

# In Defense of Ambiguity

Pat Hayes  
Florida IHMC

# Theses

- There are two distinct relationships between names and things. *Reference* is different from *access*. The architecture of the Web determines access, but has no direct influence on reference
- Reference can be established by ostention or by description. Description is inherently ambiguous; ostention can be done only to accessible entities.
- Therefore, references to non-accessible entities - the vast majority of references - must be by description, and hence must be ambiguous.
- Reference to accessible entities still differs from access. Establishing reference by ostention requires naming conventions. Access is one form of ostention.

# Two relations between names and things

- It is important to distinguish **two different relationships** between a name and a thing.
- ‘**accesses**’, meaning that the name provides a causal pathway to the thing, mediated by the Web.
- ‘**denotes**’ or ‘**refers to**’, meaning that the name is being used to mention the thing.
- WebArch uses ‘identifies’ to mean both or either of these, apparently in the belief that they are synonyms. But they are not, and to think of them as being the same is to be **profoundly confused**. A great deal of the muddle we now find ourselves in can be directly traced to this confusion.

**‘accesses’**, meaning that the name provides a causal pathway to the thing, mediated by the Internet.

- Understand ‘access’ as inclusively as possible, so this includes uses of an IRI to access a website, an http endpoint, a ‘representation’ (in the REST sense) of a resource, a server, a file, an html page, a mailbox, a webcam, ... *anything at all* that can receive, send or be directly influenced by, or indeed itself *be*, any piece of information that can be transferred by a transfer protocol, either now or in the foreseeable future.
- Cast this net as broadly as you like, the **accessible** things will always be an extremely small subset of the set of all things that can be **referred to**. Moreover, although one can of course refer to accessible things, *most acts of reference will be to things not in this class*, because most of the world’s business is concerned with other things than the architecture of the internet. (Example: the weather in Oaxaca.)
- Historical note: the word ‘resource’ meant ‘accessible thing’ in the early writings of Engelbart and others. This corresponds fairly accurately to the English meaning of ‘resource’, unlike subsequent W3C usage. The newer notion of ‘information resource’ might be similar to ‘accessible thing’, but it may be narrower.

# Access vs. Reference.

1. Reference has to do with the semantics of language; access has to do with network architecture.
2. Successful reference is part of a communication act between cognitive (language-using) agents. Successful access requires neither cognition nor linguistic communication.
3. Access is mediated by transmission over a communication network, and uses energy. Reference is not a physical relationship and does not require any supporting architecture.
4. Only 'information resources' (?) can be accessed; anything at all can be referred to, even non-existent things.
5. Access, in order to be useful, should be unambiguous. Reference to natural entities, in contrast, is *inherently* ambiguous.

# Reference on the Web is the same as reference off the Web

This is simply obvious, I claim. The Web is a transport mechanism for representations. But how a representation represents, and how the names in it refer, has nothing particularly to do with how the representations are transported. ( Korzybsky: “The map is not the territory”.)

Reference has to do with the meaning of *texts*, broadly construed. The Web has extended the notion of text to hypertext, but adding markup to a text does not change its meaning qua text (\*); and the use of URIs in RDF/OWL does not even qualify as markup. With a few exceptions (owl:imports, rdfs:seeAlso, etc,.) the Semantic Web languages would operate exactly unchanged if the identifiers in them were not URIs at all, and if the Web did not exist.

(\*) One exception might be the ‘gestural’ use of a hyperlink , as in “for more on Julius Caesar, see <a href=“<http://www.ihmc.us/users/phayes/Julius.html>”>here</a>, where the word “here” refers to the accessed entity, much as one might gesture at a library shelf in ordinary conversation. But even such uses, while new in written language, have clear roots in spoken language use.

# Reference on the Web is not determined by architecture

Web architecture does not determine what any names, including URIs, *refer* to. It only determines what they *access*. The relationship between access and reference is essentially arbitrary; it is ours to decide.

Since most of the things referred to by names are not accessible, references to them can only be determined by description, perhaps based on other pre-existing naming conventions. If, for example, a URI is intended to denote the city of Paris, this reference cannot be established by ostention, since Paris isn't accessible. It has to be done by description. And descriptions can never pin down a referent exactly. There will always be some slack, some possible doubt about what exactly is being referred to.

# Reference by description is inherently ambiguous

This claim may seem to fly in the face of common sense, so I will defend it. My point here is basically a re-statement of Quine's thesis of the radical indeterminacy of translation, but applied to communication. However, I inferred it from many observations in actual ontological practice.

What does it mean to say that a name refers to a thing? There is nothing physical, no architectural infrastructure, which could possibly support such a claim: no 'pathway' *from* the name *to* its referent. And yet it seems clear that language does use names successfully to refer.

We say that a name refers when a *use* of the name is sufficient to communicate a thought or a proposition about the referent, during an act of communication. The actual mental processes which constitute 'mental identification' in human communication are mysterious. The best formal account we have, which also applies to the Semantic Web, says that this is a process of inference.

The same considerations apply whether the agents involved are humans speaking English or Webbots drawing conclusions in OWL.



# Reference by description is inherently ambiguous

It is difficult, perhaps impossible, to make enough inferences so as to be sure that what *the recipient* understands a name to refer to is *exactly the same thing* as what the sender had in mind. Communication using referring names is *always* hostage to slips of reference. A formal version of this claim is based on Gödel's theorem: as long as the language used for communication is sufficient to express arithmetic, it will have 'nonstandard' models.

But the point is made better informally, by looking at normal human communication. Take the most direct and unambiguous kind of name, a public proper name of a public entity, say "Paris". This might refer to the central part of greater Paris defined by the *arrondissements*, or to the larger metropolitan area. It might refer to the state of the city *now*, or it might refer to the entire history of the city. In the right context, it can be used to refer to the inhabitants of Paris, the buildings in Paris, the customs of Paris, etc.. Another example: "Everest". (How much does Everest weigh? Where are its edges)

# Reference by description is inherently ambiguous

Ordinary discourse is full of ambiguities which are rarely noted, because they have no practical importance, but which are rendered vivid by trying to agree about 'common sense' well enough to write it down in a formal notation. My favorite example (apologies to those who have heard it before) was a disagreement about whether or not a fitted carpet was 'in' an office or 'part of' the office.

Two competent, intelligent adult native speakers of English each discovered, to their mutual amazement, that another could believe such an obviously false claim. Over an hour of discussion it gradually emerged, by a process of induction from many examples, that they understood the meaning of 'office' differently: for one it meant, roughly, an inhabitable place; for the other, something like a volume of space defined by the architectural walls.

These two people never knew, until this event, that they had different mental meanings for 'office' (in fact, more generally, for 'room'). Presumably this was possible because they had never previously engaged in a communicative act in which their conceptual differences made any difference to the communication: but note that each had, in fact, been understanding the English word 'office' differently. So 'office' and 'room', in this instance, are ambiguous. Probably every word in English is ambiguous.

# Saying more can make reference by description even more ambiguous

Even when a stable situation of mutual reference has been reached, it can be upset by the addition of new vocabulary. Suppose two agents agree, somehow, that “Bill” refers to a particular person. Still, they might have divergent notions of what it means to be a person. Such divergences are already found in ‘standard’ ontologies. For example, Dolce requires a high-level distinction to be made between continuants and occurrents, and a person would naturally be classed as a continuant; but other ontologies reject this contrast and subsume all temporal entities under one heading.

So are all names of persons rendered ambiguous by the presence of this high-level ontological divergence of opinion, so that we have to distinguish Pat\_Hayes-the-continuant from Pat\_Hayes-the-4d-history? For formal reasoning purposes, the difference is important, and indeed confusion between these concepts can produce immediate logical contradictions, such as inferring that I am both 62 years old and 7 years old (because I was once, and a continuant retains its identity through time.) And what of the reasoner who is simply not concerned with this distinction: do we need a Pat\_Hayes-lite as well?

# Saying more can make reference by description even more ambiguous

Adding richer formal ontologies does not reduce ambiguity of reference, but **INCREASES** it, by providing for finer ontological distinctions.

If all you want to say about persons is that they have mailboxes and friends, then you can treat 'person' as a simple category. If however you want to reason about people's lifestyles, travel plans, medical histories and legal rights, then you may be obliged to distinguish finer categories of personhood. If you wish to include persons in a comprehensive 'upper-level' ontology, then more arcane questions about personhood and how it relates to existence must be answered.

The more one is able to say and is obliged to reason about, the more possible distinctions there are to be made: and as the knowledge increases in scope, the more possibilities arise for making distinctions that were formerly impossible.

# Saying more can make reference by description even more ambiguous

What makes this kind of consideration particularly acute is the view that URIs should be **global** and **eternal** in scope. This makes things worse. It means that if there is *any possibility* of some such ontological distinction being made by anyone, anywhere, at any time, then in order to avoid ambiguity of reference, the URI must make *all* the potentially disambiguating decisions ahead of time, as it were: which is of course manifestly impossible, because there is always the possibility of some new distinction being made later.

It is impossible to achieve unambiguous global reference of names by using descriptions. So we should not set out to attempt it, nor pose it as a goal or a 'good practice'. The presence of the Web does not alter the nature of description or of reference, so the conclusion applies there also.

# Since ambiguity is inevitable, let's make lemonade

Since reference by description is inherently ambiguous, it might be better to set out to *utilize* the inevitable rather than try to *minimize* it.

The 'URI crisis' is said to result from phenomena like using the same name to refer to me and my website, or the same name to refer to the weather and the weather report. These are all examples of *overloading (punning)*: using a single name to refer to multiple referents.

Overloading isn't always as bad as it is rumored to be. It can be simply a way of using names efficiently. Natural language is rife with lexical ambiguity which does not hinder normal communication (rose, bank); programming languages routinely do it (+ means both integer and real addition); Common Logic allows a single name to denote an individual, a function and a relation (and IKL adds *proposition* to the list).

# Since ambiguity is inevitable, let's make lemonade

In general, overloading is harmless when it is possible to figure out from the information available at the time which of the various referents is intended by that particular use of the name. Often this is easy when the referents are in different ontological categories, as with a person and a web page.

One way to think about overloading is that an overloaded term denotes a single 'complex thing' which has all the usual referents as parts or facets, selected by implicit selectors. This often works surprisingly well: for example, the syntactic of logic (and RDF and OWL) is enough to 'select' the appropriate referent in the CL and IKL formal semantics. For people and websites it seems to work, also, for most purposes: I am not affected by an attempt to http GET me instead of my website, and my website isn't going to spend any of my salary. Inference is not affected by ambiguity of reference, of course; and the normal infrastructures of interaction with people and websites are sufficiently robust to disambiguate this kind of overloading when they need to be disambiguated.

# Since ambiguity is inevitable, let's make lemonade

Using the same URI to *refer to* me and to *access* my website isn't even an example of overloading. It only becomes that if we assume that the name which accesses the website therefore also refers to the website. But why do we make this assumption? There is nothing in the HTTP specs which seems to have anything particularly to do with *reference*.

A suggestion: names (IRIs) refer to accessible things just when the thing is actually *assigned* that name; and assigning is done only by an explicit naming convention, by pointing to the object and asserting, *this* is called <name>. Unlike reference by description, this kind of naming can be completely unambiguous.

There are two ways to point to an accessible object: by *being* it, or by using a URI which accesses it. (Access is a kind of 'pointing gesture'.) The two kinds of naming convention this makes possible are similar respectively to wearing a name badge, and to having someone point at you and say, "I will call you Spiderman".



# Since ambiguity is inevitable, let's make lemonade

Assigning a name to an accessible entity inside the entity itself makes the name 'permanent' and global in scope. Example: named graphs. For XHTML this could be done by a property in the header. It is under the control of the owner of the resource.

Assigning a name by using a URI to access the thing being 'christened' is under the control of the remote user of the URI, and the name assignment can be local in scope. (Note, this is consistent with the idea that URIs have global scope as *accessing names*.)

But the key point is that we need some actual mechanism for assigning referents to names, even for objects on the Web. Right now we don't have *any*, because access isn't the same thing as reference.

# Addendum

Web architecture does not determine what any names, including URIs, refer to. It only determines what they access. The relationship between access and reference is ours to decide.

WebArch http-range-14 seems to presume that if a URI accesses something directly (not via an http redirect), then the URI must refer to what it accesses.

This decision is so bad that it is hard to list all the mistakes in it, but here are a few.

It presumes, wrongly, that the distinction between access and reference is based on the distinction between accessible and inaccessible referents.

It places the responsibility for deciding the relationship between referring name and access name at the wrong end of the communication channel.

It uses a distinction in how a text is delivered (an http code) to disambiguate the text itself; a category mistake analogous to requiring the the postman dance a jig when delivering an official letter. Since the text bears no trace of its delivery, no disambiguation is achieved by this.

Since the vast majority of names, even on the Web, refer to things which are not accessible, this requires the vast majority of referring URIs to perform a completely pointless act of redirection.

It produces harmful effects by mis-using http codes for an alien purpose. Already, this decision has caused XML schema applications which use the Dublin Core vocabulary to fail standard validation tests. The harmful effects of this redirection requirement will continue, and achieve nothing useful in return.