The ambiguous logic of Ambiguity

Kees van Deemter
Institute for Perception Research,
P.O.Box 513, 5600 MB Eindhoven
e-mail: deemter@heiipo5.bitnet
HOLLAND

1 Introduction: The problem of ambiguity in Natural Language Processing

In order for a natural language expression to be interpreted, its intended meaning must be discovered. Sometimes, however, the intended meaning cannot be determined uniquely from the expression alone. If this is so, the expression is called ambiguous.

Ambiguous phenomena are a serious problem for natural language processing. Ambiguity has been identified as the main obstacle for automatic translation of natural language [Bar-Hillel 60], as well as for other natural language processing tasks. Yet, in a way, ambiguity has never been taken seriously, for it is usually assumed that all ambiguities can be resolved. Barwise and Perry, for instance, claim that, although sentences can be ambiguous, *utterances* cannot ([Barwise and Perry 83], pp. 39-41). Each particular utterance of an expression is an utterance of it in a certain "way" that removes all ambiguity. In accordance with this idea, if ambiguity is treated in practical systems, it is usually treated in the manner in which an illness is treated: in order to get rid of it. (See, for instance, the survey [Hirst 87].) The result of ambiguity resolution is a formula from which all ambiguity has been removed and which does its job — database query, or whatever — in standard ways.

However, I think that complete ambiguity resolution is not always possible. It is of little use to ask whether an utterance situation must, *in principle*, always contain enough information to afford complete disambiguation. What counts is that *in practice*, it is often notpossible to extract enough information. There is an abundance of arguments for this position. Most importantly, (1) the use of background knowledge to discard impossible or unlikely readings is notoriously problematical (e.g. [Bar-Hillel 60], or [Carter 87]). Furthermore, (2) there are indications that the use of prosodic cues for the disambiguation of spoken language is not very promising, either [Huber 89]. In addition, (3) a text may be perceived incompletely: if, in a sequence of sentences, one sentence disambiguates the second, then if the first sentence is not perceived, the second one remains ambiguous. Furthermore, (4) ambiguity may sometimes be *exploited*, when something is left in the vague on purpose. And finally, (5) even when purely syntactical disambiguation procedures exist, disambiguation has been proved to be sometimes computationally intractable [Ristad and Berwick 89].

Under the circumstances, it is natural to ask whether disambiguation is always necessary. In this paper, I will inquire how ambiguous expressions may be put to use at a point where they are still ambiguous. Thus, it might be possible to 'treat' ambiguities without getting rid of them. Reasoning with ambiguous expressions could *augment* resolution strategies. Thomason once worded generally held views when he hypothesized that

> "... there is no serious point to constructing an artificial language that is not disambiguated" ([Thomason 73], note 5).

The design and study of such languages is precisely the strategy that I want to advocate.

The scenario for the rest of this paper goes as follows. In section 2, a view of ambiguity will be put forward that describes ambiguity as incomplete information. Section 3 sketches the outlines of a family of "ambiguous logics". Some conclusions are presented in section 4.

2 Ambiguity as incomplete information

An ambiguous expression is, by definition, less informative than any of its disambiguating paraphrases. Therefore, like so many other phenomena in the semantics of natural language, ambiguity has to do with incomplete information. Consequently, the logic of ambiguity is bound to have close analogies with the logic of partial information, a branch of logic that has been studied extensively in recent years [Blamey 86], [Langholm 88], [Fenstad e.a. 87], [Muskens 89].

In principle, the viewpoint of incomplete information can be exploited for the analysis of ambiguity in many different ways. Perhaps the most straightforward account would be skeptical of all the different readings that semanticists have attributed to purportedly ambiguous expressions (cf. e.g. [van der Does and Verkuyl]) and postulate *one* meaning, while the other "readings" are simply specializations of this common meaning. However, such a straighforward account cannot always work. Suppose a grammar generates only one meaning $||X||$ for an expression X, and suppose that this meaning incorporates the commonalities between all the readings of X, in the sense that $||X||$ is the strongest meaning that logically subsumes all possible meanings that X may ever have. Then imagine a pair of sentences A, B, where A is ambiguous between readings A_1, A_2, where A_1 is logically *weaker* than A_2, while B can only mean A_1. Then $||A|| = ||B|| = A_1$, since A_1 is precisely what A_1 and A_2 have in common[1]. But now consider the negations *not A* and *not B*. The sentence *not A* is ambiguous between $\neg A_1$ and $\neg A_2$. The weakest of these two is not $\neg A_1$ but $\neg A_2$, and therefore $\neg A_2$ is what the two have in common, so $||\neg A|| = \neg A_2$. On the other hand, $||\neg B|| = \neg A_1$, so $||\neg A|| \neq ||\neg B||$, even though $||A|| = ||B||$. Consequently, the current account cannot define negation in a compositional fashion, which shows that taking the commonalities is not a viable approach to ambiguity. Ambiguity is not merely a kind of nonspecificity: somehow, the different readings of an ambiguous expression must be taken into account.

Another overly simplistic idea would be to replace the notion of "the meaning" $||\phi||$ of an expression ϕ by the notion of "the set of meanings" of ϕ (in suggestive notation, $||||\phi||||$). This account would fail to take into account that a choice for one meaning as the intended meaning of an expression depends on choices for the meanings of its subexpressions. Even more seriously, there may be dependencies between the interpretations of different occurrences of an ambiguous expression. For instance, if one ambiguous word occurs twice (once in a question and once in the answer to the question, for instance), there seems to

[1]More precisely, A_1 is the logically strongest statement that subsumes both readings (i.e. A_1 and A_2) of A.

be a tendency to interpret both occurrences in the same way. It should be stressed that if dependencies are not taken into account, then the logical consequences are drastic. For instance, in order for $(\phi \vee \neg\phi)$ to be a theorem of an ambiguous logic, the two occurrences of ϕ must be kept constant. In this paper, I will act on the assumption that there is a tendency towards *equal* interpretation of different occurrences, and then the crucial difficult question becomes: how much coherence is obligatory? For sometimes, different interpretations for different occurrences of constants must be allowed. For instance, *green* must be able to have different readings in *I am not so green as to eat green bananas* [Landsbergen Scha 79]. Note, however, that the logic of the situation *enforces* the "unexperienced"-reading upon the first occurrence, whereas the second occurrence can only be used in the sense of a green colour, thus overruling what seems to be a natural tendency. Nevertheless, it might be a rule that *whenever possible*, different occurrences of a constant in a text must be interpreted in the same way. Varieties of this hypothesis will be investigated in later sections.

A nice perspective that combines the viewpoint of incomplete information with the necessity of coherent interpretation can be obtained as follows. First, I adopt Barwise and Perry's idea of a "way to interpret" an expression. (Henceforth: a *mode of interpretation*, or simply a *mode*.) However, we drop the assumption that an utterance situation must always provide complete information. Ambiguity, then, is the case where there is incomplete information about how an expression is interpreted. Consequently, it is modelled by an *incomplete mode* of interpretation. The idea of a mode suggests a certain constancy: it seems plausible that larger stretches of text can be interpreted according to one and the same mode of interpretation. It is this strategy of *incomplete modes* that will be employed in the following sections.

3 The logic of Ambiguity

In order to develop a logic for ambiguous expressions which is based on the notion of a mode of interpretation, I will first give formal content to this notion (section 3.1), after which, in section 3.2, I will set up and compare a number of alternative logics for a language with ambiguous expressions. I will, for now, largely restrict our attention to the case of *lexical* ambiguity, although it will be indicated how the same ideas may also be applicable to, for instance, derivational ambiguities.

3.1 Ways to interpret ambiguous formulas

In this section, the notion of a *mode* of interpretation will be formalized: an analogon of the notion of an interpretation *model* that is suitable to deal with ambiguous constants.

3.1.1 Lexical ambiguity

We will proceed in three stages. First, I will define the notion of an *admissible mode*. Then, a class of "most preferred" admissible modes will be carved out , which are called *coherent* modes. And finally, the notion of a *disambiguation* for ambiguous constants will be defined.

Firstly, an interpretation function must be allowed to have different values for different occurrences of a constant. Therefore, I will assume that any two occurrences of a given constant are different expressions of the language. Different occurrences of a constant α can be distinguished explicitly by the use of superscripts (α^1, α^2), for purposes of reference to them. Furthermore, an interpretation function $\Im$ must be defined on occurrences of constants. Now, if we call a model new-style a *mode*, then a mode must contain a domain and an interpretation new-style. In addition, the set of possible meanings for a given ambiguous constant must be constrained. Perhaps the most natural way to encode this is by Meaning Postulates, which must now hold for all occurrences of constants. For instance, the meaning postulate

$$\forall x(Pitcher^i(x) \leftrightarrow Vase(x)) \vee \forall x(Pitcher^i(x) \leftrightarrow Baseballplayer(x))$$

will now say that each occurrence *pitcher*^{i} of the word *pitcher* must either mean the same as *vase*, or the same as (somewhat inaccurately) *baseballplayer*. In addition, conventional Meaning Postulates have to spell out the relationships between nonambiguous constants. For instance,

$$\forall x(Vase(x) \rightarrow Inanimate(x)),$$

expresses that, in an admissible model with interpretation $\Im$, all occurrences *vase*^{i} and *inanimate*^{j} of the words *vase* and *inanimate* must be interpreted in such a way that $\Im(\text{vase}^i) \subseteq \Im(\text{inanimate}^j)$. Thus, I define

A mode $m = < D, \Im >$ is *admissible* if all meaning postulates are true in m.

As a side effect of the meaning postulates, the semantic *types* of the nonambiguous constants induce a set of possible types for ambiguous constants. For instance, if *baseballplayer* is of type t_h (h for *human*) while *vase* is of type t_f (f for furniture), then the meaning postulate formulated above implies that *pitcher* can be of either type.

Consider a text τ and a classical model $M = < D, I >$. When ambiguous constants enter the picture, I will say that an interpretation new-style $\Im$ *extends* (is an extension of) a classical interpretation I if it treats all occurrences of unambiguous constants in accordance with I:

$\Im$ *extends* I iff, for all occurrences α^i of unambiguous constants α, $\Im(\alpha^i) = I(\alpha)$.

Now, given these notions, how can one define the notion of a *coherent* mode? Note that, in classical logic, an interpretation function provides a *perfectly* coherent "way to interpret" nonlogical constants, just as the assignment to variables provides a way to interpret variables. However, for the modeling of ambiguity, coherence must now be qualified since models must be allowed to make exceptions. As we have seen, an interpretation that interprets the constants in the sentence *I am not so green as to eat green bananas* must be allowed to interpret the two occurrences of the word *green* differently. There are several ways to implement this idea. For instance, one may use a default rule that prefers one meaning, but comes up with a second option whenever the default is overruled by contextual factors. However, I will formulate an even more thoroughly "coherent" proposal in which as many occurrences of constants as possible are interpreted equally.

Coherence can now be enforced by requiring that $\Im$ contains a minimum of exceptions, given a text, the classical interpretation that $\Im$ extends, and the meaning postulates. Let the *agreement number* of an interpretation new-style, with respect to a text, denote the number of occurrences of constants in that text that are interpreted equally:

$\Im$'s *agreement number*, relative to a text τ, equals
$|\{< \alpha^i, \alpha^j >: i \neq j \;\&\; \alpha^i \text{ and } \alpha^j \text{ occur in } \tau \;\&\; \Im(\alpha^i) = \Im(\alpha^j)\}|$.

The notion of a *coherent mode* can now be defined as follows:

Let M be a model $< D, I >$. Then a mode $m =< D, \Im >$, where $\Im$ extends I, is ***coherent*** with respect to a text τ and M if
— (i) m is ***admissible***, and $\Im$ does not lead to type conflict in τ.
— (ii) there is no extension $\Im'$ of I that fulfils (i) and that has a higher agreement number, relative to τ, than $\Im$.

We will also call an interpretation *coherent* with respect to a text if it is contained in a coherent mode for that text. Conversely, I will say that a mode m *extends* a model M if the interpretation function contained in m extends the one contained in M. To illustrate the definition of a coherent mode, suppose there are two occurrences, α^1, α^2 of the constant α in a text τ, while the only coherent interpretations for α^1 that do not lead to a type conflict in τ are $\{a, b, c\}$, and the only coherent type-correct interpretations for α^2 are $\{b, c, d\}$, then a coherent model has two different options: either $\Im(\alpha^1) = \Im(\alpha^2) = b$, or $\Im(\alpha^1) = \Im(\alpha^2) = c$. This illustrates that a text can allow more than one coherent extension. It may even happen that two different constants have occurrences in a text, and that the definition of a *coherent model* allows either one of them (but not both) to be interpreted equally on all its occurrences. To illustrate, consider the text

$$\tau = F(a)\&F(a_2)\&G(a),$$

where a is ambiguous between a_1 and a_2, and F is ambiguous between F_1, which is applicable to a_2 but not to a_1, and F_2, which is applicable to a_1 but not to a_2.[2] Suppose, finally, that G is applicable to a_1, but not to a_2. Then τ contains contradictory information about the types of F and a, as it were. As a result, either F or a must allow exceptions in its interpretation[3]. The earlier natural language example shows what happens when context forces two occurrences of the same constant to have different interpretations: the first occurrence of *green* is forced to mean *unexperienced*, while this meaning is impossible in combination with the CN *bananas*.

The current "dynamic" formulation of coherence pushes coherence as far as possible. For instance, restrictions on the possible values of an occurrence α^i may even influence what values a later *or earlier* occurrence α^j can take. This can be illustrated if, in the earlier example of the text τ, the conjunct $F(a_2)$ is omitted. In the resulting text, G causes its argument, the *second* occurrence of a, to be interpreted as a_1, and this enforces the same reading a_1 upon the *first* occurrence of a. In other respects, however, the present proposal is a relatively conservative choice, since the only kinds of "pressure" on coherent models that it takes into account are coherence and type-conflict. A more adequate account would also take into account whether an interpretation is plausible in other ways. Logically inconsistency is a case in point. For instance, suppose F is an ambiguous predicate. Then one might argue that an interpretation that attributes the same interpretation to both occurrences of the predicate F in the formula $F(a)\&\neg F(a)$ is not coherent, since no model that does this can make the formula true. However, this idea would introduce some obvious technical complications into the definitions that I would rather avoid, except in some occasional remarks.

[2]It may be that predicates F_1 and F_2 exist in the language, but this is not essential. In the latter case, F_1 and F_2 can be seen as shorthands for arbitrary meanings.

[3]To be precise, coherent modes can either have F^1 (i.e. the first occurrence of F) $= F^2$ (i.e. the second occurrence of F) $= F_1$, but then a must allow an exception: $a^1 = a_2$ and $a^2 = a_1$; or $a^1 = a^2 = a_1$, but then F must allow an exception: $F^1 = F_2$ and $F^2 = F_1$. Note that both interpretations have the same *agreement number*.

However, we are still one step removed from a full formalization of a "way to interpret" ambiguous constants. For, a mode of interpretation gives a full interpretation of unambiguous as well as ambiguous constants, whereas we only want to formalize how the meaning of ambiguous constants depends on the meaning of unambiguous ones. Therefore, a "way to interpret" ambiguous constants can be formalized as a disambiguation function μ that has a model as it argument and a *mode* as its value. The set **Dis** of possible disambiguations is defined as follows:

> **Dis** is the set of functions μ that have a model $M =< D, I >$ as argument and a mode $\mu(M)$ as value, and such that, for all M, it holds that $\mu(M) =< D, \Im >$, for a certain $\Im$ that *extends* I.

So, where a mode is the "ambiguous" equivalent of a classical model, a disambiguation function (henceforth, a disambiguation) formalizes the idea of a way to interpret ambiguous constants. Whenever the distinction between ambiguous and unambiguous constants is relevant, the more complex terminology of disambiguations has to replace the notion of a mode.

3.1.2 Combining lexical and other ambiguities

In this section, I will indicate very briefly how the idea of a "way to interpret" an expression might be generalized beyond the case of ambiguous constants.

A "way to interpret" a derivationally ambiguous expression is, of course, basically a derivation. There are several ways to model a derivation. One option is to view a derivation tree as a relation between occurrences of syntax rules and a number of things that are either occurrences of syntax rules or occurrences of basic expressions (i.e. of constants) of the language. This relation of immediate dominance holds between an occurrence of a *basic expression* and an occurrence of a syntax rule if the basic expression is one of the arguments of the syntax rule; and it holds between an occurrence S^2 of a *syntax rule* and an occurrence S^1 of a syntax rule if S^2 is the rule occurrence that formed one of the arguments of S^1. For instance, the derivation of the logical formula $((\phi \vee \psi)\&\chi)$ can be described by the relation $\{<\chi, \&>, <\vee, \&>, <\phi, \vee>, <\psi, \vee>\}$. (Here, & denotes the syntax rule that introduces conjunctions, etc.)

It seems that the notion of coherence applies to derivations as well as to constants and syntax rules. For instance, the argument

```
All old plants and trees were green
-------------------------------------------
All very old plants and trees were green
```

is intuitively valid, since if *old* applies to the entire CN *plants and trees* in the premiss, then so must *very old* in the conclusion, but if it applies only to *plants*, then the same holds for *very old*. As before, coherence must allow exceptions, since one occurrence of a syntax rule may be forced to have one reading, while another occurrence of the same rule may be forced to have another reading. Therefore, let the ***agreement number*** of a derivation relation R denote the number of pairs $(< x^i, y^j >, < x^k, y^n >)$ in R, where x^i and x^k are different occurrences of the same basic expression or syntax rule, while y^j and y^n are different occurrences of the same syntax rule. Then a stipulation of the following type may be postulated:

> R is a **coherent derivation relation** for a text τ iff
> — (i) R is a derivation relation that encodes a possible derivation for τ, and
> — (ii) there may not exist a derivation relation R' for τ that fullfills (i) and that has a higher agreement number than R.

Thus, assume that R is the derivation relation that encodes the derivations of the two sentences that make up the above-mentioned argument, and assume that R is *coherent.* Furthermore, call the rule that combines an adjective and a Common Noun S_{adj}, and call the rule that combines two Common Nouns to form a conjunctive Common Noun S_{and}. Then *coherence* requires that these two syntax rules stand in the same relation of dominance in both of their occurrences. In other words, the two sentences must have nearly isomorphic derivation trees. As a consequence, the above-mentioned inference must hold, as is easy to see.

Complete modes of interpretation. In this way, we have arrived at two components of a "way to interpret" expressions: a coherent *mode* and a coherent *derivation relation.* To catch up with discussions in section 2, I will call a pair that consists of an Interpretation function and a Derivation relation a (complete) *mode* of interpretation. Let the *agreement number* of m, with respect to τ, be the sum of (the agreement number of M w.r.t. τ) + (the agreement number of R). As promised, here is a tentative definition of coherence that corrects the flaws of "apartheid" in earlier stipulations:

> A mode $m = < \Im, R >$ is an **coherent** (complete) **mode** of interpretation, given a text τ, iff
> - (i) $\Im$ fulfils all Meaning Postulates and all components of m together succeed in attributing a meaning to τ, and
> - (ii) m contains a minimum of exceptions in the sense that there is no mode m' that fulfils (i) and that has a higher agreement number than m.

Incomplete modes of interpretation. Now that I have defined the notion of a complete mode, how can *incomplete* modes be defined? I propose to use an elimination approach [Landman 86] for the modeling of incompleteness and to view related concepts accordingly. Thus, a set of complete modes m constitutes an incomplete mode $\mathbf{m}$, and an incomplete mode $\mathbf{m}$ is *at least as strong* as an incomplete mode $\mathbf{m}'$ if $\mathbf{m} \subseteq \mathbf{m}'$.

In the next section, I will use the devices developed here to address questions of truth and logical consequence in an ambiguous language.

3.2 Reasoning under lexical ambiguity

In dealing with inferential properties of ambiguous expressions, attention will be restricted to lexical ambiguity. The inclusion of derivational ambiguities would lead to additional complications, since derivational ambiguity jeopardizes the notion of a *subformula* and therefore also the possibility of direct recursive definitions. Therefore, a complete mode consists of a domain and an interpretation function new-style only and we will see how a logic can be set up.

3.2.1 Varieties of logical consequence

The first milestones on the way to a theory of logical consequence are truth and falsity. Truth with respect to a complete mode is basically a classical notion, except for the fact that the clause for an atomic formula must allow different interpretations for different occurrences of constants:

$R^k(a_1^i, ..., a_n^j)$ is true with respect to a mode $m =< D, \Im > \Leftrightarrow$
$< \Im(a_1^i), ..., \Im(a_n^j) > \epsilon \Im(R^k)$.

There are several plausible options for truth as well as falsity with respect to an *incomplete* mode $\mathbf{m}$. In particular, one may require truth (falsity) in *all* or in *some* complete modes, and one may either or not restrict attention to modes that are coherent with respect to the formula in question:

Strong: ϕ is true with respect to an incomplete mode $\mathbf{m} \Leftrightarrow \phi$ is true in all admissible $m \,\epsilon\, \mathbf{m}$ (or: in all $m \,\epsilon\, \mathbf{m}$ that are coherent w.r.t. ϕ and the model that m extends),

Weak: ϕ is true with respect to an incomplete mode $\mathbf{m} \Leftrightarrow \phi$ is true in at least one admissible $m \,\epsilon\, \mathbf{m}$ (or: in at least one $m \,\epsilon\, \mathbf{m}$ that is coherent w.r.t. ϕ and the model that m extends),

and similarly for falsity. All these options can be motivated convincingly. For instance, the weak versions give a speaker the benefit of the doubt. If one does not know what a sentence is intended to mean, and if it allows a true interpretation, then the polite thing to do is consent. The strong versions, on the other hand, take a careful approach: they reckon with the worst, in order to avoid misunderstandings. This illustrates that, in an ambiguous setting, logical concepts become ambiguous themselves.

Now that versions of *truth* and *falsity* have been defined, the next step is to define logical consequence. A priori, there are many possibilities. The principle that logical consequence should *preserve truth* does not suffice here, since there are so many sorts of truth around. The situation is similar to the one for truth and falsity. Firstly, one may assume complete disambiguation and define logical consequence relative to a certain disambiguation μ, or one may abandon the assumption of complete knowledge and *quantify* over possible disambiguations. Secondly, in the latter case, one may either quantify existentially or universally. Thirdly, in all these cases, one may either or not require that a mode of interpretation is coherent for the text of the inference. Finally, as we have seen in section 3.1, one may take into account whether a certain disambiguation can lead to a consistent interpretation of the premisses of an argument.

Let me illustrate these choices. To save space, I use some abbreviations: I will write $m \models \phi$ for "ϕ is true with respect to m" and $\mathbf{M}$ will stand for the set of admissible modes. Assume that A is a formula and τ is a text, that is, a sequence of formulas. Then a version of ambiguous consequence that assumes complete disambiguation without taking coherence into account can be formalized as follows:

(a) *Ambiguous consequence relative to a disambiguation; noncoherent:*
$\tau \models_\mu A \Leftrightarrow_{def} \forall m \epsilon \mathbf{M}$ ($m = \mu(M)$, for certain M) $\Rightarrow$ ($m \models \tau \Rightarrow m \models A$).

A coherent version may be formulated as follows:

(b) *Ambiguous consequence relative to a disambiguation; coherent:*
$\tau \models_\mu A \Leftrightarrow_{def} \forall m \epsilon \mathbf{M}$ ($m = \mu(M)$, for certain M) $\Rightarrow$
m is coherent w.r.t. $\tau \cup \{A\}$ and $M \Rightarrow$ ($m \models \tau \Rightarrow m \models A$).

A more subtle, partial version of coherence will be illustrated below, where versions of inference are discussed that are not parametrized by a disambiguation. Consistency can be taken into account by requiring the existence of models that make the premisses true with respect to the disambiguation in question. Thus, a "consistent" version of (a) may be formalized as follows:

(a)$'$ = *version (a) + consistency:*
$\tau \models_\mu A \Leftrightarrow_{def} \exists m \epsilon \mathbf{M}(m = \mu(M)$, for certain M & $m \models \tau$ &
$\forall m \epsilon \mathbf{M}$ $(m = \mu(M)$, for certain $M) \Rightarrow (m \models \tau \Rightarrow m \models A)$.

Given my assumption, throughout this article, that very often no disambiguation is available, "supervaluational" approaches, which quantify over possible disambiguations, are of more concern to us than parametrized ones. In supervaluational accounts, one may restrict quantification to a subset $\mathbf{Dis}'$ of $\mathbf{Dis}$, to formalize incomplete disambiguation of constants. Now, ***weak*** versions would require logical consequence with respect to *some* complete modes in $\mathbf{Dis}'$, while strong ones require logical consequence with respect to *all* of them. I will mostly leave the domain of quantification implicit and write, for instance, $\forall\mu$ to mean $\forall\mu\epsilon\mathbf{Dis}'$, for some suitable $\mathbf{Dis}' \subseteq \mathbf{Dis}$.

If coherence is *not* taken into account, the following definitions result:

(1) ***All admissible modes, strong:***
$\tau \models A \Leftrightarrow_{def} \forall\mu, M\ (\mu(M)\epsilon\mathbf{M} \Rightarrow (\mu(M) \models \tau \Rightarrow \mu(M) \models A))$.

(2) ***All admissible modes, weak:***
$\tau \models A \Leftrightarrow_{def} \exists\mu\forall M\ (\mu(M)\epsilon\mathbf{M} \Rightarrow (\mu(M) \models \tau \Rightarrow \mu(M) \models A))$.

The behavior of these notions will be studied in the next section. Note that the requirement of consistency can also be superimposed on unparametrized inference. For instance, quantification in (1) and (2) may be restricted to those μ for which $\exists M(\mu(M) \models \tau)$. Furthermore, as we have seen, quantification can be restricted to coherent modes. The most straightforward option is to give these definitions the following counterparts:

(3) *Total coherence, strong:* $\tau \models A \Leftrightarrow_{def} \forall M, \mu, m(m = \mu(M) \Rightarrow$
$((m$ coherent w.r.t. $\tau \cup \{A\}$ and $M) \Rightarrow (m \models \tau \Rightarrow m \models A)))$.

(4) *Total coherence, weak:* $\tau \models A \Leftrightarrow_{def} \exists\mu\forall M, m(m = \mu(M) \Rightarrow$
$((m$ coherent w.r.t. $\tau \cup \{A\}$ and $M) \Rightarrow (m \models \tau \Rightarrow m \models A)))$.

The strengths of the logics. The logics that I have described are widely different, extensionally. In particular, noncoherent varieties are very weak. It is easy to see that the two noncoherent versions of unparametrized inference are no stronger than their coherent counterparts:

$\tau \models_1 A \Rightarrow \tau \models_3 A$.
$\tau \models_2 A \Rightarrow \tau \models_4 A$.

In fact, $\tau \models_1 A$ "hardly ever" holds. For instance, it does not hold generally that $\phi \models_1 \phi$, since an occurrence of an ambiguous constant in the premiss can be interpreted differently from an occurrence in the conclusion. And if a formula ϕ of predicate logic contains only ambiguous constants, then it cannot be a theorem at all. For instance, $\not\models_1 \forall x(F(x) \rightarrow F(x))$. Only those theorems hold that do not depend on ambiguous material at all. A typical theorem is $G(a) \models_1 G(a) \vee F(a)$, where G and a are unambiguous constants. It does not matter whether F is ambiguous or not, since theoremhood does not depend on this disjunct. Thus, noncoherent ambiguous consequence is not a very prolific notion.

Weak versions of ambiguous consequence ($\models_{2,4}$) have different characteristics. Most notably perhaps, weak notions tend to be at least weakly ***inconsistent***. For instance, the weak version of the totally coherent calculus is inconsistent: one can easily find ϕ for which $\models_2 \phi$ ***and*** $\models_2 \neg\phi$. The "weak and admissible" calculus is even strongly inconsistent, since a formula of the form $\phi \& \neg\phi$ can be validated by chosing different interpretations for constants in the two occurrences of ϕ. Note that especially *weak* inconsistency is not a straightforward argument *against* these logics, since it may be argued that beliefs ϕ and $\neg\phi$ are only ***syntactic*** contradictories, to be resolved by an interpreter by chosing different interpretations for the two occurrences of ϕ.

It may be expected that the totally coherent versions of inference are classically behaved, but this turns out differently. Consider, for instance, the parametrized version of coherent inference. Assume that the predicate constant F is ambiguous between F_1 and F_2, while only F_1 is applicable to the argument a and only F_2 is applicable to the arguments b and c. Now consider

$$F^1(a) \models^?_\mu \exists x(F^2(x)) \vee F^3(b) \vee F^4(c),$$

where F^1, F^2, F^3 and F^4 denote different occurrences of the constant F. Assume, finally, that μ is such that, for any model M, F^2 is interpreted as the reading F_2, while F^1 is interpreted as the reading F_1. Observe that, for each model M, $\mu(M)$ is coherent with respect to the text of the inference, since the context enforces different meanings for the different occurrences of F. To be precise, F^1 is forced by its argument a to mean F_1, but F^3 and F^4 are forced by their arguments to mean F_2. Therefore, F^2 must obey majority rule and mean F_2. Consequently, the argument is not valid in the coherent version of parametrized inference, even though it is classically valid. The same example also shows that the *unparametrized* version of total coherence is weaker than classical logic, since there is at least one coherent disambiguation (namely μ) that invalidates it. As we will see in the next section, the reason for the invalidity of the inference in the two coherent systems is characteristic: context can destroy validity.

Yet, the unparametrized version of total coherence is closely related to classical logic. In fact, classical theorems in the strong-and-coherent versions of ambiguous consequence can only fail to hold as a result of "majority rule". If there is no contradictory information about the meaning of constants, then perfect (i.e. classical) coherence is enforced. One of the consequences of this observation is that if ϕ contains only one constant, say F, then the defined notion behaves classically:

$\models_3 \phi(F)$ if and only if $\phi(F)$ is classically valid.

3.2.2 Structural Rules

In the present section, I will show that the logics described in section 3.2.1 are not just weaker than classical logic but are more radically different from classical paradigms, since they violate one or more of the so-called *structural rules* of logic. These rules do not concern the formal properties of specific operators of a language, they govern the way in which old results can be used and combined to yield new ones. A case in point is ***monotonicity***, which expresses that the addition of new premisses to an argument does not jeopardize the inference.

Violation of structural rules may rightly be called a trend in current logical research. For instance, it is well-known that *nonmonotonic logics* have been proposed as models for diverse varieties of reasoning. In addition, structural rules have been used to characterize the logical properties of *categorial grammars* [van Benthem 90]. In the sequel of this section, the validity of the structural rules for some of the systems of ambiguous consequence will be tested. I will continue to concentrate on unparametrized versions, since these are more interesting from my point of view. Finally, I will simplify and assume that quantification, in all the supervaluational versions, is over the entire set **Dis** of disambiguations. The results for versions that employ real subsets of **Dis** can then easily be derived as well. It is easy to see that the order of presentation of the arguments of an inference is irrelevant even in the coherent versions of ambiguous logic that I have proposed, since I have opted for a "bidirectional" version of coherence. In other words, we have

— **permutation invariance:** $\eta, \chi, \phi, \kappa \models \psi \;\Rightarrow\; \eta, \phi, \chi, \kappa \models \psi$

This allows a simple statement of all the other structural rules:

— **reflexivity:** $\eta \models \eta$
— **monotonicity:** $\eta \models \psi \Rightarrow \eta, \chi \models \psi$
— **contraction:** $\eta, \chi, \chi \models \psi \Rightarrow \eta, \chi \models \psi$
— **expansion:** $\eta, \chi \models \psi \Rightarrow \eta, \chi, \chi \models \psi$
— **cut:** $\eta \models \psi$ and $\psi, \chi \models \phi \Rightarrow \eta, \chi \models \phi$

Structural rules do not concern the specific apparatus of the language (connectives, quantifiers) but the general ways in which old results may be rearranged to produce new ones. Their validity will be tested by means of expressions of a very simple language that has only the connectives of propositional logic; atomic formulas in the style of predicate logic will be used however, in order to illustrate the interplay between nonlogical constants of different semantic types. The results of this enquiry will be presented schematically towards the end of this section.

Firstly, let us consider notions of ambiguous consequence that take all admissible modes into account. The definitions are repeated for convenience:

(1) *All admissible modes, strong:*
$\tau \models A \Leftrightarrow_{def} \forall \mu, M\ (\mu(M) \epsilon \mathbf{M} \Rightarrow (\mu(M) \models \tau \Rightarrow \mu(M) \models A))$.

(2) *All admissible modes, weak:*
$\tau \models A \Leftrightarrow_{def} \exists \mu \forall M\ (\mu(M) \epsilon \mathbf{M} \Rightarrow (\mu(M) \models \tau \Rightarrow \mu(M) \models A))$.

In the case of (1), it is easy to see that *reflexivity* is violated. For instance, the inference $F(a) \models_1 F(a)$ becomes invalid, since a disambiguation μ may attribute different values to the two occurrences of the predicate constant F. All the other structural rules are valid, as one may easily verify. The situation for (2) is different. To begin with, *Reflexivity* holds, since any disambiguation μ that happens to attribute the same values to corresponding occurrences of ambiguous constants must satisfy the clause for $\eta \models_1 \eta$. Similarly, most other structural rules remain valid. Consider, for instance, *monotonicity*. If there is a disambiguation μ that satisfies the clause for $\eta \models_2 \psi$, then it is easy to see that μ must also satisfy the clause for $\eta, \chi \models_2 \psi$. The only rule to fail is the *cut rule*.

Consider the following counterexample: F is ambiguous between F_1 and F_2, G is ambiguous between $G_1 = F_1$ and G_2, while H is ambiguous between $H_1 = G_2$ and H_2. Now

$F(a) \models_2 G(a)$ and $G(a) \models_2 H(a)$, but $F(a) \not\models_2 H(a)$,

since G can be "unified" with F and with H, while F itself cannot be unified with H.

Thus, transitivity, a special case of the cut rule, is invalidated by the fact that ***different*** interpretations may be attached to different occurrences of a constant within a larger argument. On to the more interesting, ***coherent*** versions of ambiguous consequence:

(3) ***Total coherence, strong:*** $\tau \models A \Leftrightarrow_{def} \forall M, \mu, m(m = \mu(M) \Rightarrow ((m \text{ coherent w.r.t. } \tau \cup \{A\} \text{ and } M) \Rightarrow (m \models \tau \Rightarrow m \models A)))$.

(4) ***Total coherence, weak:*** $\tau \models A \Leftrightarrow_{def} \exists \mu \forall M, m(m = \mu(M) \Rightarrow ((m \text{ coherent w.r.t. } \tau \cup \{A\} \text{ and } M) \Rightarrow (m \models \tau \Rightarrow m \models A)))$.

Let us first consider the strong version. One effect of coherence is that the relation is ***reflexive***. For, whenever a premiss contains a constant that also occurs in the conclusion, the agreement number of the argument will be maximized if the two occurrences are interpreted in the same way. But, none of the other structural rules — except for ***permutation invariance***, of course — holds. For instance, the calculus is ***nonmonotonic***.

Suppose, the individual constant a has readings a_1 and a_2, while H is ambiguous between $H_1 = F$ and H_2; assume, in addition, that F is inapplicable to a_2 and that both the constant G and H's reading H_2 are inapplicable to a_1. Then, using H^1 and H^2 for different occurrences of the constant H

$F(a) \models_3 H^1(a)$, but $F(a), G(a) \not\models_3 H^2(a)$.

The reason is that the conjunct $G(a)$ destroys the coherence between the different occurrences of a, thereby making available both a_1 and a_2 as possible interpretations of $H^2(a)$, whereas the occurrence of a in $H^1(a)$ is forced to be interpreted as a_1.

Perhaps unexpectedly, "majority rule" causes even the number of occurrences of a given premiss to be significant. Thus, ***expansion*** and ***contraction*** do not hold.

Suppose, F is ambiguous between F_1 and $F_2 = G$. Assume that the reading F_1 is inapplicable to a, but applicable to b and c, while the reading $F_2 = G$ is inapplicable to c but applicable to a and b. Then

$F(a), F(a), F(b), F(c) \models_3 G(b)$, but $F(a), F(a), F(b), F(c), F(c) \not\models_3 G(b)$.

For, coherence forces the occurrences of F in the first inference to mean $F_2 = G$, but in the second inference, there is as much pressure on F to mean F_1 as there is to mean G. As a result, the occurrence of F in the premiss $F(b)$ becomes free to mean either F_2 (which can be unified with G) or F_2, which cannot be unified with G.

A counterexample against the ***cut rule*** can be construed as follows:

Suppose a is ambiguous between a_1 and a_2, and assume that a_1 is a contant that occurs in the language; furthermore, assume that G is ambiguous between G_1, a reading that is inapplicable to a_1, and $G_2 = F$, a reading that is inapplicable to a_2 but applicable to a_1. Then

$\models_3 F(a_1) \rightarrow F(a)$ and $\models_3 F(a) \rightarrow G(a)$, but $\not\models_3 F(a_1) \rightarrow G(a)$.

This can be seen as follows: the occurrence $F(a)$ in the second premiss forces the reading a_1 upon a. But since only the reading F of G is applicable to a_1, $F(a) \rightarrow G(a)$ really says $F(a) \rightarrow F(a)$, which is, of course, a theorem. But this constraint on the meanings of G and a does not extend to the conclusion $F(a_1) \rightarrow G(a)$. Consequently, the conclusion contains a formula that can be invalidated by chosing an appropriate interpretation.

As it happens, the weak variety of the coherent calculus behaves in the same way, as far as the structural rules are concerned. For instance, the calculus is ***nonmonotonic***.

To see this, suppose one has two ambiguous predicate constants, F and G. Suppose the meaning postulates say that a is unambiguous, while F is ambiguous between F_1 and F_2, and G is ambiguous between G_1 and $G_2 = F_2$. Furthermore, assume that F_1 is applicable to the constant b, while F_2 is not. Then we have

$$F^1(a) \models_4 G^1(a), \text{ but } F^2(a)\ ,\ F^3(b) \not\models_4 G^2(a).$$

The reason is that predicate occurrences F^1 and G^1 can be unified to mean F_2 (choose a disambiguation that extends an arbitrary classical interpretation to an interpretation new-style that interprets both F and G as F_2), while for occurrences F^2 and G^2, this possibility is ruled out by the presence of the premiss $F^3(b)$, since an interpretation can only be coherent for $F(a), F(b)$ if it makes F applicable to b. Therefore, F^2 and F^3 must mean F_1.

Also, *contraction* of premisses can affect a logical argument, since by dropping an occurrence of a premiss, a previously coherent interpretation may become incoherent, as the following example illustrates. As before, different occurrences of the constant a are numbered consecutively for further reference.

$$F(a^1), F(a^2), G(a^3), G(a^4) \models_4 H(a^5), \text{ but}$$
$$F(a^6), F(a^7), G(a^8) \not\models_4 H(a^9).$$

For suppose H is ambiguous between H_1 and $H_2 = G$, while a is ambiguous between a_1 and a_2. Assume, furthermore, that H_1 is applicable to both of these readings of a, while G is only applicable to a_2, and F is only applicable to a_1. In this situation, the occurrence a^5 is still ambiguous between a_1 and a_2, since there are equal amounts of pressure on a^5 to mean a_1 and a_2. Consequently, there is a disambiguation that validates the first argument. (Namely the disambiguation in which a^5 means a_2, and H means the same as G.) However, the occurrence a^9 can only be interpreted as a_1, since that interpretation is enforced by the majority of occurrences of a in the second argument. As a result, the second argument cannot be validated, for H and G cannot be unified.

A counterexample against the *cut rule* can be construed as follows:

$$\models_4 F(a_1) \rightarrow G(a), \text{ and } \models_4 (F(a_1) \rightarrow G(a)) \rightarrow (F(a_1) \rightarrow G(a_2)),$$
$$\text{but } \not\models_4 F(a_1) \rightarrow G(a_2).$$

For, let again a be ambiguous between a_1 and a_2, F be ambiguous between F_1 and F_2, and G be ambiguous between G_1 and $G_2 = F_2$. Then the second premiss of this instance of modus ponens can only be validated by interpreting $G(a)$ as $G(a_2)$, an interpretation that does not validate the first premiss. The conclusion does not follow, as is easy to see.

Here is a schema that summarizes the behavior of six of the proposed systems with respect to structural rules of logic:

	Reflex-ivity	Monoto-nicity	Permu-tation	Contrac-tion	Expan-sion	Cut rule
a	no	no	yes	no	no	no
b	yes	no	yes	no	no	no
1	no	yes	yes	yes	yes	yes
2	yes	yes	yes	yes	yes	no
3	yes	no	yes	no	no	no
4	yes	no	yes	no	no	no

The schema illustrates that, of course, *monotonicity* implies *expansion*. But it also suggests that *expansion* and *contraction* must lead to the same results. More to the point, those structural rules that change the context, so to speak, of the premisses or the conclusion of an argument, are invalidated by all notions of logical consequence that employ our "less than perfect" version of coherence. Given majority rule, context changes meaning, one might say.

In sum, all varieties show nonclassical behavior, both in terms of structural rules and in terms of logical strength, as we have seen. There are two possible reactions to this situation. The first is to admit that reasoning in natural language is a hazardous affair, and that, not unexpectedly, ambiguity is one of the culprits. The other reaction is to look for repair. For instance, one may formulate syntactic constraints on the premisses of an argument. For instance, it may be argued that an argument as a whole must show some form of coherence, over and above the coherence of its individual steps, and that this must be taken into account. Thus, the weak and coherent system (4) does not allow that contradictory statements (cf. section 3.2.1) are true with respect to one and the same disambiguation. Likewise, one may "rescue" the *cut rule* in $\models_4$, by requiring that the same disambiguation is employed for ambiguous constants in all the premisses *and* in the conclusion.

It is hard to say which of the various systems for ambiguous logic is to be preferred. Indeed, the *ambiguity* of the notion of ambiguous consequence may be a central conclusion of this article.

Yet, not all the logical properties that I have used are equally consequential. A choice might be argued along the following lines. It seems reasonable to reject those systems that are not *reflexive*, since reflexivity is, in a sense, the basis of a logic. (If $\phi \not\models \phi$, then also $\phi \not\models \phi \vee \psi$, etc.) But, with some hesitation (see above), the same may be said of consistency: arguably, a logic that allows inconsistency does not deserve the name of a logic. But *if* these two dogmas are entertained, then a choice has already been enforced, since system $\models_3$ is the only system discussed that fulfils both of them. In addition, one might argue that this system, the totally coherent and strong variety of ambiguous logic, is *closest in spirit* to classical logic. For firstly, where classical logic forbids all exceptions in the interpretation of constants, $\models_3$ embodies the most far-going version of *coherence* that is compatible with ambiguity, as we have seen in section 3.1. And secondly, in quantifying universally over possible disambiguations, the system emulates the situation in classical logic, where an inference is valid only if it respects truth for *all possible* interpretations of its constants.

4 Conclusion

In the introduction, I have argued that complete disambiguation is often not *possible*. In the rest of this article I have tried to substantiate the claim that complete disambiguation is not always *necessary*. In particular, I have pointed out that the ambiguity of an expression need not make it unsuitable to figure in a logical argument.

In the course of this paper, I have often made use of the assumption of "coherence", which is a specialization of the idea that different occurrences of an expression must be *somehow* related in meaning. One need not accept this assumption, but then less attractive notions of ambiguous consequence tend to result. (Such "noncoherent" notions were modelled, among others, by the notions $\models_1$ and $\models_2$.) The hypothesis of *coherence* says that different occurrences of a nonlogical constant are interpreted equally, whenever this is possible. However, different hypotheses can easily be substituted for this one. For instance, in the area of structural rules, a preference for *unequal* rather than *equal* interpretation can easily be seen to lead to quite similar results.

One aspect of my account that should perhaps be stressed is its *dynamic* character. For, given the assumption of coherence, a suitable interpretation for an ambiguous constant can only be determined "in context". As a result, the process of disambiguation shares much with the better-known process of anaphora resolution. In particular, the bidirectional notion of coherence — which leads to permutation invariance for the premisses of an argument, cf. section 3.2.2 — implies that even material to the right of an occurrence of an ambiguous constant can influence its interpretation, as long as this material belongs to the same *text*. This parallels the kataphoric phenomena that have been studied in the anaphoric literature. Future research must further constrain the notion of a text as the domain for coherence, in order to provide theories of coherence with more empirical content.

Quite a few logical questions are triggered by my proposals that cannot be resolved in this paper. Firstly, any *direct* algorithmic implementation of the definition of, say $\models_3$, would tend to be quite expensive, computationally. Therefore, some definitions are in need of computationally more efficient equivalents. One unsolved question that arises here is the question of finite axiomatizability. Second, note that a new kind of *expressibility* questions comes up: given any set of properties of models, is there a formula ϕ that is ambiguous, in a given ambiguous logic, between precisely these properties? Likewise, one may ask, for each ambiguous formula in an ambiguous logic, whether all its readings can be expressed unambiguously. Thirdly and finally, nonstandard logics usually induce nonstandard constants. So the question is: are there any attractive new constants that are connected with nonstandard logics for ambiguity? All these questions must be set aside for future research. There is a lot of work ahead in the area of ambiguous reasoning.

Acknowledgement

I thank Jan Landsbergen for useful discussions on the topic of ambiguity. And, as ever, Johan van Benthem, whose interest in the ideas of this paper — a shortened version of chapter 3 of my Ph.D. thesis — has been very inspiring and whose help in giving them form has been unambiguously indispensable.

References

[Bar-Hillel 60] The Present Status of automatic translation of languages, in: Advances in Computers 1, pp.91-163.

[Barwise and Perry 83] Barwise,J. and Perry,J. Situations and Attitudes, The MIT Press/Bradford.

[van Benthem 90] Language in Action. Categories, Lambdas and Dynamic Logic. North Holland, Amsterdam, Studies in Logic.

[Blamey 86] Partial Logic, in: D.Gabbay and F.Guenthner (eds.), Handbook of Philosophical Logic, Vol. III, Alternatives to Classical Logic.

[Carter 87] Interpreting Anaphors in Natural Language Texts, Ellis Horwood, Wiley & Sons, New York.

[van der Does and Verkuyl] The semantics of plural noun phrases, Preliminary draft.

[Fenstad e.a. 87] J.E.Fenstad, P.Halvorsen, T.Langholm, and J.Van Benthem, Situations, Language and Logic, Reidel Publ. Comp., Dordrecht, The Netherlands.

[Hirst 87] Semantic Interpretation and the Resolution of Ambiguity, Cambridge Univ. Press, Cambridge, Mass.

[Huber 89] Prosodic Contributions to the Resolution of Ambiguity, Proceedings of the Conference Nordic Prosody (v), Abo, Finland.

[McDermott and Doyle 80] Non-monotonic logic (i). In: Artificial Intelligence 13, 1980, pp. 41-72.

[Landman 86] Towards a Theory of Information, Phd. Thesis, Universiteit van Amsterdam, The Netherlands.

[Landsbergen Scha 79] Formal Languages for Semantic Representation, in: S.Allen and J.S.Petofi, Aspects of Automatized Text Processing., Helmut Buske Verlag, Hamburg.

[Langholm 88] T.Langholm, Partiality, Truth and Persistence, CSLI Lecture Note No.15.

[McCarthy 80] Circumscription: a form of non-monotonic reasoning. In: Artificial Intelligence 13, 1980, pp.27-39.

[Muskens 89] Meaning and Partiality, Phd. Thesis, Universiteit van Amsterdam, The Netherlands.

[Ristad and Berwick 89] Computational Consequences of Agreement and Ambiguity in Natural Language. In: Journal of Mathematical Psychology, 33, 1989.

[Reiter 80] A Logic for Default Reasoning. In: Artificial Intelligence 13, 1980.

[Thomason 73] Formal Philosophy, Yale University Press, New Haven and London.