THE ROLE OF EXPECTATIONS IN REASONING

Peter Gärdenfors

Lund University Cognitive Science
Kungshuset, Lundagård
S–223 50 Lund
Sweden
E-mail: Peter.Gardenfors@fil.lu.se

Abstract: Logical theory traditionally assumes the following: (1) Logical inference is a relation between sentences (or propositions), not between thoughts (or anything cognitive). (2) The validity of an argument is only dependent on the logical structure of the sentences and independent of their meaning.

In practical reasoning, however, these assumptions are not valid. In this paper I want to show that by taking expectations into account, one can achieve a much better understanding of how logic is put to work by humans. In particular, one obtains a very simple analysis of nonmonotonic reasoning. I will also discuss the cognitive origins of expectations.

Then a man said: Speak to us of Expectations.

He then said: If a man does not see or hear the waters of the Jordan, then he should not taste the pomegranate or ply his wares in an open market.

If a man would not labour in the salt and rock quarries, then he should not accept of the Earth that which he refuses to give of himself.

Such a man would expect a pear of a peach tree.

Such a man would expect a stone to lay an egg.

Such a man would expect Sears to assemble a lawn mower.

Kehlog Albran, The Profit

1. UBIQUITOUS EXPECTATIONS

We all have expectations. For example, you presumably expect a speaker at a scientific meeting to be standing properly on his feet while talking. You may also expect that he is not red in his face, although this expectation may be much weaker than the previous one.

Sometimes, however, our expectations are violated. If a speaker is standing on his hands while giving his presentation, you would, I venture, be surprised. But given this violation of your previous expectations, you would now expect him to be red in his face, even if you don't observe it. This example shows that we are able to reason and draw conclusions even when our primary expectations are gainsaid.

As a matter of fact, expectations are ubiquitous, although they are not often made explicit. You expect there to be a floor when you enter a room; you expect a door handle not to break when you press it; you expect your morning newspaper to arrive on time; and you don't expect Sears to assemble a lawn mower. The main thesis of this article is that expectations play a crucial role in everyday reasoning. In brief, we use them as supplementary premises when we make inferences. In particular, I will argue that much of *nonmonotonic logic* can be reduced to classical logic with the aid of an analysis of the expectations that are hidden in the arguments. I will also discuss the cognitive origins of expectations.

In classical logic, the role of expectations is eschewed. My diagnosis of this neglect, begins by bringing out into the open some of the philosophical and methodological assumptions that has governed the traditional approach to logic since Frege.

Classical assumption 1: Logical inference is a relation between sentences (or propositions), not thoughts (or anything else related to cognition).

According to the traditional view, logical arguments are described as a relation between a set Γ of premises

and a conclusion C, in symbols Γ 7 C. The premises and conclusions are expressed as sentences in some language, preferably a formal language.

What kind of relation is 7? Traditionally this is answered as follows:

Classical assumption 2: The validity of an argument is only dependent on the logical structure of the sentences in Γ and C and independent of their meaning, their truth or the context.

The inference relation 7 can be specified in two major ways: (i) proof-theoretically, by specifying the axioms and derivations rules that generate 7;[1] and (ii) semantically, by formulating the truth conditions for the logical relation to hold. In either case, the assumption 2 is valid.

However, what makes something an expectation and not an axiom or a standard premise is that it is *defeasible*. If an expectation is in conflict with a premise, it yields. Consequently, expectations don't fit with the classical assumptions.

Hence expectations are suppressed in classical logic, but in practical reasoning they are not. Everyday arguments are full of hidden premises that need to be made explicit in order to make the argument logically valid. In each separate case, it may be possible to add the hidden assumptions to make the derivation comply with the classical assumptions. But nobody does this. The reason for this is that expectations are normally shared among speakers, and unless they are countervened, they serve as a common background for arguments.

Consequently, if we want to describe a logical inference relation that better conforms to everyday reasoning, we need a notion of inference that is not constrained by the classical assumptions. So if these assumptions are abandoned, a crucial question becomes: *What can be used as premises in an argument?* For classical logic, the answer is: only the set Γ. For practical reasoning the answer may be different. As I shall argue in Section 3, the answer to this question may help us understand much of what has been called nonmonotonic logic.

2. LURIA'S CAMELS

However, before turning to an analysis of the role of expectations in nonmonotonic reasoning, let me present an extreme case (from the classical point of view) of handling premises. Consider the following dialogue taken from Luria (1976, p. 112):

> Subject: Nazir-Said, age twenty-seven, peasant from village of Shak-himardan, illiterate.
> The following syllogism is presented: **There are no camels in Germany. The city of B. is in Germany. Are there any camels or not?**
> Subject repeats syllogism exactly.
> **So, are there camels in Germany?**[2]
> "I don't know, I've never seen German villages."
> *Refusal to infer.*
> The syllogism is repeated.
> "Probably there are camels there."
> **Repeat what I said.**
> "There are no camels in Germany, are there any camels in B. or not? So probably there are. If it is a large city, there should be camels there."
> **But what do my words suggest?**
> "Probably there are. Since there are large cities, there should be camels."
> *Again a conclusion apart from the syllogism.*
> **But if there aren't any in all of Germany?**
> "It is a large city, there will be Kazakhs or Kirghiz there."
> **But I am saying that there are no camels in Germany.**
> "If this village is in a large city, there is probably no room for camels.
> *Inference made apart from syllogism.*

On the basis of this and a number of similar interviews, Luria draws the conclusion that illiterate people are far inferior to literates when it comes to logical reasoning.[3] However, I don't agree with his diagnosis. I believe that what is at stake here is exactly the question of what is allowed as premises in an argument.

In contrast to Luria, my hypothesis is that the peasants that he is interviewing don't allow themselves to use anything as a premise for an argument unless they have *personal experience* of the premise. So in the dialogue above, Nazir-Said *ignores* the information provided in the syllogism, since he no direct knowledge of the matter ("I don't know, I've never seen German villages)".

In my opinion, the situation becomes very clear in a dialogue like the following (Luria 1976, p. 108–109):

> The following syllogism is presented: **In the Far North, where there is snow, all bears are white. Novaya Zemlya is in the Far**

[1] Some of the axioms may be 'non-logical', which normally means that they are taken from a 'theory'.

[2] Luria's text probably contains an error here. The question should have been "Are there any camels in B.?"

[3] Also cf. Luria (1979).

North and there is always snow there. What colors are the bears there?

"There are different sorts of bears."

The syllogism is repeated.

"I don't know: I have seen a black bear, I've never seen any others ... Each locality has its own animals: if it's white, they will be white; if it's yellow, they will be yellow."

Appeals only to personal, graphic experience.

But what kind of bears are there in Novaya Zemlya?

"We always speak only of what we see; we don't talk about what we haven't seen."

The same.

But what do my words imply? The syllogism is repeated.

"Well, it's like this: our tsar isn't like yours, and yours isn't like ours. Your words can be answered only by someone who was there, and if a person wasn't there he can't say anything on the basis of your words."

The same.

But on the basis of my words – in the North, where there is always snow, the bears are white, can you gather what kind of bears there are in Novaya Zemlya?

"If a man was sixty or eighty and had seen a white bear and had told about it, he could be believed, but I've never seen one and hence I can't say. That's my last word. Those who saw can tell, and those who didn't see can't see anything!" (At this point a young Usbek volunteered, "From your words it means that bears there are white.")

Well, which of you is right?

"What the cock knows how to do he does. What I know, I say, and nothing beyond that!"

The upshot is that Luria's evidence does not indicate that there is anything wrong with the logical abilities of the illiterate peasants. The main difference is that they don't take a statement provided by a stranger as something that can be used in reasoning; *only personal experience is allowed* (there seem to be strong moral feelings about this among the Uzbeks).

The illiterate Uzbeks thus violate the second classical assumption since the validity of an argument is extremely dependent on the personal experience of the person who presents the conclusion, to the extent that the explicit premises in Γ may be totally ignored. Thus we have here a notion of logical validity that is quite different from the classical. This does, however, not mean that the illiterates in Luria's investigation are illogical – it only means that they are reasoning by other rules.

Luria's interviews show that when people learn to read, they will learn to play a different logical game. My understanding, which does not completely fit with Luria's, is that when becoming literate one learns to see the text as an abstract entity, independent of a particular speaker and his practical experience and his motives for proclaiming the words. The text is seen as a symbolic structure. It is only in relation to such a symbolic entity that the classical assumptions make sense. For somebody who only hears spoken words, uttered by a particular person, it is much more difficult (and of little practical importance) to view the words as symbolic structures. In other words, literacy teaches us to separate abstract arguments from their practical context.

I have presented my analysis of Luria's interviews as an extreme case of deciding which premises may be used in a logical argument – the case when even the premises explicitly stated in Γ may be disregarded. The point of the example is that, when practical reasoning is concerned, one must be careful in specifying what counts as a premise for an argument.

3. NONMONOTONIC REASONING BASED ON EXPECTATIONS

In some recent articles, David Makinson and I have argued that the areas of nonmonotonic logic and belief revision are very closely related (see Makinson and Gärdenfors (1990), Gärdenfors (1990), (1991a), and Gärdenfors and Makinson (to appear)). In particular, we show in Gärdenfors (1991a), and Gärdenfors and Makinson (to appear) how various forms of nonmonotonic inferences can be given a unified treatment in terms of how *expectations* are used in reasoning. This section begins with a summary of that analysis, but also discusses some limitations of the assumptions.

3.1 Motivation

The guiding idea is that when we try to find out whether C follows from Γ, the background information that we use for the inference does not only contain the premises in Γ, but also information about what we *expect* in the given situation. For instance, if we know that someone is a Spanish woman, we anticipate her to be dark and temperamental. Such expectations can be expressed in different ways: by default assumptions, statements about what is normal or typical, etc. These expectations are not premises that have to be accepted, but they are *defeasible* in the sense that if the premises Γ are in conflict with some of the expectations, we don't use them when determining whether C follows from Γ.

I want to show that expectations are used basically in the same way as explicit premises in logical arguments; the difference is that the expectations are,

in general, more defeasible than the premises.[4] Consequently, the expectations used in nonmonotonic inferences need no special notation, but they can be expressed in the same language as regular beliefs. This is one side of the unified treatment of nonmonotonic reasoning. For simplicity I shall work with a standard propositional language L which will be assumed to be closed under applications of the *boolean connectives* ¬ (negation), & (conjunction), ∨ (disjunction), and ∅ (implication). I will use α, β, γ, etc. as variables over sentences in L. I will assume that the underlying logic includes *classical propositional logic* and that it is *compact*. Classical logical consequence will be denoted by 7 and the set of classical consequences of a set Γ will be denoted Cn(Γ).

In this section and the following, all the different expectations will be formulated in L. In contrast to many other theories of nonmonotonic reasoning there are thus no default rules or other additions to the basic language, such as modal operators, that will be used to express the defeasible forms of information. Another, non-propositional, way of handling expressions will be presented in Section 6.

The key idea behind nonmonotonic reasoning can be put informally as follows:[5]

α nonmonotonically entails β iff β follows logically from α together with "as many as possible" of the set of our expectations as are compatible with α.

In order to makes this more precise, we must, of course, specify what is meant by "as many as possible".[6] But before turning to technicalities, let me illustrate the gist of the analysis by a couple of examples. "α nonmonotonically entails β" will, as usual, be denoted α □ β .

As a first example, let the language L contain the following predicates:

Sx: x is a speaker at a conference
Hx: x is standing on his hands
Rx: x is red in the face

Assume that the set of expectations contains
Sb ∅ ¬Rb and Sb & Hb ∅ Rb, for all individuals b. Assuming that the set of expectations is closed under logical consequences it also contains Sb ∅ ¬Hb and, of course, the logical truth Sb & Hb ∅ Sb. If we now learn that b is a speaker at a conference, that is Sb, this piece of information is consistent with the expectations and thus we can conclude that Sb □ ¬Rb according to the recipe above.

On the other hand, if we learn both that b is a speaker and, surprisingly enough, is standing on his hands, that is Sb & Hb, then this information is *inconsistent* with the set of expectations and so we cannot use all expectations when determining which inferences can be drawn from Sb & Hb. The most natural expedient is to give up the expectation Sb ∅ ¬Rb and the consequence Sb ∅ ¬Hb. The contracted set of expectations which contains Sb & Hb ∅ Rb and its logical consequences, in a sense contains "as many as possible" of the sentences in the set of expectations that are compatible with Sb & Hb. So, by the general rule above, we have Sb & Hb □ Rb. This shows that □ is indeed a nonmonotonic inference operation.

3.2 Expectation orderings

Expectations function as hidden assumptions. However, when evaluating their role in arguments it is important to note that our expectations about the world do not all have the same strength. For example, we consider some rules to be almost universally valid, so that an exception to the rule would be extremely unexpected; while other rules are better described as rules of thumb that we use for want of more precise information. An exception to the latter type of rule is not unexpected to the same degree as in the former case. In brief, our expectations are all defeasible, but they exhibit varying *degrees of defeasibility*.

In order to make these ideas more precise, I shall assume that there is an ordering ≤ of the sentences in L. '$\alpha \leq \beta$' should be interpreted as 'β is at least as expected as α' or 'α is at least as surprising as β'.
'$\alpha < \beta$' will be written as an abbreviation for
'not $\beta \leq \alpha$' and '$\alpha \approx \beta$' is an abbreviation for
'$\alpha \leq \beta$ and $\beta \leq \alpha$'.

According to the key idea of this section α □ β means that β follows from α together with all the propositions that are 'sufficiently well' expected in the light of α. How well is 'sufficiently well'? A natural idea is to require that the added sentences be strictly more expected than ¬α in the ordering. This is the motivation for the following definition.

Definition. □ is an *expectation* inference relation iff there is an ordering ≤ satisfying (E1) – (E3) such that the following condition holds:

(C□) $\quad \alpha \mid\!\sim \gamma$ iff $\gamma \in Cn(\{\alpha\} \approx \{\beta: \neg\alpha < \beta\})$

To a large extent, the formal properties of the nonmonotonic inference relation defined in this way depends on the properties that are assumed to hold for

[4] But cf. the examples from Luria above.
[5] I will confine the analysis to the case where there are only finitely many premises which can be conjoined to a single α. However, as shown by Freund, Lehmann and Makinson (1990), there is a canonical way of extending any such finitary relation to cover infinite sets of premises.
[6] This idea is related to the idea of 'minimal change' within the theory of belief revision (see Gärdenfors (1988), pp. 66–68).

the ordering <. Gärdenfors and Makinson (to appear) assume that it satisfies the following postulates:

(E1) If $\alpha \leq \beta$ and $\beta \leq \gamma$, then $\alpha \leq \gamma$ (*Transitivity*)

(E2) If α 7 β, then $\alpha \leq \beta$ (*Dominance*)

(E3) For any α and β, $\alpha \leq \alpha$ & β or $\beta \leq \alpha$ & β (*Conjunctiveness*)

The first postulate on the expectation ordering is very natural for an ordering relation. The second postulate says that a logically stronger sentence is always less expected. From this it follows that the relation $\leq$ is reflexive. The third constraint is crucial for the representation results proved in Gärdenfors and Makinson (to appear), but presumably the one that is most open to query. It concerns the relation between the degree of expectation of a conjunction α & β and the corresponding degrees of α and β.

Note that the three conditions imply *connectivity*: either $\alpha \leq \beta$ or $\beta \leq \alpha$. For by (E3) and (E2) either $\alpha \leq \alpha$ & $\beta \leq \beta$ or $\beta \leq \alpha$ & $\beta \leq \alpha$ and we conclude by (E1). From (E2) it also follows that α & $\beta \leq \alpha$ and α & $\beta \leq \beta$, so (E3) entails that α & $\beta \approx \alpha$ or α & $\beta \approx \beta$. This means that we cannot interpret the degrees of expectation directly in terms of their *probabilities*, since (E3) is violated by any probability measure. The word 'expectation' as it is used in this paper should thus not be confused with the notion of 'expected utility' in decision theory. 'Expected utility' has to do with expectations of the *values* of various outcomes, while the notion of expectation studied here concerns *beliefs* about the world. In my opinion, this use of 'expectation' comes much closer to the everyday use.

Recalling that by the three conditions on expectation orderings we have $\neg\alpha \leq \beta_i$ for all $i \leq n$ iff $\neg\alpha \leq \beta_1$ & ... & β_n, it is immediate, using the compactness of Cn, that (C□) is equivalent to:

(C□') α □ γ iff either α 7 γ or there is a β □ L with α & β 7 γ and $\neg\alpha < \beta$

This condition may be surprising. It says that γ follows from α if there are some expectations that are consistent with α which together with α *classically* entails γ. In other words: *Nonmonotonic logic is nothing but classical logic if relevant expectations are added as explicit premises!* I believe that this observation can remove a lot of the mystery surrounding nonmonotonic inferences. If the analysis presented here is correct, a lot of the paraphernalia of nonmonotonic logic will not be required anymore. Among other things, one needs no new notation for defaults,[7] no inference rules that are special to nonmonotonic logics, and no particular model theory.

3.3. Weaker assumptions about expectations

Gärdenfors and Makinson (to appear) prove that expectation inference relations, which are based on expectation orderings fulfilling (E1) – (E3), satisfy a number of postulates for nonmonotonic inferences. One of the strongest postulates is the following:

Rational Monotony:
If α ¸ $\neg\beta$ and α □ γ, then α & β □ γ.

This postulate cannot be proved without assuming that the strong condition (E3) holds for the expectation ordering. Conversely, in the completeness proof for expectation inference relations based on orderings fulfilling (E1) – (E3), the proof that (E3) is fulfilled makes essential use of Rational Monotony.

Like (E3), Rational Monotony is a strong postulate, the validity of which is sometimes challenged.[8] For example, Ginsberg (1986) presents a counterexample to a corresponding principle for conditionals which can also be used against Rational Monotony. His example involves the following statements:

α: Verdi is not French.
β: Bizet is French.
γ: Satie is French.
δ: Bizet and Verdi are compatriots.
ε: Bizet and Satie are compatriots.

In the example, α, β, and γ are seen as background facts, that is, sentences that have a strong degree of expectation (and we may assume that they have roughly the same degree of expectation.

Now, let us first assume δ as a premise and consider the nonmonotonic consequences of this assumption. Since δ is inconsistent with the conjunction of α and β, at least one of these expectations must be given up. However, it is not certain that β is given up, and consequently one cannot conclude $\neg\varepsilon$ from δ, so that we have δ ¸ $\neg\varepsilon$. Furthermore, since δ and γ are independent statements, the addition of δ does not affect the validity of γ. Hence it is reasonable to suppose that δ □ γ.

Next, let us start from δ & ε as a premise, i.e., that all three composers are compatriots. In this inferential situation, at least one of the expectations α, β, or γ must be rejected. However, since they are assumed to be of equal strength, it may be that γ is rejected. Consequently, it does not hold that δ & ε □ γ. Summing up, we have δ □ γ and δ ¸ $\neg\varepsilon$, but not

[7] Defaults will be analyzed in terms of expectations in the following section.
[8] For a defence, see Lehmann and Magidor (1992).

$\delta \,\&\, \varepsilon \;\vdash\; \gamma$, which gives us a counterexample to Rational Monotony, and, indirectly, a counterexample to (E3).

The intuitive validity of the Ginsberg's counterexample depends on the fact that δ and γ are *independent* statements. The notion of independence is difficult to formalize,[9] but in the present context it can at least be stated that independence entails absence of expectations in the sense that if δ and γ are independent, then $\delta \;\vdash\; \gamma$ if and only if $\not\vdash \gamma$ and $\gamma \;\vdash\; \delta$ if and only if $\not\vdash \delta$.

As noted above, it follows from (E3) that the expectation ordering is total. However, the notion of independence requires that some sentences not be comparable with regards to their degree of expectation. Since the validity of (E3) is tightly connected to the validity of Rational Monotony, this is then the cause of the violation of Rational Monotony in Ginsberg's example. For these reasons, it interesting to study weaker versions of (E3) that do not entail that the expectation ordering be total, but, say, only that it be a partial ordering. In particular, Rott (1992) investigates the following two principles:

(E3?) If $\alpha < \beta$ and $\alpha < \gamma$, then $\alpha < \beta \,\&\, \gamma$

(E3|) If $\alpha \,\&\, \beta < \beta$, then $\alpha < \beta$

In connection with belief revision procedures, he is able to show some representation theorems involving these principles. Neither of his theorems utilizes the equivalence of Rational Monotony (i.e. the postulate K*8 from Gärdenfors (1988)).[10] It remains an open question as to the extent his results can be transferred to the context of nonmonotonic inferences.

4. DEFAULTS AS EXPECTATIONS

One of the main motivations for studying nonmonotonic reasoning is that this kind of theory is necessary if we want to understand reasoning by *default assumptions*. In this section I want to show that an ordering of expectations contains enough information to express, in a very simple way, what we require with respect to default information.[11] The principal idea is that a default statement of the type 'F's are normally G's' can be expressed by saying that 'if something is an F, then it is less expected that it is non-G than that it is G'. This formulation is immediately representable in an expectation ordering by assuming that the relation $Fb \varnothing \neg Gb < Fb \varnothing Gb$ holds for all individuals b.

To illustrate the general idea of expressing defaults of the form 'F's are normally G's' as an expectation relation $Fb \varnothing \neg Gb < Fb \varnothing Gb$ for all individuals b, assume that all we know about b is that Fb. We want to decide the nonmonotonic consequences of this fact. It can be determined, via (C$\vdash$), that $Fb \;\vdash\; Gb$. It can also be determined that $Fb \;\not\vdash\; \neg Gb$. Further information about b, for example that Hb, will mean that we no longer need to check whether $Fb \varnothing \neg Gb < Fb \varnothing Gb$, but rather whether $Fb \,\&\, Hb \varnothing \neg Gb <$
$Fb \,\&\, Hb \varnothing Gb$, which may give a different answer. This is exactly how we want a default rule to operate.

To give an analysis of a familiar example, the so called Nixon diamond, suppose that L contains the following predicates:

Rx: x is a republican
Qx: x is a quaker
Px: x is pacifist

Assume that we have the default rules "republicans are normally not pacifists" and "quakers are normally pacifists." According to the rule given above, we express these defaults by a number of ordering relations of the form $Rb \varnothing Pb < Rb \varnothing \neg Pb$ and $Qb \varnothing \neg Pb < Qb \varnothing Pb$, respectively, for various individuals b.

From this we conclude, as above, that if all we know about McCarthy is that he is a republican, then we expect him to be a non-pacifist (and we don't expect him to be a quaker); and if all we know about Fox is that he is a quaker, then we expect him to be a pacifist (and don't expect him to be a republican). Now, suppose that, contrary to our expectations, Nixon is a both a quaker and a republican, that is $Qa \,\&\, Ra$. What can be concluded concerning his pacifism?

If we know that $Qa \,\&\, Ra$ and we want to decide whether $\neg Pa$ or Pa follows nonmonotonically, then this can be determined, via (C$\vdash$), by looking for the strictly greater of $Qa \,\&\, Ra \varnothing \neg Pa$ and $Qa \,\&\, Ra \varnothing Pa$ in the expectation ordering. Three cases are possible: (1) $Qa \,\&\, Ra \varnothing \neg Pa < Qa \,\&\, Ra \varnothing Pa$. In this case, we conclude that $Qa \,\&\, Ra \;\vdash\; Pa$. (2) $Qa \,\&\, Ra \varnothing Pa <$
$Qa \,\&\, Ra \varnothing \neg Pa$. For similar reasons, we conclude that $Qa \,\&\, Ra \;\vdash\; \neg Pa$. (3) $Qa \,\&\, Ra \varnothing Pa \approx$
$Qa \,\&\, Ra \varnothing \neg Pa$. In this case (or in the case when they are incomparable, if the ordering is not supposed to be total), then neither $Qa \,\&\, Ra \;\vdash\; \neg Pa$, nor
$Qa \,\&\, Ra \;\vdash\; Pa$ will hold.

None of these three possibilities is ruled out by the two ordered pairs $Ra \varnothing Pa < Ra \varnothing \neg Pa$ and
$Qa \varnothing \neg Pa < Qa \varnothing Pa$. The reason is that it follows from (E2) that $Ra \varnothing \neg Pa \leq Qa \,\&\, Ra \varnothing \neg Pa$ and that $Qa \varnothing Pa \leq Qa \,\&\, Ra \varnothing Pa$. Consequently, the maximum of $Qa \,\&\, Ra \varnothing \neg Pa$ and $Qa \,\&\, Ra \varnothing Pa$ will

[9] However, see Gärdenfors (1978) for a general analysis and Gärdenfors (1991b) for an application of this analysis to belief revision processes.
[10] For a survey of some of the results concerning other principles for expectation orderings, also cf. Gärdenfors and Rott (to appear).
[11] Cf. Morreau (1992) for a related analysis of defaults.

be at least as high as each of Ra ∅ ¬Pa and Qa ∅ Pa in the expectation ordering. But on the other hand, the two comparisons do not suffice to determine which, if any, of Qa & Ra ∅ ¬Pa and Qa & Ra ∅ Pa is the greater. So, the information available does not permit us to conclude anything concerning ¬Pa or Pa.

To sum up, the nonmonotonic consequences one can draw from the premise that Qa & Ra depends on which is chosen to be the maximal element of Qa & Ra ∅ ¬Pa and Qa & Ra ∅ Pa in the expectation ordering. The default relations Ra ∅ Pa < Ra ∅ ¬Pa and Qa ∅ ¬Pa < Qa ∅ Pa are not sufficient to determine this choice.

5. BUT

We have expectations, but we are sometimes surprised. The choice of the word "but" instead of "and" in the previous sentence indicates that the information contained in the second half of the sentence *violates our expectations*. This, I want to argue, is the core meaning of "but".

In introductory courses in logic, one often uses the formalism of propositional logic to analyse the conjunctions of natural language. It is shown how words like "and", "or", "not", "if ..., then", "unless", "even if", etc. can be expressed in formulas. But "but" is seldomly given a proper analysis. At best it is said that it has the same logical meaning as "and". For any user of language, it should be obvious that this is false. Among linguists, it is commonplace that "but" expresses a violation of expectations.[12] However, they leave it at that, since they have no way of representing and analysing expectations.

Using the tools of the previous two sections, the analysis of "but" I want to propose is the following:

(CBut) *A sentence of the form "α but β" is acceptable in a context C if and only if α and β are both acceptable in C, and in C it holds that $\alpha \sqsupset \neg\beta$.*

I don't propose any truth conditions for "but", simply because I don't think there are any. As will be clear soon, the use of but is very context sensitive, so I believe the proper analysis of "but" should be in terms of the conditions under which a sentence is *accepted* in a given context. The context also determines what the current expectations are.

Let me apply the analysis to some examples:

(1) She is rich and ugly.

(2) She is rich but ugly.

The difference between the content of (1) and (2) is that in (2) the speaker presupposes that rich women are normally not ugly, while such an expectation is not indicated in (1). The use of "but" *signals* an expectation. However which expectation is signalled is not determined from the sentence alone, but depends on the whole context.

The role of the context can sometimes be quite subtle. Compare the following two sentences from Robin Lakoff (1971), p. 133:

(3) John hates ice-cream but so do I.

(4) John hates ice-cream but I like it.

On a standard reading of (3), it carries the expectation that people normally like ice-cream. The "but" indicates no contrast between me and John, but a contrast between me and people in general. However, in a natural context where (4) is uttered, there is no expectation concerning how normal people like ice-cream. On the other hand, mentioning John's dislike for ice-cream creates an expectation, albeit a weak one, that I too should dislike it and it is this "inductively generated" expectation that is denied by the "but". The same argument applies to another of Lakoff's examples:

(5) John is tall but Bill is short.

She calls examples like (4) and (5) instances of "semantic oppositions" and argues that it is one of the functions of "but" to express such oppositions (1971, p. 133). However, I find it more natural to view sentences (4) and (5) as a special kind of denial of expectations, which is the other meaning of "but" that Lakoff identifies and which is expressed more formally in my analysis above.

In this section, I have outlined an analysis of the meaning of "but" based on expectations. In addition to this application, I believe that a large part of the discussion within linguistics and philosophy concerning *presuppositions* of sentences can be given a more unified treatment in terms of the expectations of the speaker. However, such an analysis will not be attempted here.

6. HOW ARE EXPECTATIONS TO BE REPRESENTED?

Expectations have, so far, been treated as primitive notions. But where do they come from? In this final

[12] For example Lakoff (1971). She distinguishes (p. 133) between two main meanings of "but": (1) semantic opposition as in "John is tall but Bill is short"; and (2) denial of expectation as in "John is tall but he he's no good at basketball". However, as I will argue later, the semantic opposition meaning is a special case of violated expectation.

section I will discuss the origins of expectations and alternative ways of modelling them.

In Section 3, expectations were modelled by expectation orderings which are orderings of propositions. An important epistemological question for the analysis presented in that section is how this ordering is determined. In Gärdenfors and Makinson (to appear) it is shown that it is possible to define an expectation ordering by using a nonmonotonic inference operation by the following equation:

(C≤) $\alpha \leq \beta$ iff either $\alpha \,\&\, \beta \in Cn(\square)$ or $\neg(\alpha \,\&\, \beta) \mathrel{|\!\sim} \alpha$.

The case when $\alpha \,\&\, \beta \in Cn(\square)$ is just the limiting case when $\alpha \,\&\, \beta$ is logically valid. The main case when $\neg(\alpha \,\&\, \beta) \mathrel{|\!\sim} \alpha$ means basically that if $\alpha \,\&\, \beta$ is expected and we assume that $\neg(\alpha \,\&\, \beta)$, then α is no longer expected, which is the criterion for α being less expected than β. In Gärdenfors and Makinson (to appear), we prove in Theorem 3.3 that if $\square$ is any inference relation that satisfies the full set of postulates, including Rational Monotony, then the ordering $\leq$ defined by (C≤) is indeed an expectation ordering over L that satisfies (E1) – (E3).

However, this results does not give a satisfactory solution to the problem of the origin of an expectation ordering — it is like putting the cart in front of the horse. The proposed definition is worthless from a methodological point of view since the nonmonotonic inferences are what is to be explained with the aid of expectations.

A more constructive answer is to view expectations as emerging from *learning processes*. In our roles as cognitive agents, we do not simply observe the world around us, but we also *generalize* in several ways, by discovering patterns and correlations, by forming concepts, etc. The generalizations breed expectations. Expectations are in this way accumulated by *inductive* methods rather than by deductive reasoning. In an evolutionary perspective, expectations can be regarded as a way of summarizing previous experiences in a cognitively economical way.

The analysis presented in Section 3 represents expectations by an ordering of *propositions*. However, if expectations are created by inductive methods, propositional representations of expectations need not be the most appropriate form. In Gärdenfors (to appear a), I argue that there are three levels of inductive reasoning: The symbolic, the conceptual and the subconceptual. On the symbolic level, inductive inferences are represented by propositions, while on the conceptual level observations and inductive processes are represented by *conceptual spaces* consisting of a number of quality dimensions. On the subconceptual level, finally, observations are described in terms of the perceptual receptors of the mechanism (human, animal, or artifical) performing the inductive generalizations. In contrast to traditional philosophy of science and AI approaches, I argue that the most important aspects of inductive processes are to be found on the conceptual and subconceptual levels. Consequently, the origins of expectations should be sought on these levels too.

A currently popular method of modelling processes on the subconceptual level is by using *neural networks*. When a neural network is trained, the weights of the connections between the neurons are changed according to some learning rule. The set of weights of a network obtained after the training period is thus an implicit representation of the "expectations" of the network.

In this context, it can be noted that Balkenius and Gärdenfors (1991) show that by introducing an appropriate schema concept and exploiting the higher-level features of a "resonance function" in a neural network, it is possible to define a form of nonmonotonic inference relation. It is also established that this inference relation satisfies some of the most fundamental postulates for nonmonotonic logics. The upshot is that a large class of neural networks can be seen as performing nonmonotonic inferences based on the expectations of the network. The construction presented in that paper is an example of how symbolic features can emerge from the subsymbolic level of a neural network.[13]

However, neural networks constitute only *one* way of modelling expectations. Apart from their role in logical reasoning, it seems to me that the notion of expectation is central for many cognitive processes. Hence, it is of great interest for cognitive science in general to investigate different models of expectations. With the exception of "expected utility", the concept does not seem to be much studied within cognitive psychology.[14] One further exception is Dubois and Prade's (1991) work on the connections between expectation orderings and *possibility logic*, which points to a different direction. In conclusion, I would like to recommend that the notion of expectation be studied from a variety of approaches within cognitive science. There are numerous potential applications of such studies.

ACKNOWLEDGEMENTS

Research for this article has been supported by the Swedish Council for Research in the Humanities and Social Sciences. Earlier versions of this paper has

[13]For further discussion of of how logic emerges from the dynamics of information, see Gärdenfors (to appear b).

[14]And expected utility has to do with expectations of *values*, not expectations about *knowledge* as is the contents of the expectations studied in this paper.

been presented at the conference on Logic at Work, Amsterdam, December 17–19 1992, and the European Conference on Analytical Philosophy, Aix-en-Provence, April 23–26, 1993. I wish to thank the participants at these meetings, in particular Dov Gabbay and Fiora Pirri, for helpful comments.

REFERENCES

Balkenius, C. and P. Gärdenfors (1991), "Nonmonotonic inferences in neural networks," in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference KR'91*, J.A. Allen, R. Fikes, and E. Sandewall, eds. (San Mateo, CA: Morgan Kaufmann), pp. 32–39.

Dubois, D. and H. Prade (1991), "Epistemic entrenchment and possibility logic," *Artificial Intelligence* 50, pp. 223–239.

Freund, M., D. Lehmann and D. Makinson (1990), "Canonical extensions to the infinite case of finitary nonmonotonic inference relations," in G. Brewka and H. Freitag, eds. *Arbeitspapiere der GMD nº 443: Proceedings of the Workshop on Nonmonotonic Reasoning*, pp. 133–138.

Gärdenfors, P. (1978), "On the logic of relevance," *Synthese* 37, pp. 351–367.

Gärdenfors, P. (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (Cambridge, MA: The MIT Press, Bradford Books).

Gärdenfors, P. (1990), "Belief revision and non-monotonic logic: Two sides of the same coin?," in *ECAI 90: Proceedings of the 9th European Conference on Artificial Intelligence*, L. Carlucci Aiello, ed. (London: Pitman Publishing), pp. 768–773.

Gärdenfors, P. (1991a), "Nonmonotonic inferences based on expectations: A preliminary report," in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, J. A. Allen, R. Fikes, and E. Sandewall, eds. (San Mateo, CA: Morgan Kaufmann), pp. 585–590.

Gärdenfors, P. (1991b), "Belief revision and relevance," *PSA 1990,* Volume 2, pp. 349–365.

Gärdenfors, P. (to appear a), "Three levels of inductive inference," to appear in the *Proceedings of the 9th International Congress of Logic, Methodology, and Philosophy of Science* (Amsterdam: North-Holland).

Gärdenfors, P. (to appear b), "How logic emerges from the dynamics of information," to appear in *Logic and the Flow of Information*, J. van Eijck and A. Visser, eds.

Gärdenfors, P. and D. Makinson (to appear), "Nonmonotonic inference based on expectation," to appear in *Artificial Intelligence.*

Gärdenfors, P. and H. Rott (to appear), "Belief revision," to appear as Chapter 4.2 in *Handbook of Logic in AI and Logic Programming*, *Volume IV: Epistemic and Temporal Reasoning*, D. Gabbay, ed. (Oxford: Oxford University Press).

Ginsberg, M. L. (1986), "Counterfactuals," *Artificial Intelligence* 30, pp. 35–79.

Lakoff, R. (1971), "If's, and's, and but's about conjunction," in *Studies in Linguistic Semantics* Vol. 3, R. Cole and J. L. Morgan, eds. (New York: Academic Press), pp. 114–149.

Lehmann, D. and M. Magidor (1992), "What does a conditional knowledge base entail?" *Artificial Intelligence* 55, pp. 1–60.

Luria, A. (1976), *Cognitive Development: Its Cultural and Social Foundations* (Cambridge, MA: Harvard University Press).

Luria, A. (1979), *The Making of Mind – A Personal Approach to Soviet Psychology* (Cambridge, MA: Harvard University Press).

Makinson, D. and P. Gärdenfors (1990), "Relations between the logic of theory change and nonmonotonic logic," in G. Brewka & H. Freitag, eds. *Arbeitspapiere der GMD nº 443: Proceedings of the Workshop on Nonmonotonic Reasoning*, pp. 7–27. Also in *The Logic of Theory Change*, A. Fuhrmann and M. Morreau, eds. (Berlin: Springer-Verlag, Lecture Notes in Artificial Intelligence nº 465, 1991), pp. 185–205.

Morreau, M. (1992), *Conditionals in Philosophy and Artificial Intelligence*, Ph. D. dissertation, Universiteit Amsterdam.

Rott, H. (1992), "Preferential belief change using generalized epistemic entrenchment," *Journal of Logic, Language and Information* 1, pp. 45–78.