

# “the”, “a”, and “”

Jerry R. Hobbs  
Information Sciences Institute  
University of Southern California  
Marina del Rey, California 90292  
USA  
hobbs@isi.edu

September 21, 2010

## Abstract

The definite determiner “the” conveys a relation between the entity referred to by the noun phrase and the description provided by the noun phrase, and says that the entity is uniquely mutually identifiable in context by the speaker and hearer by virtue of that description. This characterization splits into six different cases. The indefinite determiner “a” and the bare plural (the empty string of the title) conveys just the opposite. It says the entity referred to by the noun phrase is not mutually identifiable in context by the description provided by the noun phrase. This could result either from the property described being unknown or from the entity referred to being unknown. Each of these possibilities breaks into several subcases. Several hundred examples of definite and indefinite noun phrases in a diverse corpus are analyzed from this perspective.

## 1 Introduction

The research described here is first effort to develop a neat formal account of the use of the determiners “the” and “a” and the bare plural, in terms of the framework of “Interpretation as Abduction” (IA) (Hobbs et al., 1993). This assumes a style of representation that has been called “ontological promiscuity” (Hobbs, 1985), in which events and properties are reified and sets and type elements of sets are first-class individuals. All morphemes are viewed as conveying a proposition which can be represented in first-order logic, and the logical form of a sentence is a flat conjunction of simple propositions, with roughly one proposition per morpheme.

The first principal claim of the paper is that the word “the” conveys a relation between an entity  $x$  referred to by the noun phrase and a description  $p$  of the entity provided by the noun phrase. The information conveyed by “the” is that  $x$  is mutually identifiable in context by virtue of the property  $p$ . The

second principal claim is that the word “a” and the bare plural do they opposite. They say that the entity  $x$  is not mutually identifiable in context by virtue of the property  $p$ .

In the IA approach the interpretation of a text is the least cost proof of its logical form, allowing assumptions at a cost for propositions that can't be proved. In choosing the least-cost proof we want to favor proofs that use axioms that are currently salient, are shorter, maximize redundancy, minimize assumptions, and use the most recent propositions in the previous text or text structure.

For example, consider the text

John bought a new car. The engine is already broken.

The existence of a car is assumed in interpreting the first sentence— $car(c)$ . A part of the logical form of the second sentence is the proposition  $engine(e, z)$  that there is an engine  $e$  of something  $z$ . Suppose in our knowledge base we have the fact that cars have engines.

$$(\forall x)[car(x) \supset (\exists y)engine(y, x)]$$

One interpretation is obtained by simply assuming  $engine(e, z)$ , that is, there is an engine of something. Another interpretation is obtained by backchaining on the axiom and assuming the engine  $e$  is the engine of some car. This is more expensive because it is longer. A third interpretation is obtained by using the fact  $car(c)$  and the axiom to prove the existence of the engine. That is, the engine mentioned in the second sentence is the engine of the car mentioned in the first sentence. This proof is slightly longer, but it involves no assumptions, so it is the least-cost proof.

Now consider the text

John bought a new car. I saw the red Honda yesterday.

Here the logical form of the second sentence included  $red(h) \wedge Honda(h)$ . Suppose we have an axiom that says that cars manufactured by Honda Corporation are Hondas.

$$\begin{aligned} (\forall x)[car(x) \\ \wedge manufacture(HondaCorp, x) \\ \supset Honda(x)] \end{aligned}$$

Then we can find a partial proof of the existence of a red Honda, from the car mentioned in the first sentence. What is lacking in that proof is that that car is red and was manufactured by the Honda Corporation. But we can assume these two propositions, and still have the least-cost proof of the existence of the red Honda, so we do. These assumptions that the hearer makes in order to see the text as coherent are implicatures. It is new information.

It needs to be emphasized that we are seeking the best interpretation of the *whole* text, not just the definite noun phrases. In the text

Go down Washington Street three blocks.  
Turn left.  
My house is the third one on the right across the street from the drugstore.

“the street” does not refer to Washington Street, but to the street you turned left onto in the second sentence. This is because we have to prove not only the existence of a street, but also the existence of relations between the events and properties described in the successive sentences.

What is missing in the above analyses is the information conveyed by the definite determiner. The word “the” in the first example conveys the information that the engine can be uniquely mutually identified in context by virtue of its description as an engine. In the second example the car can be uniquely mutually identified in context by virtue of its description as a red Honda.

To explicate a notion of “mutual identifiability” we need to spell out a core theory of mutual belief. The key features of such a theory would be the following:

1. If a set of agents mutually believes  $p$  then the individual agents believe  $p$ .
2. If a set of agents mutually believes  $p$  then they mutually believe they mutually believe  $p$ .
3. Agents can do logic inside mutual belief.
4. An agent’s world knowledge is tagged by what groups of agents mutually believe it.
5. Copresence implies mutual belief in what is co-perceived, so previous discourse is mutually believed.

An agent  $a$  identifies an entity  $x$  if  $a$  knows a property  $q$  that is true of  $x$  and of nothing else. Further constraints are generally required on the property  $q$  in various contexts. An entity  $x$  is identifiable by  $a$  by virtue of a property  $p$  if  $a$ ’s knowing  $p$  causes  $a$  to identify  $x$ . The simplest case is where the property  $p$  and the property  $q$  are the same; this is the case of mutually known entities. An entity  $x$  is mutually identifiable by a group  $s$  of agents by virtue of a property  $p$  if it is mutually believed by the agents in  $s$  that if any of the agents in  $s$  know  $p$ , that will cause the agent to identify  $x$ . To repeat what was said above, the word “the” conveys that the entity referred to by the noun phrase is mutually identifiable by virtue of the description provided by the noun phrase.

One way of being identifiable via a description is by being the unique entity of that description. Examples of this include known unique entities (“the world”), entities with a functional relation with another entity, either due to the function (“the top of the table”) or due to the entity (“the engine of the car”), superlatives (“the tallest man in the room”), or sets described by plural noun phrases (“the men in the room”).

More common are cases where the hearer will be able to identify the entity uniquely in the natural course of understanding the discourse. The use of the

definite determiner here is an expression of confidence in the hearer. Speakers are always monitoring the hearer’s understanding via some folk theory of discourse understanding. The word “the” and “a” convey predicates in that theory.

## 2 Categories of Uses for “the”

I have been examining several hundred examples of the use of “the” in a diverse corpus including a novel, business news, biomedical literature, equipment failure reports, poetry, song lyrics, and transcripts of decision-making meetings. These examples can be classified into six categories.

**1. Mutually known entities:** This category is for entities that are not previously mentioned or implied in the discourse but are part of the speaker’s and hearer’s mutual world knowledge. In

a place that is estranged from all other places in *the world*

readers know that there is a uniquely salient world. In

have *the secretaries* bring us some sandwiches

the participants know there is a unique set of secretaries for their group. In

*the start air compressor lube alarm* sounded

the writer and intended reader know about the alarm. In

When I have seen *the hungry ocean* gain  
Advantage on the kingdom of the shore

the writer and intended reader know there is a uniquely salient ocean; “hungry” is new information, determined by implicature.

**2. Directly Anaphoric “the”:** This is the case where the definite noun phrase refers to something mentioned explicitly before and involves the same description of the entity that has been used before. Some examples are as follows:

The town itself is dreary; . . . Otherwise *the town* is lonesome, . . .

A federal agency has proposed a new standard for protecting the privacy of computer data, but the developers of a rival technique say *the standard* would allow . . .

We examine a mathematical model for viral multiplication . . . *The model* shows . . .

It’s not necessary that we do that before I give him the demo. I wouldn’t mind doing *the demo* immediately.

The description in the definite noun phrase is often a truncated version, e.g., “model” for “mathematical model”.

In certain contexts there is a pretense of direct anaphoricity. If there is no previous mention, then the best interpretation often results from assuming the existence of the entity and assuming it is mutually known. This is common at the beginning of novels.

*The town* itself is dreary; not much is there except *the cotton mill*,  
...

Here the existence of the town and the cotton mill are implicatures.

**3. Indirectly Anaphoric “the”:** This is the case where the definite noun phrase refers to something mentioned explicitly before but does so by means of a different description. In

Paul Jain, Media Vision’s founder and CEO, blamed *the company’s*  
problems on ...

the definite noun phrase “the company” refers to Media Vision, but via a different description.

**4. Bridging:** In this case the entity has not been referred to previously, but its existence can be inferred from something that has. This was called “bridging” by Clark (1975). The previous reference to the related entity may be in a noun phrase or in some other word or phrase. In

not much is there except the cotton mill, the two-room houses where  
*the workers* live, ...

the existence of the workers can be inferred from the existence of the cotton mill. In

After an individual is infected with HIV, virus is typically found in  
*the blood*.

an individual is a human and humans have blood. In

When I have seen the hungry ocean gain  
Advantage on the kingdom of *the shore*

the shore is implied by the ocean. In

If we get squeezed, I’ll eat *the time that we lose*

if “squeezed” is interpreted correctly as having a smaller amount of time, then “the time that we lose” is inferable.

In many cases part of a description is implied by a previous mention, and part of the information must be assumed. This is the case of bridging with implicature. In

The town itself is dreary; not much is there except the cotton mill,  
*the two-room houses* where the workers live, ...

the houses can be inferred from the town; towns have houses. But the fact that the house have only two rooms is new information that must be assumed as an implicature. In

After an individual is infected, ... viral antigens are often undetectable during *the long but variable incubation period*.

the incubation period can be inferred from the infection, but the fact that it is long but variable is new information that must be assumed.

**5. Determinative Definite Noun Phrases:** These constitute the very common case where the existence of the entity can be inferred from the complete noun phrase that describes it. In

*the incubation period* of AIDS

we know that AIDS is a disease and hence has an incubation period. The full noun phrase itself contains all the information required to infer the existence of the referent. In

*the process* of acquiring speech

acquiring is a process and the “of” is the “of” of identity. In

*the very center* of town

a town is a region and hence has a center. In

*the developers* of a rival technique

we know about techniques that they have developers.

Determinative definite noun phrases can involve implicature as well. In

*the tenants* from the nearby farms

a tenant is a farmer that rents. Farms have farmers. The fact that those farmers rent is assumed rather than proved. In

explain *the variable likelihood* of transmission

transmission being an event has a likelihood. The fact that it is variable has to be assumed. In

*the kingdom* of the shore

a kingdom is a region ruled by a king, A shore is a region. The (metaphorical) fact that this region is ruled by a king is assumed as an implicature.

**6. Generic Definite Noun Phrases:** In this case the definite noun phrase provides a property and refers to the type element of the set of all entities having that property, and is consequently uniquely mutually identifiable. Examples are as follows:

*The normal human being* is predestined to walk.

*The consumer* is making out like a bandit.

the suppressing capacity of *the immune system*

How many lambs might *the stern wolf* betray

An examination of 238 examples from this corpus showed that the distribution was as follows:

Known:	10%
Directly Anaphoric:	26%
Indirectly Anaphoric:	9%
Bridging:	17%
Determinative:	36%
Generic:	2%

There were slight genre differences. About 12% of the examples involved some sort of implicature.

### 3 Bare Singular NPs and Stuff

We first have to discuss stuff. Often bare NPs are bare precisely because they refer to pieces of stuff. In an abstract theory of stuff, we need to be able to say that an entity  $x$  is a blob of stuff of type  $p$  if  $x$  is made up of a set of granules of type  $q$  that bear some contiguity relation  $c$  with each other. The basic property of stuff is that a part of a blob of stuff of type  $p$  is also a blob of stuff of type  $p$ , down to the level of granules.

For example, traffic is stuff. The traffic on the 405 at rush hour is a blob of stuff, and so is any part of that traffic, down to the level of individual cars. Water and sand are two other examples of stuff.

Stuff can be physical, like traffic, water, sand, energy, corn, inventory, virus and dust. But activities are also viewed as stuff, such as speech, litigation, trading, therapy, music, and lunch. The granules of speech, for example, may be the individual articulations. Conditions can also be viewed as stuff, such as heredity, ruin, privacy, disease, fortune, and jeopardy. There is cognitive stuff like grief, information and interest, and there is economic stuff like money and stock.

It is well-known that leaving the determiner off a count noun can convert the referent from an individual to a blob of stuff (Talmy, 19??), as in

There was *spider* all over the windshield.

Not all bare singular NPs are stuff. In the data examined, there were three kinds of non-stuff bare singular NPs. The first was the category of bare count NPs. These occur in predicate complement position,

He is *president*.

or with certain prepositions,

on *foot*,

or in certain other idiomatic constructions,

*pistol* in hand

*hand* in hand

Some bare singular NPs should probably be viewed as proper names, such as “midnight” and “heaven”. Finally, bare singular NPs occur in telegraphic contexts such as headlines and informal messages.

DOES *AGENCY CHOICE* LEAVE *HOLE* IN PRIVACY OF DATA

*Power pack* failed.

The 240 examples of bare singular NPs examined in this study broke down into these categories as shown in the following table:

Stuff:		73.6%
Physical:	19.0%	
Activity:	24.4%	
Condition:	16.9%	
Cognitive:	11.2%	
Economic:	2.1%	
Bare Count NPs:	12.0%	
Proper Names:	2.1%	
Telegraphic:	12.4%	

Thus, in three fourths of the cases, the absence of a determiner conveys the information that the entity referred to by the NP is a blob of stuff. The other cases are either idiomatic or in informal contexts.

## 4 Categories of Uses for “a” and Bare Plural NPs

We have said that the indefinite determiner “a” and the bare plural NP convey a relation between the entity referred to by the noun phrase and the description provided by the rest of the noun phrase, and says that the entity  $x$  is *not* uniquely mutually identifiable in context by the speaker and hearer by virtue of the property  $p$  conveyed by that description. There are two ways this can happen. The first is if the entity  $x$  is not previously known;  $x$  is a new entity.

*A lawyer* walked into the room.

The second is that the property  $p$  is not previously known;  $p$  is a new property.

Pat is *a lawyer*.

We can divide the instances of the indefinites examined into these two categories. The class of indefinites due to a new property can be further classified into those in which the new property is in a predicate complement; those in which there

is a new property of an inferable entity; and measure phrases. The class of indefinites due to a new entity can be further classified into specific and new entities; entities that exist in a modal context; universal or generic entities; and what I will call Skolemized generics. Each of these categories is discussed in turn.

## 4.1 Indefinites Due to New Property

**1. New Property in Predicate Complement:** In the sentence

Walking is *a general human activity* . . .

the predicate complement “a general human activity” introduces what is purported to be a new or at least newly relevant property of walking. In

a feeling which was *a mixture of exasperation and sadness*

a new property, that of being a particular kind of mixture, is attributed to the feeling. In

The company reported a sharp decline in sales, *a jolt that cut its stock price in half*.

the phrase “a jolt . . .” is an appositive on the whole clause, and appositives are essentially reduced copular relative clauses, so that the NP is essentially a predicate complement. This phrase could be expanded to “which was a jolt . . .”.

These are all examples of the indefinite article applied to nouns functioning as count nouns. Bare singular NPs occur in the same environments conveying a new property, as in

different treatments (*chemotherapy* or *immunotherapy*)

Some say thy fault is *youth* . . .

show him why this is *important research*

We also find the same category in bare plural NPs.

Viral isolates must be thought of as *populations of closely related genomes*.

But thou, to whom my jewels *trifles* are

**2. New Property of an Inferred Entity:** This category covers the case where there is an entity mentioned elsewhere in the text which implies the existence of an  $x$  such that  $p(x)$  holds, and the indefinite NP conveys  $p(x) \wedge q(x)$ , where  $q(x)$  is new information about  $x$ . For example, in

Pat has *a sore throat*.

we know that Pat is human and that humans have throats. That is not new information. But it is new information that the throat we know Pat has is sore.

Three examples of indefinite determiners in this category in the data examined are the following:

The moon shone with *a soft, clear light*.  
Thursday's decline continued *a precipitous two-month slide*.  
Investigation revealed *a broken tooth* on the hub ring gear..

We know the moon has light; the new information is that it is soft and clear. From the continuation of a decline, we can infer a slide; the new information is that it is precipitous and has lasted two months. Gears have teeth; the new information is that one is broken.

The following are examples of bare singular NPs in this category:

the discovery of *great genetic diversity* in viral isolates  
Media Vision shares plummeted in *frantic NASDAQ trading*.  
Investigation revealed *adequate lube oil*.

Viruses have genetic diversity; that it is great is new. shares are traded; that they are traded frantically and on NASDAQ is new. An air compressors, known from a previous sentence or the context, have lube oil; that it is adequate is new.

Finally, some examples of bare plural NPs in this category:

a timid person with *gentle manners and nervous ways*  
Believed due to *worn bushings*.  
Now stand you on the top of *happy hours*

A person has manners and ways; it's new that they are gentle and nervous. An air compressor has bushings; it's new that they are worn. A person experiences hours of time; it's new that they are happy.

**2. Measure Phrases as New Properties:** States and events have durations. Entities have quantities. Sets have cardinality. A measure phrase conveys a new property, viz., the measure, of a duration, quantity, or cardinality. Consider

It needs but *a moment's* reflection to convince us . . .

Reflection is an event, and events have duration. The possessive specializes to the relation  $duration(m, e)$ , or saying that the moment  $m$  is the duration of the reflection event  $e$ . This duration is new information.

In

Not much is there but *a few peach trees*.

“few” tells us something about the cardinality of the set of trees. In

Despite *a year* of filing complaints with regulators . . .

the year tells us the duration of the sequence of filing events. In

*A number* of slow gas turbine engine starts has been noted . . .

“a number” says something about the cardinality of the set of starts. In

Betwixt mine eye and heart *a league* is took,

the word “betwixt” calls to mind a path from the eye to the heart, and the NP “a league” tells us its length. In

The we can bump [the discussion] up *a level* for general issues.

the verb and particle “bump up” indicate a motion upwards, and the indefinite NP “a level” tells us how much.

An example of a bare singular NP in this category is

a jolt that cut its stock price in *half*

Here “half” provides the new information about the quantity of the cut. An example of a bare plural NP in this category is

characterized by *constant or slowly decreasing numbers* of CD4<sup>+</sup> cells.

The bare plural NP tells about the cardinality of the set of cells.

The following table shows how often in the examined data each of the three constructions is used to convey a new property.

	“a”	Bare Sing	Bare Plur
Total Examples:	252	240	174
Percent in This Category:	37.3%	10.4%	13.8%
Predicate Complement:	14.9%	6.2%	2.9%
New Prop of Inferable:	10.2%	3.8%	10.3%
Measure Phrase:	12.2%	.4%	.6%

Thus, over a third of indefinite NPs in the data were indefinite because they conveyed a new property. Fewer bare NPs were in this category.

## 4.2 Indefinites Due to a New Entity

The second major category of indefinites is those in which a new and previously unidentifiable entity is introduced. There are four subcategories in this class, corresponding to ways in which an entity can be unidentifiable.

1. **Specific:** A new specific entity, not previously mentioned nor inferable from a previously mentioned entity is introduced.

Pat bought *a car*.

2. **Modal:** An entity is introduced in a modal context, in which many specific entities may satisfy the relevant properties.

Pat wants to buy *a car*.

3. **Universal or Generic:** A universally quantified variable is introduced, with properties shared by all members of a set.

*A car* is expensive.

4. **Skolemized:** An entity is introduced that is functionally dependent on a universally quantified variable.

Everyone bought *a* car.

In the notation of Hobbs (1985, 19??), in every case the proposition

$car'(e, x)$

is part of the logical form, where this says that  $e$  is the property of  $x$  being a car. In the Specific case, we need to infer in addition the real existence of  $e$  and  $x$ .

$Rexist(e), Rexist(x)$

In the modal case, we need to infer that  $e$  and  $x$  are possible, or occur in the appropriate modal context.

$possible(e), possible(x)$

In the Universal case, we need to infer that  $x$  is the type element of a set whose defining property is  $e$ .

$dset(s, x, e)$

In the Skolemized case, we need to infer that there is a set  $s$  on whose type element  $y$  the entity  $x$  is functionally dependent.

$dset(s, y, e_1), person'(e_1, y), buy(y, x), FD(x, y)$

Here  $s$  is the set denoted by “everyone” and  $x$  is the car that each person bought.

**1. Specific Indefinites:** Some examples of specific indefinites with the indefinite determiner from the data examined are as follows:

She argued for an hour with *a farmer* over *a plow shaft*.  
The company will report *a sharp decline* in sales.  
We examine *a mathematical model* for viral multiplication.  
We have *a demo* that takes half an hour.

Some examples of bare singular NPs in this category are

He returned to town, caused *ruin*, and went on his way again.  
Unit shows high usage of *oil*.

Some examples of bare plural NPs are

There were *tables* with *cloths* and *paper napkins*.  
*Analysts* are saying the company's growth . . .  
*Longitudinal studies* of *patients* with AIDS reveal . . .

**2. Modal Indefinites:** There are several types of modal contexts, including the following:

- Logical: *not, if, never, lacks, ...*
- Possibility: *can, may, able to, ...*
- Goal: *need, desire, require, push for, ...*
- Cognition: *assume, imagine, consider, ...*
- Communication: *rumor, convince, ...*

Some examples of modal indefinite NPs with “a” are the following. The indefinite is in italics; the word inducing the modal context is in bold.

**If** we have *a chance* to present it, ...  
 You **may** see *a price increase* of 30 percent.  
*A new lube oil pump* **is required**.  
**Compose** from such flashes *an image* of these years.  
 We can **convince** him to give us *a similar level of funding*.

Some examples of bare singular NPs in the modal category are

It **avoided** *inventory buildup*.  
 Where *peace* and *love* **can** still be found.  
**Let’s say** we use that hour for *discussion* and *lunch*.  
 They **seem** to be exchanging one long gaze of *grief*.

Some examples of bare plural NPs in the modal category are

**If** these ideas are correct, they have *implications* ...  
**allow** intelligence agencies to spy on *private companies*  
 fueled the **need** for *better encryption methods* .

An interesting example that is ultimately in the modal category is the following:

You need *a minimum* of 40 minutes.

The relevant part of the logical form is

$$need(u, x) \wedge minimum(x, s) \wedge of(x, 40mins)$$

The word “need” provides the modal context. It is important that we do not interpret “of” as signalling the predicate-argument relation, i.e., the minimum minute of the 40 minutes. Rather it is the “of” of identity. The minimum and 40 minutes are the same. To get this interpretation, we need to coerce the logical object  $x$  of *need* into an interval of time; what is needed (to give a presentation) is not a minimum but an interval. This interval is in some set of intervals, viz., the intervals that could be used for the presentation. Each of these intervals has a duration. The set  $s$  is the set of these durations. Thus, we need to expand the logical form into the following, where the coercion is in bold.

$$\text{need}(u, \mathbf{i}) \wedge \text{duration}(\mathbf{d}, \mathbf{i}) \wedge \text{onScale}(\mathbf{d}, \mathbf{s}) \wedge \text{minimum}(x, s) \wedge \text{eq}(x, 40\text{mins})$$

In a sense, we have to expand the original sentence into the sentence “You need an interval from a set of intervals whose durations like on a subscale  $s$  of the scale of durations where the minimum point on the subscale is  $x$ , which equals 40 minutes.”

Then the minimum  $x$  is new because the scale  $s$  is new.

The percentages of modal indefinite NPs found in the data examined in each type of modal context is given in the following table.

	“a”	Bare Sing	Bare Plur
Total Examples:	252	240	174
Percent in This Category:	15.7%	17.5%	20.7%
Logical:	3.1%	3.8%	2.9%
Possibility:	2.0%	8.3%	9.8%
Goal:	7.1%	3.3%	7.5%
cognition:	2.0%	2.1%	.6%
Communication:	1.6%	0%	0%

**3. Universal Indefinites:** Some NPs with indefinite determiners seem to predicate properties of all the members of some set, usually the set defined by the property given by the rest of the NP. Some examples from our data are as follows:

If you walk along a street on *an August afternoon* ...  
 ... it will be difficult for *a vaccine* to ...  
 Shall I compare thee to *a summer’s day*

Similarly bare singular NPs can refer to all portions of the stuff they describe.

*Creative stock* dropped to 5 3/8 a share.  
 In walking, *culture* is not brought into play.  
 The model explains many aspects of *infection*

Bare plural NPs can refer to a set of all entities in the set defined by the property given by the NP.

On *Saturdays* the tenants come in for a day  
*Media Vision shares* plummeted to 11.  
 a discussion about *current interests* and *available funds*

But there are some subtle issues that are raised by universal or generic indefinites.

In the approach to sets taken in Hobbs (1985, 1996, 2009), each set has a “type element” whose principal feature is that all of its properties are inherited by the real elements of the set. (The type element *is* a member of the set, but in an uninteresting way that is equivalent to its being the type element and prevents its entering into cardinality calculations.) Then in the sentence

The numerous tall men assembled.

if  $s$  is the set of men and  $x$  is the type element of that set, then “numerous” and “assembled” are properties of  $s$  and “tall” and “man” are properties of  $x$ .

Another approach would be to say that there is no such thing as a type element, but that  $x$  is a “witness element” which is a real element of the set but one about which we know nothing except all those properties that are shared by all the elements of the set. For example,  $tall(x)$  would predicate tallness to only one element of the set, but since we know about  $x$  only properties true of all members of the set, we can conclude  $tall(y)$  for any other member  $y$ .

It may seem that there is little to choose between these two approaches. In both cases we can draw the essential inference, that if a property holds for  $x$  then it holds for any other member. But in fact by admitting both approaches, we can explain some curious linguistic facts.

First note that generics can be indicated both with definite and indefinite determiners.

*The lion* has a tail. *A lion* has a tail.

There seems to be no difference between these two in meaning. From both we can infer that all lions (defeasibly) have tails. But I would argue that the definite NP refers to the type element of the set of lions, whereas the indefinite NP refers to a “witness element”, a random element about which we know no more than is true of all lions. Then the definite NP would refer to something mutually identifiable – the unique type element of the set. The indefinite NP would refer to something that is not mutually identifiable – a random witness element about which we do not know enough to distinguish it from any other element of the set. This hypothesis preserves our account of the meanings of the definite and indefinite determiners, and explains some curiosities.

One curiosity involves what seem to be anaphoric indefinites. An example from our data with the indefinite determiner is

Once down *a man* [the whiskey] glows inside him . . .  
Such things happen when *a man* has drunk . . .

The second occurrence of “a man” could have been replaced by “the man”; it seems that we’re referring to the same man. But “a man” works at least as well here. An example of an anaphoric universal bare singular NP is

[Speech] seems as natural to man as *walking* . . .  
. . . *walking* is an inherent biological function of man.

The author (Sapir) could have used a definite NP for the second reference to walking – “that walking” – but he didn’t, and he didn’t need to. An example of an anaphoric universal bare plural NP is

. . . increased susceptibility of *AIDS patients* to infections.  
. . . In *AIDS patients*, viremia can be reduced . . .

In the first example, “a man” can be replaced in both instances by “a random man”. There is no necessity for those two individuals to be the same. When

the two mentions must refer to the same individual for coherence reasons, the anaphoric indefinite is not possible. consider the four sentence pairs

A person can try hard to climb that mountain.  
But *the person* can never reach the top.

A person can try hard to climb that mountain.  
But *a person* can never reach the top.

A person can try hard to climb that mountain.  
Then *the person* just may succeed.

A person can try hard to climb that mountain.  
Then *a person* just may succeed.

In the first pair the two mentions of a person are coreferential; that's why the second is definite. In the second sentence of the second pair, the NP "a person" refers to any random person, maybe the same as the one mentioned in the first sentence and maybe not. Coherence does not force coreference. In the third and fourth pairs however, coherence does force coreference. The two persons must be the same. In this case, the anaphoric indefinite is unacceptable. It's not true that any random person can succeed, but only the person who tries hard. (However, see the discussion of lambda-abstracted generics below for a way of constructing exceptions.)

A second curiosity that distinguishes between definite and indefinite generics involves interpretations as sets or kinds. Sentences can predicate properties of objects, as in

*The lion* has a tail.  
*Lions* have a tail.  
*A lion* has a tail.

In this case, all three versions are acceptable. In the first two the property is predicated of the type element, and in the third to a witness element. But sentences can also predicate properties of sets, as in

*The lion* is common.  
*Lions* are common.  
\* *A lion* is common.

The property "common" applies to the set of lions, not individual lions. In the first the type element is coerced into the set whose type element it is. In the third sentence, there is no simple relation between the set and any random lion that can be used as a coercion function. Kinds can be viewed as mapping from times and/or possible worlds into sets. Thus, "becoming extinct" is a property of the kind "lion", not of the set of lions or of an individual lion.

*The lion* is becoming extinct.  
*Lions* are becoming extinct.  
*A lion* is becoming extinct.

A similar account works here. “Lions” refers to the set of lions, and this is coerced into the kind. “The lion” refers to the type element of the set of lions, and this is coerced into the kind. “A lion” refers to any random real lion, and there is no coercion function available for mapping this into the kind.

Another kind of universal or generic indefinite occurs in measure phrases. We have seen indefinites in measure phrases above, but they have two possible uses. In the phrase “a penny a pound”, the indefinite “a penny” is in the category of a new property of an inferable entity. The indefinite “a pound” is in the category of a non-mutally identifiable entity, specifically, a universal or generic entity. In the phrase

5 3/8 *a share*

the NP “a share” ranges over all shares, or applies to any random share. Other example in our data of this usage are

make *a gallon* of oil go further than it used to it’s funded at 150K *a year*

Another kind of universal or generic indefinite might be called lambda-abstracted indefinites. In this case the indefinite NP functions as a property that is applicable not only to the entity focused on, but to a set of similar entities as well. Often it is used to describe something that is true of a specific entity but by virtue of a more general implicational relation. In

Pat had been in a plane for 14 hours. Such *a trip* would make anyone exhausted.

the phrase “a trip” might seem to be an anaphoric reference to Pat’s being in a plane for 14 hours, and thus definite. But the function of “a” here is to remove the reference from Pat’s trip alone and extend it to any trip that involved 14 hours on a plane. Thus, “a” explicitly conveys the property that its referent is not the mutually identifiable trip Pat took, but something less specific. Moreover, it expresses the general implicational relation from being on a plane 14 hours to being exhausted.

Examples of lambda-abstracted indefinites found in the data examined include the following.

Different mutants within such *a quasispecies* . . . The genetic variability is not so much *a complication* as . . . Speech is so familiar *a feature* of daily life that . . . And summer’s lease hath all too short *a date*

An interesting example of the use of lambda-abstracted indefinites is the following sentence from a business news article:

Weiss added that he was confounded by *a situation* in which *a secret owner* could tie the hands of *a board* that did not acknowledge his authority.

The situation, the owner, and the board are all identified in the previous sentences. But by using the indefinites (presumably that part is a direct quote), Weiss indicates that anybody would be confounded by these circumstances.

**4. Skolemized Generics:** The final category of indefinites due to new entity is what I call “Skolemized generics”, after the logical device of “Skolem functions” to represent existentially quantified variables inside the scope of universally quantified variables. In these cases the indefinite entity is functionally dependent on a universally quantified variable, and thus is not a mutually identifiable member of a set.

The universally quantified variable could be introduced by a plural NP or an NP beginning with “each” or “every”, as in

Several women were there and they had *a Nehi* or *a swallow* of whiskey.

It could be introduced by a generic, as in

... the individual ... he is born in the lap of *a society* that ...

It could be introduced by a clause describing habitual action, as in

So on two days he is dull and cross, but on the third day he livens up and sometimes has *an idea* or two.

Bare NPs also occur as Skolemized indefinites. Some examples from the data of bare singular NPs are as follows:

On fine autumn days, she ground *sorghum* ...  
... after an individual is infected with HIV-1, *virus* is typically found in the blood ...  
When I have seen ... *brass eternal* slave to mortal rage

The bunch of sorghum depends on the days, the virus depends on the individual, and the brass depends on the seeing events. Some examples of bare plural NPs as Skolemized indefinites are as follows:

There were ... *great gatherings* on Saturday nights.  
This incubation period is characterized by ... *minor and short-lived upsurges* of viremia ...  
She sold *chitterlins* and sausage in the town near-by.

The set of gatherings depend on the Saturday nights, the set of upsurges depend on the incubation period, and the set of chitterlins depend on the habitual selling events.

The following table presents statistics on the uses of indefinites in the data examined in the subcategories of entities not mutually identifiable because a new entity is introduced.

	“a”	Bare Sing	Bare Plur
Total Examples:	252	240	174
Total in Some Class:	100%	85.8%	96.0%
New Property:	37.3%	10.4%	13.8%
New Entity:	62.7%	75.4%	82.2%
Specific:	22.5%	14.6%	22.4%
Modal:	15.7%	17.5%	20.7%
New Universal:	12.5%	15.8%	20.7%
Anaphoric Universal:	.8%	5.4%	12.1%
Skolemized:	12.2%	22.1%	6.3%

## 5 Conclusion

The English determiner ‘the’ can be characterized as conveying a relation between an entity and a description, where the relation is that the entity is uniquely mutually identifiable in context by the speaker and hearer. This allows room for some of the information conveyed by the noun phrase to be new and resulting from implicature, or assumption, in the interpretation process. The various ways an entity can be mutually identifiable were catalogued. This should provide a good starting point for the commonsense theory of discourse processing that an account of mutual identifiability must tap into.

The counterpart hypothesis is that the indefinite determiners, including “a” for singular count nouns and the empty string for plurals and singular mass nouns, also conveys a relation between the entity the noun phrase refers to and the property it expresses. This relation is that the entity is *not* mutually identifiable by the speaker and hearer in context by virtue of the property. This breaks into two broad categories—NPs that are indefinite because the property is previously unknown and NPs that are indefinite because the entity is previously unknown. Various subcategories of each of these two have been described.

An important lesson from this investigation is that relatively simple characterizations of the major determiners is possible. But this requires a great deal of infrastructure to have been built up, including accounts of mutual belief and identifiability, interpretation by abduction, stuff, and sets and generics.

## Acknowledgments

This research was funded in part by the Advanced Research and Development Activity.

## References

- [1] Clark, Herbert, 1975. “Bridging”. In R. Schank and B. Nash-Webber (Eds.), *Theoretical Issues in Natural Language Processing*, pp. 169-174. Cambridge,

Massachusetts.

- [2] Hobbs, Jerry R. 1985. "Ontological Promiscuity." *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics*, pp. 61-69. Chicago, Illinois, July 1985.
- [3] Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. "Interpretation as Abduction", *Artificial Intelligence*, Vol. 63, Nos. 1-2, pp. 69-142.