparrows, cranes, salmon etc. Driven by their specific information need a searcher wishes to be informed about a particular situation or set of situations. Observe very closely that the information need imposes a *preferential* ordering on situations. That is, the searcher will readily prefer situation s over situation t because, for example, (s)he wants to be informed about s more than (s)he wants to be informed about t. This observation is fundamental to the theory presented in this article; the nonmonotonicity of aboutness is a consequence of it. Figure 1 depicts a preferential ordering over a set of migration situations. In this example the searcher prefers to be informed about bird migration situations over salmon migration ones, which in turn are preferable over migration situations involving humans.

Based on the above intuition, the preference relation can be formalized as follows. Let $s \sqsubset_N t$ denote that situation s is preferable to situation t in the light of information need N. It is natural to assume that $\sqsubset_N$ is irreflexive, that is, it makes no sense to say that a situation s is preferred

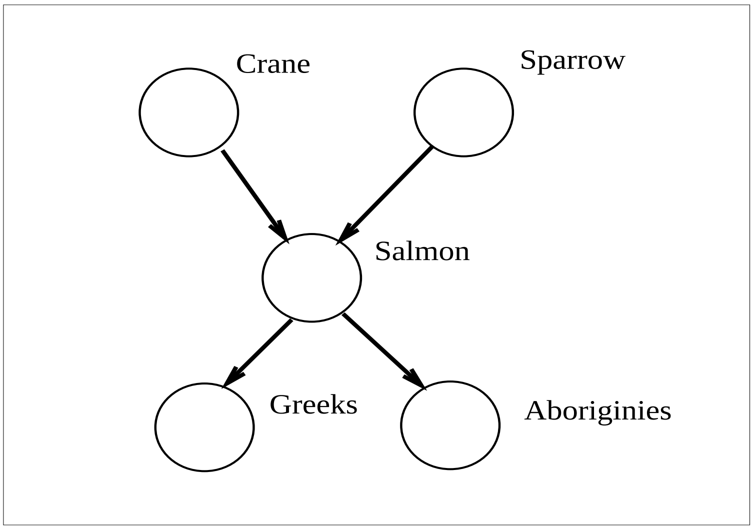

Figure 1: Example preferential ordering on migration situations

over itself. Certainly preference is antisymmetric: if s is preferred over t, then it is not the case that $t \sqsubset_N s$. A final property to state about the $\sqsubset_N$ relation is that it is transitive. Stated simply, $\sqsubset_N$ is a strict partial order on a set of situations $\mathcal{S}$. The structure $\langle \mathcal{S}, \sqsubset_N \rangle$ will be referred to as a preferential structure. Assuming that the searcher does not want to be informed by an infinite number of situations, we will take the set $\mathcal{S}$ to be finite. As a consequence a so called smoothness condition is guaranteed, the details of which are not relevant to this article (see [KLM90] for precise details).

Preferential structures are being investigated within meta-theories for nonmonotonic reasoning [KLM90, Sho89]. In these investigations, the preference relation, denoted $\sqsubset$ has to do with "normality". For example, the situation in which Tweety is a bird and flies, is considered more normal than a situation where Tweety is a bird and doesn't fly. Here "situation" refers to an interpretation of a standard logic (propositional or first-order predicate, either classical or modal).[1] This logic, which is based on a preferential structure $\langle \mathcal{S}, \sqsubset \rangle$, is termed a *preferential logic* and is denoted by $L_{\sqsubset}$. Irrespective of the intuitive meaning given to the preference relation, the following theorem can be stated [Sho89]:

Theorem 3.1 (Shoham) Let $L_{\sqsubset}$ be a preferential logic such that $\sqsubset$ is smooth. Then $L_{\sqsubset}$ is monotonic if and only if $\sqsubset$ is an equivalence relation.

Monotonicity of logic $\mathrm{L}_{\sqsubset}$ refers to the following property, where i, j and k are formulae of $L_{\sqsubset}$:

$$\frac{i \models k}{i \wedge j \models k}$$

In Shoham's theorem the situations are interpretations of the logic $L_{\sqsubset}$. Nevertheless, we believe that the theorem is very relevant for information retrieval. Up to now, aboutness has been considered as being monotonic with regard to information composition [BH94]. By way of illustration: if we know that information carrier i is about carrier k, then composing information j to i will not disturb the aboutness relation with k. Stating this as a rule yields:

$$\frac{i \models_a k}{i \oplus j \models_a k}$$

[1] The situations considered in information retrieval are probably not of this type. It is a focus of ongoing investigation to to provide a suitable model-theoretic basis for information retrieval. Recent efforts have centred around situation theory [HB94, LR92, Rij93]

Observe that the above rule bears a striking resemblance to the monotonicity property referred to in Shoham's theorem. Now place this theorem within the context of information retrieval and the following question arises. Given that $\sqsubset_N$ in $\langle \mathcal{S}, \sqsubset_N \rangle$ is *not* an equivalence relation (because it is a strict partial order), does this imply that aboutness is *not* monotonic with regard to information composition?

4 Defaults in Information Retrieval

Assume for the moment that there is a searcher with an information need dealing with migration. Furthermore, assume the information need is satisfied by being informed about the migration of birds. Observe that if the searcher enters the query migration, (s)he *expects* information about the migration of birds. In other words, that the migration is of birds is *assumed* i.e. taken as a default. This default is expressed as follows:

$$\mathsf{migration} \quad \mid\!\sim_N \quad \mathsf{bird}$$

which reads "in the light of the information need N, migration is *implicitly* about birds". Observe that the default is parameterized with the information need N. Remember that it is this need that is imposing the ordering on situations.

Defaults arise out of user biases implicit in their cognitive search model. More significantly, defaults imply a preference to be informed about some situations over others. In other words, defaults are intimately tied to the underlying preferential structure. The nature of this relationship will be explored in the rest of this section.

Consider information carrier i. Each situation s that i discriminates can be considered a model of i because situation s "supports" the information carried by i (denoted $s \models i$). That is, we have adopted a situation-theoretic stance; the information carried by i is "true" in situation s [Dev91]. Adapting Shoham's definition [Sho89] allows a so called preferential situation to be introduced as follows:

Definition 4.1

A situation $s, s \in \mathcal{S}$ preferentially satisfies an information carrier i (written $s \models_N i$) if

- $s \models i$
- *there is no situation t such that $t \sqsubset_N s$ and $t \models i$*

The situation s is said to be a preferred *situation which supports i.* ⊡

With regard to the preferential structure depicted in figure 1, the preferred situations are the crane and sparrow situations.

Defaults can now be defined in terms of preferential entailment as follows:

Definition 4.2

An information carrier i preferentially entails information carrier j ($i \mid\!\sim_N j$) if for all situations $s, s \in \mathcal{S}$:

- *if $s \models_N i$ then $s \models j$*

Information carrier i is said to be implicitly about *carrier j in the light of information need N.* ⊡

This definition states that $i \mid\!\sim_N j$ if the preferred situations of i are also situations supporting the information carrier j. For example, the default: migration $\mid\!\sim_N$ bird expresses that preferred migration situations are also bird situations.

5 Reasoning with Defaults under Information Composition

When a searcher wishes to learn about bird migration (s)he typically wants to become informed about such aspects as their migration patterns, where they migrate to and why etc. In terms of a preferential structure the preferred bird migration situations will support information of the above sort. Within this specific information need, the searcher would seemingly not want to be informed about salmon migration. With this as background, consider the following application of compositional monotonicity which is now based on defaults:

$$\frac{\textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird}}{\textsf{salmon} \oplus \textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird}}$$

This example demonstrates that aboutness is *not* monotonic with regard to information composition because the preferred salmon migration situations will not typically be bird situations. Observe that the following is intuitively acceptable:

$$\frac{\textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird}}{\textsf{spring} \oplus \textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird}}$$

What is at work here? It seems that the default $\textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird}$ *precludes* any discussion about salmon in the context of migration. That is,

$$\textsf{migration} \quad \perp_N \quad \textsf{salmon}$$

Note the extension to the notion of information preclusion introduced earlier. Preclusion is now parameterized with the information need. In our example, the above preclusion follows from the preference for bird migration situations. Generalizing from the example leads to the conclusion that preclusion is also tied to the underlying preferential structure.

Definition 5.1

An information carrier i preferentially precludes information carrier j ($i \perp_N j$) if for all $s, s \in \mathcal{S}$:

- *if $s \models_N i$ then $s \not\models j$*

$\boxdot$

In other words, the intersection of the preferred i-situations and the j-situations is empty.

Preferential preclusion relationships open the door to one type of conservative monotonicity with regard to aboutness. *Rational Compositional Monotonicity* states that information composition may only occur when no preclusion relationships are violated.

$$\frac{i \mathrel{\mid\!\sim_N} j \qquad i \not\perp_N k}{i \oplus k \mathrel{\mid\!\sim_N} j}$$

Assuming that part of the migration information need would typically involve some aspect of the seasonal movements of birds, we cannot conclude that all preferred migration situations *do not* support some information about seasons. That is, $\textsf{migration} \not\perp_N \textsf{spring}$. As a consequence, the following is a valid application of Rational Compositional Monotonicity:

$$\frac{\textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird} \qquad \textsf{migration} \not\perp_N \textsf{spring}}{\textsf{spring} \oplus \textsf{migration} \mathrel{\mid\!\sim_N} \textsf{bird}}$$

Cautious Compositional Monotonicity also provides a conservative approach to information composition, namely, it can only occur within the scope of suitable defaults:

$$\frac{i \mathrel{\mid\!\sim_N} j \quad i \mathrel{\mid\!\sim_N} k}{i \oplus j \quad \mathrel{\mid\!\sim_N} k}$$

For example, given the following defaults:

$$\begin{array}{lcl} \textsf{language} & \mid\!\sim_N & \textsf{computer} \\ \textsf{language} & \mid\!\sim_N & \textsf{programming} \\ \textsf{language} & \mid\!\sim_N & \textsf{PASCAL} \end{array}$$

the following are some conclusions using Cautious Compositional Monotonicity:

$$\begin{array}{rcl} \textsf{computer} \oplus \textsf{language} & \mid\!\sim_N & \textsf{programming} \\ \textsf{programming} \oplus \textsf{language} & \mid\!\sim_N & \textsf{PASCAL} \\ \textsf{computer} \oplus (\textsf{programming} \oplus \textsf{language}) & \mid\!\sim_N & \textsf{PASCAL} \end{array}$$

6 Using Default Reasoning for Information Filtering

One of the areas of practical application of this theory is information filtering. The basic idea is the following. A traditional system is used to retrieve an initial set of information objects. The filter returns only returns those objects from this set whereby the query is a nonmonotonic consequence of the associated characterization. More formally,

$$res(q) \quad = \quad \{o \mid \chi(o) \mid\!\sim_N q\}$$

where $\chi(o), \chi \subset \mathcal{L}_T$ denotes the characterization of information object o. It remains to define what is meant by *nonmonotonic consequence*. This is the set of information carries that can be concluded using a set of rules or axioms. We are currently investigating the following set: Cautious Compositional Monotonicity, Rational Compositional Monotonicity, Reflexivity ($i \mid\!\sim_N i$) and Default Consistency. In contrast to Reflexivity, Default Consistency is not a theorem of the underlying model theory (Definitions 4.1,4.2,5.1). The rule is in essence normative; it expresses the constraint that the left hand side of defaults should not preclude each other:

$$\frac{i \mid\!\sim_N k \quad j \mid\!\sim_N k}{i \not\perp_N j}$$

To illustrate how nonmonotonic inference can be used for information filtering, consider the following selection of information fragments which were retrieved from the Internet in response to the query surfing.

f_1 = WWW Surfing
f_2 = Net Surfing
f_3 = Surfing Mosaic
f_4 = Surfing the Net
f_5 = surfing the Web
f_6 = surfing conditions in California

Assume that these fragments have the following characterizations:

$$\begin{array}{lcl} \chi(f_1) & = & \{\textsf{WWW} \oplus \textsf{surfing}\} \\ \chi(f_2) & = & \{\textsf{net} \oplus \textsf{surfing}\} \\ \chi(f_3) & = & \{\textsf{surfing} \oplus \textsf{mosaic}\} \\ \chi(f_4) & = & \{\textsf{surfing} \oplus \textsf{net}\} \\ \chi(f_5) & = & \{\textsf{surfing} \oplus \textsf{web}\} \\ \chi(f_6) & = & \{\textsf{surfing} \oplus (\textsf{conditions} \oplus \textsf{california})\} \end{array}$$

In this search the seeker was interested to find out about wave surfing conditions around the world, so implicit in this information need is the default:

$$\textsf{surfing} \quad \mid\!\sim_N \quad \textsf{conditions}$$

as all preferred surfing situations will support information about surfing conditions. Note also that all preferred Californian situations will be about surfing, hence,

$$\mathsf{california} \quad \mathrel{|\!\sim}_N \quad \mathsf{surfing}$$

To show the desired result of fragment f_6 being in the result set, we must show $\chi(f_6) \mathrel{|\!\sim}_N \mathsf{surfing}$. This is done as follows:

Proof:

Reflexivity renders $\mathsf{surfing} \mathrel{|\!\sim}_N \mathsf{surfing}$ and as a consequence of Definition 5.1: $\mathsf{surfing} \not\perp_N \mathsf{surfing}$. Now using Rational Compositional Monotonicity:

$$\frac{\mathsf{california} \mathrel{|\!\sim}_N \mathsf{surfing} \qquad \mathsf{surfing} \not\perp_N \mathsf{surfing}}{\mathsf{surfing} \oplus \mathsf{california} \mathrel{|\!\sim}_N \mathsf{surfing}}$$

Using Cautious Compositional Monotonicity:

$$\frac{\mathsf{surfing} \mathrel{|\!\sim}_N \mathsf{conditions} \qquad \mathsf{surfing} \mathrel{|\!\sim}_N \mathsf{surfing}}{\mathsf{surfing} \oplus \mathsf{conditions} \mathrel{|\!\sim}_N \mathsf{surfing}}$$

Application of Default Consistency yields:

$$\mathsf{surfing} \oplus \mathsf{california} \quad \not\perp_N \quad \mathsf{surfing} \oplus \mathsf{conditions}$$

Rational Monotonicity can now once again be applied:

$$\frac{\mathsf{surfing} \oplus \mathsf{california} \mathrel{|\!\sim}_N \mathsf{surfing} \qquad \mathsf{surfing} \oplus \mathsf{california} \not\perp_N \mathsf{surfing} \oplus \mathsf{conditions}}{(\mathsf{surfing} \oplus \mathsf{california}) \oplus (\mathsf{surfing} \oplus \mathsf{conditions}) \mathrel{|\!\sim}_N \mathsf{surfing}}$$

Use of distributivity on the LHS renders the desired result:

$$\mathsf{surfing} \oplus (\mathsf{conditions} \oplus \mathsf{california}) \quad \mathrel{|\!\sim}_N \quad \mathsf{surfing}$$

$\boxdot$

As the searcher is interested in wave surfing, *not* Internet surfing so the following preclusions are apparent:

$$\mathsf{surfing} \quad \perp_N \quad \mathsf{web}$$
$$\mathsf{surfing} \quad \perp_N \quad \mathsf{WWW}$$
$$\mathsf{surfing} \quad \perp_N \quad \mathsf{net}$$
$$\mathsf{surfing} \quad \perp_N \quad \mathsf{mosaic}$$

These insure that none of the fragments $f_1 \ldots f_5$ will be included in the result set as information carriers such as $\mathsf{net} \oplus \mathsf{surfing}$ are not nonmonotonic consequences.

7 Summary and Further Research

This article has demonstrated that aboutness shows nonmonotonic character under information composition. This result is not surprising as it parallels the existence of information retrieval systems based on (network-based) probabilistic inference. It is more important to consider the consequences of this discovery. Information Retrieval systems must be conservative with regard to information composition, for example, in query expansion. Up till now, there has been little theory in this area. This article provides some foundation that may help fill this gap.

This article has not addressed the question of where the defaults come from. We see two approaches as being worth investigating. The first approach is to glean initial defaults from interactions with the searcher and further refine them as a part of a relevance feedback. This approach has the advantage that the initial defaults could be tailored to the searcher's given information need. A second approach would be to have a set of predefined defaults like those being used in the Cyc project [LGP+90]. The assumption underlying this approach is that these defaults would be applicable for a large percentage of searchers.

As mentioned earlier, information retrieval lacks a model-theoretic framework. Such a framework is necessary to scrutinize aboutness inference mechanisms. This article has presented the rudiments of such a framework. It borrows from an information-based theory to logic, namely situation theory, as well as the nonmonotonic consequence relation approach found in artificial intelligence. We believe that this combination is potentially very fruitful and aim to extend the current framework to gain a deeper understanding of aboutness. In particular, we plan to integrate this work with the aboutness rules/axioms presented in our previous work. Our ultimate goal is to produce an aboutness preserving inference mechanism founded on a suitable model-theoretic framework. Such a mechanism could form the brains of intelligent information agents.

References

[BH94] P.D. Bruza and T.W.C. Huibers. Investigating Aboutness Axioms using Information Fields. In *Proceedings of the Seventeenth ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 112–121, 1994.

[Bru93] P.D. Bruza. *Stratified Information Disclosure: A Synthesis between Information Retrieval and Hypermedia.* PhD thesis, University of Nijmegen, 1993.

[Coo71] W.S. Cooper. A Definition of Relevance for Information Retrieval. *Information Storage and Retrieval*, 7:19–37, 1971.

[Dev91] K. Devlin. *Logic and Information.* Cambridge University Press, 1991.

[HB94] T.W.C. Huibers and P.D. Bruza. Situations: A general framework for studying Information Retrieval. In *Proceedings of the 16th BCS Information Retrieval Colloquium.* British Computer Society, Springer-Verlag, 1994.

[KLM90] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence*, 44:167–207, 1990.

[Lan86] F. Landman. *Towards a Theory of Information.* Foris, 1986.

[LD94] M. Lauer and M. Dras. A Probabilistic Model of Compound Nouns. In *Proceedings of the Seventh Australian Joint Conference on Artificial Intelligence (AI94)*, pages 474–481. World Scientific, 1994.

[LGP+90] D. Lenat, R. Guha, K. Pittman, D. Pratt, and M. Shepherd. Cyc: Toward Programs with Common Sense. *CACM*, 33(8), 1990.

[LR92] M. Lalmas and C.J. van Rijsbergen. A Logical Model of Information Retrieval based on Situation Theory. In *Proceedings of the 14th BCS Information Retrieval Colloquium.* British Computer Society, Springer-Verlag, 1992.

[Rij86] C.J. van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29(6):481–485, 1986.

[Rij93] C.J. van Rijsbergen. What is Information Anyway? In *Two Essays in Information Retrieval.* University of Glasgow, 1993. Research Report IR-93-03.

[Sho89] Y. Shoham. Efficient Reasoning about Rich Temporal Domains. In R.H. Thomason, editor, *Philosophical Logic and Artificial Intelligence*, pages 191–222. Kluwer Academic Publishers, 1989.