

ATTENTION IN THE INFERENCE OF DEFINITE REFERENCE

*James M. Stevens
University of Minneosta*

1. Introduction

Definite reference represents a language task that requires knowledge of a world in addition to knowledge of language itself. A definite description may be unique (e.g. ‘the largest city of South America’), in which case common knowledge of the world should suffice. Perhaps more interestingly, definite expressions are used much more frequently to refer to entities that are not unique per the description (e.g. ‘the city in South America’) but can be isolated as unique through some inference of context.

The application of a semantic theory to real-world mappings of form to meaning is not a straightforward task, as the foundations of different theories may seek different goals. Nonetheless, in this paper I will briefly consider two perspectives for the treatment of language as a valued system: denotation and stochastic, two approaches that are by no means mutually exclusive but are independent historically. A denotation model, as discussed below, will consider that forms are assigned values and complex structures can be evaluated as combinations of the values of their constituent parts (Frege 1892). A more flexible system would consider that forms map onto their values with some probability, and this probability not necessarily being the same for a form across discourse situations or even grammatical structures. The latter, probabilistic approach might be illustrated by example. The expression ‘wooden leg’ might be used to refer to some stub of wood that substitutes for an amputated leg. Although the object is wooden, it actually is not a leg (except on a metaphorical level). If we take the intersection of all entities that are wooden and all that are legs, we would not necessarily find the object we had in mind. For instance, a table leg is in some sense a leg and might be made out of wood, but it is unlikely one would call it a ‘wooden leg’. The complexity of context to derive a meaning might lead one to consider that the meaning of a proposition might not follow directly from the meaning of its constituents. The proposal here claims that by integrating **attention** into a model of interpretation, it is possible to model the inference in human language use necessary to establish the value of forms.

2. The Problem

The problem that this paper addresses is that some uses of language do not fully specify their referent in the world. A prominent example of this includes definite expressions, which may include definite descriptions (e.g. ‘the King of Spain’), definite pronouns (‘she’), demonstrative pronouns (‘that one’), among other forms (Abbot 2004). When we face expressions such as ‘she’ or ‘the man at the beach’, we often do not have enough linguistic information to specify a

referent. If language as code becomes ambiguous, this leaves us with the problem of how humans establish a value for the use of language.

We turn to how the problem of interpreting definite expressions might be addressed within a denotational model and in a statistical model. These two models represent different approaches to the consideration of language as value. The denotational approach, historically the elder as far as questions of language are concerned, is based on the search for logical forms and the truth of propositions that can be derived from the linguistic forms. The stochastic approach, although dating at least to the 18th Century (Bayes 1763), has emerged only recently as a viable framework for the treatment of language, with the advent of statistical computational methods. Whereas the main question of denotation is the logical form, the main question of a stochastic approach will be the inference method. In light of these different goals, the two approaches below are by no means to be considered in competition.

3. Denotation and definiteness

Denotation treats the value of language to be what that language indicates in the world (Frege 1892). The words ‘is the second largest French-speaking city’ would, at present, indicate Montreal. Similarly, ‘car’ would indicate an ambiguous member of the set of cars in the world. A sentence, which does not by itself indicate some thing in the world, is considered to have a truth value, depending on whether the propositions expressed by the sentence match a system of assignment (i.e. a world).

3.1 Russell and definite description

Bertrand Russell (1905) took on the special question of definite descriptions, as these do not seem to fit neatly into a set-theoretic view of value. Russell argued that a definite description (e.g. ‘the Queen of England’) do not carry any value in themselves but obtain value only within the context of a proposition. In this sense, a definite description testifies to the existence of some individual, to the exclusion of others, that meets the criteria of the expressed proposition. If these conditions are not met, then some proposition is false in what is being expressed.

Russell (1919) claims that the only difference between a definite form, ‘the so-and-so’, and an indefinite, ‘a so-and-so’ is that the former implies that it is unique in having this property. His example is that ‘the author of *Waverly*’ can be rephrased logically as ‘the value of x for which “ x wrote *Waverly* is true”’. If this definite description is used within a proposition ‘the author of *Waverly* was Scotch’, then three propositions are implicitly embedded in this phrase (p. 225):

- (1) at least one person wrote *Waverly*
- (2) at most one person wrote *Waverly*
- (3) whoever wrote *Waverly* was Scotch

The above implied propositions include the requirement of existence, the requirement of uniqueness, and the requirement that the proposition expressed

about the entity in question is true. We might consider violation of each of these requirements to see how they do seem to hold up as useful. Firstly, if the sentence was instead ‘the author of the last book ever written was Scotch’, this would be false due to the fact that the last book ever written has not yet been so. Likewise, if we say ‘the inhabitant of London felt out of place in Leeds’, this could not be interpreted as true, because there is no unique individual who inhabits London that comes to mind. Of course, we might imagine a work of fiction beginning in this manner, but the context would restrict the possible individuals to the person we are currently interested in. Finally, if we were to say that ‘the author of *Waverly* was Latvian,’ this simply is not true, as the one individual who fits the criterion of being the author of the work was a native of Scotland.

More formally, Russell (1919) considered that a definite description can be analyzed as the following (p. 226):

“The term satisfying the function ϕx exists”

means:

“There is a term c such that ϕx is always equivalent to ‘ x is c .’”

A definite description ϕx must therefore be satisfied by some individual to be true, and it must be satisfied by that individual in all cases. This still does not give the definite description denotation, because the description can only be considered within a proposition that itself is true or false.

We might consider Russell’s analysis, as consistent with a denotation model, as treating language as a signal, in that the propositional material given in such instances as ‘the King of Spain’ or ‘the author of *Waverly*’ carries enough information for a hearer to establish some individual that satisfies the conveyed information.

There has been criticism of the view that a definite description signals some unique individual.

3.2 Criticism of Russell’s proposal

Peter Strawson disagreed with Russell’s approach to definite description, in that the latter treats the value of a definite description to take value from the proposition in which it occurs, whereas Strawson (1950) argued that a definite description cannot be judged without taking into consideration the context in which it is uttered.

Strawson thus separated how language may form an expression and how it may be used in a context. Three categories of these forms include the following (p. 231):

- (4) a. a sentence
- b. a use of a sentence
- c. an utterance of a sentence

A sentence, in Strawson’s view, cannot be judged to be true or false, unless one considers the use of the sentence and the circumstances of it being uttered. Given a sentence, such as Strawson’s example ‘the King of France is wise,’

there is no means to decide that this sentence itself has a truth value, as we do not know how one might be using it. If one uses the sentence to refer to Louis XIV, there may be reason to agree. If one is using the sentence to refer to a later king, such as Louis XVI, there would likely be less consensus on the truth value. Finally, if the utterance of the sentence occurred in the 21st century in Paris, it would be an inappropriate reference to either king.

Whereas Russell's analysis is said to 'imply' that the referent is unique, Strawson claims that any reference to an individual (e.g. 'the King of France') presupposes the existence of that individual. Strawson gave the following example as a sentence that would not be uttered as either a true or false claim of the world:

(5) The present king of France is bald.

For Strawson, no one would utter the sentence in (5), because there is no such individual, i.e. France, being a republic, does not presently have a king. Thus, one cannot say that there is a true or false proposition about an individual who does not exist. Russell's analysis, on the other hand, would consider the sentence in (5) implies three propositions, one of which is false, namely that there exists some individual with the property of being the present king of France.

In Russell's defense, however, it seems that Russell's analysis was of the propositions contained within natural language and not how language is typically used outside of logic. However, which view do we take if we are to analyze the interpretation of definite expressions as they are used? Apparently, Strawson's interest lies closer to our stated goal, but he offers no formalism to model the interpretation of language.

Another criticism of Russell's analysis was that often definite descriptions do not uniquely map onto a referent (Strawson 1950; Peacocke 1975). Intriguingly, this is exactly what a definite description presumably means, i.e. that some entity uniquely fits the description. When one uses an expression such as 'the man on the train' or 'the girl', there is no expectation that there was only one man on one train or only one girl in the world at the utterance of the expression, and yet Russell's analysis might lead us to believe that. These 'incomplete descriptions' are, according to Russell (1957), interpretable because of egocentricity, in that we narrow our experiences according to context such that only one man on a train or only a single girl possibly comes to mind.

Keith Donnellan (1966) saw two functions to definite descriptions, *attributive* and *referential* uses. Donnellan gives one example of the expression 'the man drinking a martini'. If we are at temperance meeting and a speaker says that 'the man drinking a martini is not following the rules of our association,' this person is making a statement about *any* man who is drinking a martini and thus has no particular individual in mind. Thus, any individual who fits the description of the statement would make the proposition true. This is the *attributive* use of a definite description. On the other hand, one might be at a party and notice an interestingly-looking person in the corner with a martini glass in hand. Making reference to this individual, one might say something about 'the man drinking a martini'. This is the *referential* use of a definite

description, in that the expression used is merely a means to pick out someone, but interestingly the expression need not even be true. One might pick out the individual at the party as ‘the man drinking a martini’ even if his martini glass is filled with water. Russell’s analysis does not capture this distinction and it is not clear how might one do so, at least within the denotation framework, as the referential use of language does not depend on the truth of a proposition. An additional complication is how to model the fact that the attributive use of definite description in the above example does not necessarily imply a unique entity drinking a martini. Thus, Donnellan’s distinction of attributive and referential uses of definite descriptions not only complicates how to analyze something definite, in the sense of deciding which use it belongs to, but also a complication arises from the fact that neither use fits squarely with Russell’s proposal. One might argue that this problem is in fact handled by making a *de re* and *de dicto* distinction between extensional referents and intensional ones, but it is not plainly clear that this distinction truly reflects Donnellan’s separation of uses here.

3.3 Toward a solution of definite interpretation

3.3.1 Attention as predicate

A partial solution to the linguistic material not fully specifying a referent and thus not permitting a truth value is to add information to the logical form. There is a difference between what we say in language and what we interpret it to mean, depending on the context in which it is used. For this reason, we might consider that in addition to linguistic information, we also make use of the speaker’s attention to decide how to interpret a message. If the focus of the speaker’s attention is permitted as itself a predicate a in the logical form, this may allow for the restriction of reference, as desired in the incomplete description problem above.

We might thus consider definite descriptions of the form *the* ϕ in the following manner, which is here a subset of universal quantification.

- (6) $(\forall x)[\text{flower}(x) \wedge a(x) \rightarrow \text{pink}(x)]$
The flower BE pink.

The above form does not specify for number. This would take additional restriction:

- (7) $(\forall x)[\text{flower}(x) \wedge a(x) \rightarrow \text{pink}(x)]$
 $\wedge (\exists y)(\forall z)[\text{flower}(y) \wedge a(y) \wedge \text{flower}(z) \wedge a(z) \rightarrow z = y]$
The flower is pink.

- (8) $(\forall x)[\text{flower}(x) \wedge a(x) \rightarrow \text{pink}(x)] \wedge (\exists y, z)[\text{flower}(y) \wedge a(y) \wedge \text{flower}(z) \wedge a(z) \wedge z \neq y]$
The flowers are pink.

The separation of definiteness from number here does not occur in Russell’s proposal. It is therefore an advantage to have a framework for treating both

singular and plural definites. The proposal here makes definiteness a set of its own, which is linguistically signaled but must be interpreted as those entities within the speakers' attention. Definiteness, in Russell's analysis, was considered to describe a single entity, rather than a single set and for this reason could not handle definite plurals.

The proposed analysis also offers a means to explain the complementary distribution of definite phrasal pronouns (e.g. 'I', 'me', 'mine', but not 'my') and definite articles, as illustrated here in Table 1.

(9) Table 1: Definite form by presence of specified noun

Form of definite marker	
Noun present	Article (e.g. the bird; *it bird)
Noun absent	Pronoun (e.g. it; *the \emptyset)

One might consider that perhaps definite pronouns and definite articles share something of their meaning, but grammatical restrictions force the difference in form. Indeed, some languages (e.g. French) can use the same word (e.g. 'le') as a pronoun and an article. This apparent allomorphy could be explained (at least in part) by considering definiteness itself to be a restriction set. The presence of additional content (i.e. a nominal) would then determine whether the form of the definite expression includes a pronoun or article. 'it' and 'the' can be considered to differ only by the company they keep:

(10) $(\forall x)[\text{parasol}(x) \wedge a(x) \rightarrow \text{striped}(x)]$
The flower BE pink.

(11) $(\forall x)[a(x) \rightarrow \text{pink}(x)]$
It is pink.

Although there certainly is an attraction to collapsing definite articles and definite pronouns into a single 'definite marker' category that is expressed according to grammatical context, the elegance of this model would be tested when considering some languages other than English. For instance, Finnish does not have a definite article, although it does have definite pronouns (e.g. *se* 'it'). If the meaning of definiteness can be expressed pronominally in Finnish and definite pronouns are allomorphs of the 'definite marker' morpheme, then one might expect that language to also have definite articles according to the above analysis. However, the fact that languages differ in how forms express similar semantic content is not generally problematic, and the different availability of definite forms between English and Finnish may in fact be little different than other examples (e.g. that a language might have a word for 'one' but not an indefinite article).

3.3.2. Concerns with attention as a predicate

The advantages to treating attention as itself a logical predicate include the useful combination of semantic and contextual factors to reach a uniquely specified reference set through the use of the speaker's attention. However, this same benefit comes also with a cost. Firstly, the inference procedure involved with resolved the members of a is possibly intractable, if not impossible without additional cues. If the semantic representation is to serve as some form of linguistic value, then the logical forms with attention predicates do not reach any value more than original incomplete descriptions do. In other words, the attention predicate is still incomplete and thus dubiously useful in representing the value that language users take from a sentence.

An additional concern with considering attention to be a predicate is that one's attention is a rather vague criterion for categorizing a referent. We might be able to easily decide that some entity is a 'dog' or even 'fast' or 'happy', but we do not so readily discuss what the focus of our attention is. Thus, it seems unlikely that the focus of our attention is something that can so easily serve as a set as can words. Moreover, what is in our attention does not seem to have a Boolean value, i.e. that something either is or is not in attention. It might be better treated as a gradient value and differently than the values of words (although many words, e.g. 'tall', will likewise be problematic for Boolean valued predicates).

4. A stochastic attempt at definite reference interpretation.

Attention is an important area of study in the brain, and much of what is known about attention comes from visual behavior. Moore and Armstrong (2003) describe attention as an intensified response to a stimulus, when that stimulus itself has not changed in intensity. An analogous understanding of attention might be desirable in the study of language use, whereby a referent may be maintained or magnified in attention, although linguistic material does not correspondingly increase in signaling that referent.

An approach to modeling the view that attention is an increased response to a stimulus source would be to magnify probability values associated with a referent. These values would then correspond to the strength of response. Considering that a full response to a referent as 1, the probability of each referent in an ontology could be assigned a value fitting within the following range: $0 \leq P(X) \leq 1$.

If a referent is signaled, we can consider that the response at 1, but that the maintenance of the referent in attention will depend on the prominence placed on it. Intuitively, the magnification from being a subject will be more than that of an object and this more than an adjunct, all other semantic and contextual information being equal. This prominence ranking would inversely follow the obliqueness hierarchy (Beaver 2002):

(12) Subject < Direct Object < Indirect Object < Adjunct

Thus, the attention parameters α , which we will define as the degree to which something is maintained in attention, will differ in their magnification factor,

relatively to each other, depending on the grammatical role of the referring expression:

$$(13) \quad \alpha_{\text{Subject}} > \alpha_{\text{DirectObject}} > \alpha_{\text{IndirectObject}} > \alpha_{\text{Adjunct}}$$

So that this magnification is not permanent (i.e. a referent will not stay in our attention indefinitely), a decay function A will also be necessary to bring the probability down to an original low level.

For the sake of the argument, we will allow for convenient values for the attention parameters and the decay function, as a function of the number of utterance n since last signaling a referent:

$$(14) \quad \begin{aligned} \alpha_{\text{Subject}} &= 1 \\ \alpha_{\text{DirectObject}} &= .6 \\ \alpha_{\text{IndirectObject}} &= .4 \\ \alpha_{\text{Adjunct}} &= .2 \\ A &= 1/n^3 \end{aligned}$$

The attention parameters and decay function will then give us the means to model how the strength of a response is decreasingly maintained across time. If we consider the strength of response to a referent to be $P(X)$, then for all utterances at time $n > 0$,

$$(15) \quad P(X)_n = A \cdot \alpha \cdot [P(X)_{n-1}]. \quad (\text{when } X \text{ is signaled, } n = 0)$$

The above equation shows that the response at a certain time n (the time of an utterance) will be the product of the response at the previous time $n-1$ (here, we take this as the previous sentence-level utterance), the asymptotic decay function, and the attention function that depends on grammatical role.

An update function to the probability of a referent will allow for the effects of attention to be calculated across utterances. This should make for the difference in inferring a referent when all other information (such as linguistic material) is equal or ambiguous. If, for example, we were to use Bayesian reasoning to infer a referent in a discourse, our update function should disambiguate situations in which the linguistic evidence does not specify a single referent. Here below is Bayes' Law (1763):

$$(16) \quad P(X | Y) = \frac{P(Y | X) \cdot P(X)}{P(Y)}$$

Bayesian reasoning not only has led to advances in robotics and generally in artificial intelligence (Stuart & Russell 2003), but research in neuroscience also suggests that brain depends on probabilistic dependencies, for which Bayesian reasoning offers a model (Knill & Pouget 2004). We take Bayes' Law to mean, for our case, that the probability of a referent given some evidence (here language) is equal to the probability of the evidence, conditioned on the referent, multiplied by the prior probability of the referent and divided by that of the evidence.

The inference procedure notwithstanding, attention factors and a decay function that manipulate the probability of a referent allow for a systematic means of choosing a likely candidate, given an ambiguous signal. Below is a constructed discourse segment to show how an attention parameters and a decay function could help bias an appropriate interpretation:

- (17) a. Mary walked to the store with Susan.
 b. Mary wanted to buy some eggs.
 c. Susan needed pasta but had to borrow some money.
 d. Apparently, Bob had lost his job recently.
 e. She then left, visibly upset.

‘She’ in (17e) is an ambiguous expression, here referring to *Mary*, *Susan*, or perhaps yet another female individual. If we consider the probability of Mary and Susan, we might be able to depend on the attention parameters to disambiguate the referent of ‘she’ above:

- (18) a. Mary walked to the store with Susan.
 $P(\text{Mary})_a = 1$, $P(\text{Susan})_a = 1$
 $\alpha_{\text{Subject}} = .1$, $\alpha_{\text{Adjunct}} = .2$
- b. Mary wanted to buy some eggs.
 $P(\text{Mary})_b = 1$, $\alpha_{\text{Subject}} = 1$
 $P(\text{Susan})_b = A \cdot \alpha_{\text{Adjunct}} \cdot P(\text{Susan})_a = 1/1 \cdot .2 \cdot 1 = .2$
- c. Susan needed pasta but had to borrow some money.
 $P(\text{Mary})_c = A \cdot \alpha_{\text{Subject}} \cdot P(\text{Mary})_b = 1/1 \cdot 1 \cdot 1 = 1$,
 $P(\text{Susan})_c = 1$, $\alpha_{\text{Subject}} = 1$
- d. Apparently, Bob had lost his job recently.
 $P(\text{Mary})_d = A \cdot P(\text{Mary})_c = 1/8 \cdot 1 = .125$,
 $P(\text{Susan})_d = A \cdot \alpha_{\text{Subject}} \cdot P(\text{Susan})_c = 1/1 \cdot 1 \cdot 1 = 1$
- e. She then left, visibly upset.
 $P(\text{Mary})_e = A \cdot P(\text{Mary})_d = 1/27 \cdot 1/8 = .0046$,
 $P(\text{Susan})_e = A \cdot P(\text{Susan})_d = 1/8 \cdot 1 = .125$

Above the application of attention parameters and the decay function led to $P(\text{Susan})$ being higher than $P(\text{Mary})$, respectively .125 and .0046. We would infer, from this information, that *Susan* is the more likely referent of ‘she’, given an ambiguous signal.

The use of a stochastic approach to interpret a definite expression is shown here. The advantages include the ability to address the problem of incomplete description, in a way that is simpler than arbitrarily restricting the domain of reference (i.e. the egocentricity that Russell proposed). Another advantage to the stochastic approach is that it easily can incorporate other information by being amenable to Bayesian reasoning and thus to various forms of evidence. A benefit to a stochastic method of linguistic interpretation is also that there may be neurological support to such a model.

Disadvantages to the above approach include that, despite the hypothesis that cognition may work similarly, it is difficult to support this with solid empirical evidence. Also, there is an arbitrary aspect to deciding the probability values of referents and how these should be manipulated by attention. How one decides even the ranking of attention parameters proves an elusive question. Although one might understand intuitively and follow up with psycholinguistic evidence that the subject is more likely to stay in attention than the objects or grammatical adjuncts, other cues might manipulate this ranking in a subtle way that might not be modeled yet. Finally, attention itself is a somewhat intuitive and ill-defined concept that will be difficult to model in want of a more solid neurological understanding of how it works generally.

References

- Abbott, Barbara. 2004. Definiteness and Indefiniteness. In L. R. Horn and G. Ward (eds.), *The Handbook of Pragmatics*, 122-149.
- Bayes, Thomas. 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* 53:370-418.
- Beaver, Daniel. 2002. The Optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1):3-56.
- Donnellan, Keith. 1966. Reference and definite descriptions. *Philosophical Review* 77:281-304.
- Frege, Gottlob. 1892. On Sense and Nominatum. In *The Philosophy of Language*, ed. A. P. Martinich, *The Philosophy of Language*, 199-211. Oxford: Oxford University Press.
- Knill, David and Alexandre Pouget. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neuroscience* 27:712-19
- Moore, Tirin and Katherine Armstrong. Selective Gating of Visual Signals by Microstimulation of Frontal Cortex. *Nature* 421: 370-373.
- Peacocke, Christopher. 1975. Proper names, reference, and rigid designation. In *Meaning, Reference, and Designation*, ed. Simon Blackburn, 109-32. Cambridge: Cambridge University Press.
- Russell, Bertrand. 1905. On denoting. *Mind* 14: 479-93.
- Russell, Bertrand. 1919. Descriptions. In *The Philosophy of Language*, ed. A. P. Martinich, *The Philosophy of Language*, 221-227. Oxford: Oxford University Press.
- Russell, Bertrand. 1957. Mr. Strawson on Referring. In A. P. Martinich (ed.), *The Philosophy of Language*, 243-246.
- Russell, Stuart and Peter Norvig. 2003. *Artificial Intelligence: A Modern Approach*, Upper Saddle River, NJ: Pearson.
- Strawson, Peter. 1950. On referring. *Mind* 59: 320-44.