

The Big Schema of Things: Two Philosophical Visions of The Relationship Between Language and Reality and Their Implications for The Semantic Web

Allen Ginsberg
The MITRE Corporation¹
7515 Colshire Drive
McLean, Virginia
01-703-983-1604
aginsberg@mitre.org

ABSTRACT

This paper presents two philosophical views concerning the relationship between language and reality, and shows how those views can influence one's understanding of the enterprise of knowledge representation. Doing so will 1) shed light on the underlying assumptions of existing practice in the relevant community, and 2) help motivate and pave the way for an alternative understanding that has implications for the representation of knowledge in the Semantic Web. In this paper, I show how the two visions relate to the so called "URI-Identity Crisis" [2]. This involves a semi-technical discussion of the notion that URIs can be used to identify "resources," followed by a philosophical analysis of the motivation and presuppositions that underlie the perceived need to let *Uniform Resource Identifiers* (URIs) play a referential role. The analysis shows how this perception is molded by adherence to what I call "the correspondence vision" of the relationship of language to reality. I then discuss how the notion of "meaning as use," one of the underpinnings of what I call "the holistic vision," leads to an alternative approach to understanding the role of URIs in a Semantic Web language such as the *Resource Description Framework* (RDF). I conclude by showing how this new approach can naturally and flexibly represent situations where diverse communities of users having substantive disagreements lay claim to having the correct understanding of the use of a term, something that seems to be problematic for the approach to Semantic Web languages based on a correspondence vision.

Categories and Subject Descriptors

I.2.4 [Knowledge Representation Formalisms and Methods]

General Terms

Management, Design, Standardization, Languages, Theory.

Keywords

Philosophy, Language, URI, Semantic Web.

1. INTRODUCTION

The phrase "philosophical vision of language and reality" is a concise way of referring to a host of questions, including the following: what is the correct rational explication of the concept of linguistic meaning? What does it mean for a statement to be true (or false)? How are other uses of language, other than making statements, to be accommodated in a theory of meaning?¹

There are basically two overall philosophical visions of the relationship between language and reality. Before describing them it should be noted that I am not attributing these views to any particular philosopher or philosophical school of thought. In a paper of this size and nature it is impossible and unnecessary to delve into the details and nuances of the particular sub-issues involved in what I term a "vision" of the relationship between language and reality. The two visions described in this paper are overviews or synopses based on my understanding of combinations of general recurrent trends or themes in the history of philosophy.

In one vision, which I will dub the "correspondence view," language and reality are two completely independent realms. On this view, the notion of a *language-independent* "truth of the way things are", is held to be a coherent notion. When someone utters a statement what they have said is true because of the satisfaction of certain "truth conditions" for that statement. Furthermore, whether or not those particular truth-conditions obtain is a "fact" that exists in its own right independent of thought and language. A statement's truth-conditions are determinate non-linguistic entities and a statement is true if and only if its *corresponding* truth-conditions obtain. Note that on this view, given two conflicting "total theories" of the world, both of which are equally good in terms of predictive power and other observable properties, there is still an issue as to which one is "really true," because the theories truth-conditions are conflicting. For example, certain theories in physics, such as Newton's theory of gravitation, can either be expressed as a field-theory or as an action-at-a-distance theory, where it is demonstrable that under all circumstances the theories will yield the same observable

¹ The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author.

consequences. According to the correspondence view, only one of these theories can be true.

The second vision, which I will call the “holistic view,” holds that language and reality are not wholly independent of one another. Whereas the correspondence view tends to see language as something that we can use to render or represent an independent realm of truth “piece-by-piece”, the holistic view sees language and belief interwoven into a total system that we use to structure the world we experience. To quote Quine: “Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system [9].”

On the holistic view, the notion of a *language-independent* “truth of the way things are” is, at best, an empty notion devoid of any empirical significance. The idea of *correspondence* - a statement is true because of the satisfaction of certain “truth conditions” for that statement - might be acceptable as a common-sense description of *our understanding of the notion of “truth” as it functions in our language and belief system*. But that is just saying something about our use of the term “truth,” not something about the “ultimate nature of truth.” Finally, on this view, given two conflicting “total theories” of the world, both of which are equally good in terms of predictive power and other observable properties, there is no issue as to which one is “really true.” These two theories are just different ways of synthesizing experience.

1.1 Meaning-as-Truth vs. Meaning-as-Use

Aside from their explicit difference as to whether the notion of a language-independent “truth of the way things are” makes sense, or is a useful explanatory concept, the two visions described do not necessarily seem logically incompatible. However, if one stretches the correspondence view to include the idea that *meaning*, as well as *truth*, depends on correspondence to truth conditions, then a conflict as to the proper analysis of meaning does arise. This “stretching,” together with a doctrine of what can be directly observed and hence, subject to “verification,” is, more or less, what the *logical positivist* school did to arrive at the “verification theory of meaning,” which states that only statements with empirically verifiable truth-conditions are meaningful.

On the holistic view, the meaning of a statement and any of its constituent terms, is not a function of some correspondence to an independent reality, nor does it depend upon the possibility of “reducing” the statement to a set of statements that are capable of being verified in principle (by being able to observe the truth-conditions of the latter). Rather it involves a complicated “web” of connections to other statements and beliefs containing other terms. To quote Wittgenstein “...the meaning of a word is its use in the language game [12].”

1.2 Correspondence and Common-sense

The correspondence view has a certain common-sense appeal. Consider the statement “The Moon is made of cheese.” The terms “The Moon” and “cheese” and “made of” all seem to be meaningful by virtue of their referring to non-linguistic items, and therefore the statement as a whole is meaningful because it has relatively clear truth-conditions.

However, naive adherence to the correspondence view leads to consequences that may not seem altogether commonsensical. Consider the statements “John is a person,” “Jane is an employee,” and “3 is a number.” Each of these meaningful

statements seems to refer to non-linguistic items of a sort different than a concrete object such as the Moon. On a naive correspondence view, statements using the terms ‘person’, ‘employee’, ‘number’, and ‘3’ are capable of being true or false (meaningful) only because those terms refer to some non-linguistic reality. Whether one sees the latter as being independently-existing “forms” or “ideas,” in the manner of Plato, or one sees them as being mental objects (concepts), the point is that there *is* something that constitutes the referent of these terms, and that is why they have meaning. A naive correspondence view therefore leads to *realism* with respect to the status of abstract objects.

1.3 An Example

In order to begin to focus on the ways in which these views can influence our understanding *and* practice of knowledge representation, let us consider an example in some detail. Consider the time-honored case of the concept of an *Employee*. An *Employee* is typically defined as “a person who works for another in return for financial or other compensation.” This, of course, implies that

Every employee is a person. (S1)

Depending upon the particular style of knowledge representation used, this connection between *Employee* and *Person* may be formalized in many different ways, e.g., as a rule, as a subclass relationship, etc. For example in OWL this could be expressed by:

```
<owl:Class rdf:ID="Employee">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Person"/>
  </rdfs:subClassOf>
</owl:Class>.
```

The formal semantics of the variants of OWL and how it applies to this sort of syntactic construct is well-known and not at issue. The issues are the following:

1. Is statement (S1) true?
2. If so, is that in virtue of a correspondence between the terms of the statement and items belonging to an external reality? (Where this correspondence is a mapping that could be specified in the manner of a formal semantics as in OWL).

First, let us consider question (2). If (S1) is true and/or meaningful in virtue of some correspondence to independently existing truth-conditions, what are those conditions? The naive correspondence-based answer is simply to say that there are language-independent *concepts* of *Employee* and *Person* and there is a language-independent *relationship* of *Subsumption*, and these three items are related in a language-independent way that corresponds to the truth-conditions for (S1), namely, *the concept of an Employee is subsumed by the concept of a Person*.

With regard to issue (1) ask yourself if there are no realistic circumstances under which you might come to question the truth of the statement. Can we *a priori* rule out the possibility that someday computers or dogs or cats might become employees but still not be persons? Imagine a world in which the cause of animal-rights gains so much momentum that an “animal bill of rights” becomes legally binding in some region. This bill of rights might accord animals a legal status below that of

personhood, but high enough that animals used for certain “jobs,” e.g., bomb-sniffing dogs, must be treated as legal employees.

Someone who holds the correspondence view might say that those are circumstances in which people might decide to *use* words differently, but that the statement “every employee is a person” would still be true under the original mapping to reality. After all, the concepts and relationships that make (S1) true have an existence independent of any particular language, albeit as abstractions. Those abstractions are (presumably) not subject to change.

Looking at this from the holistic view, one could argue that a term such as “employee” is a clear example of one that should not be viewed as getting its meaning primarily from a correspondence mapping to an external reality, especially in the manner of the naïve correspondence to a universe of Platonic abstractions. On this view, not only were there large stretches of human history in which no employees existed, there were also times when the concept *Employee* did not exist. The *Employee* concept is a creation of our culture. The exact specification of this concept is never totally fixed because our system of concepts, our language, is a highly interconnected web of statements and beliefs that is continually adjusted in the face of our ongoing experience. If we started to admit computers and other non-persons as employees, then they would *be* employees. To argue about whether this would amount to a change in our language versus a change in reality is to argue about nothing. There simply would be no truth to the matter as to whether the class of things that *really are employees* had been discovered to be larger than previously thought, or that the word “employee” no longer referred to the concept *Employee*.

1.4 Correspondence, Holism, and Knowledge Representation

As noted above, these two visions are painted with a very broad brush and are not intended to accurately characterize the views of any particular philosopher or school of thought. Also, I am not claiming that practitioners of computational knowledge representation explicitly espouse one vision or the other. However, if one were obliged to say which vision is more in tune with modern efforts in computational knowledge representation, I, for one, would argue that the latter are more or less completely aligned with a correspondence view.

There are a number of good reasons for that. The correspondence vision is extremely *modular* in nature: it leads to an expectation that we can focus on modeling a piece of reality in isolation, and that language can be used to mirror the structure of that piece in isolation. It is hard to think of anything more central to both the professional practice and aesthetic sensibility of computer scientists than modularity.

Another reason is that the correspondence vision seems to be a kind of commonsense version of the practice of model-theoretic formal semantics. A formal semantics of the OWL version of statement S1, for example, would involve assigning sets of objects to the class-terms used. The statement is true if and only if the set assigned to *Employee* is a subset of the set assigned to *Person*. It is easy to move from that level of abstraction to something that sounds similar but is drastically different: assign *Employee* the set of all employees and assign *Person* the set of all persons. Again the OWL version of S1 is true if and only if the former is a subset of the latter. With such an interpretation it

might seem as if we have not only succeeded in making the OWL version of S1 come out true, but that we have also given it a meaning. Perhaps we have given the OWL statement a meaning in some abstract sense, but it is dangerous to think that we have thereby succeeded in giving it a meaning equivalent to the meaning a competent speaker of English assigns to S1 itself.

It would be folly to suggest that progress in the technologies of the Semantic Web hinges upon which of the positions regarding linguistic meaning is *correct*, and that these issues must be resolved before progress can be made. However, as is often the case with fundamental philosophical issues, experience teaches that there is probably something worth considering in both positions. That this is indeed the case for the issue in question is the burden of the next two sections of this paper.

2. ANALYSIS OF THE URI-IDENTITY CRISIS

The “URI-Identity crisis” refers to a problem that arises from the requirement that a URI can and/or should identify “resources” that are not necessarily “network retrievable [11].” According to [11], for example, “human beings, corporations, and bound books in a library can also be considered resources.” As demonstrated in [8], simply allowing URIs to take on this role leads to an ambiguity in determining whether an occurrence of a URI simply identifies some addressable web-content, e.g., a web page, or is intended to refer to something that does not have a “network location.”

Although the idea of using URIs in such a manner was contemplated very early in the history of the Web, it was only with the advent of the Resource Description Framework (RDF) and the vision of the Semantic Web that a pressing need for such identifiers arose in the community of interest. That is because the motivation behind the RDF is to allow machine-readable statements about things “outside the web,” such as particular human beings, employees of a corporation, etc., to be constructed, where these RDF statements are formal versions of statements that human beings might make about the same things. So a means to refer to resources unambiguously was needed and URIs were thought to fill that need.

In order to make the ensuing discussion precise and less verbose, for the remainder of this paper, the term ‘URI’ by itself is used to refer to the notion of a certain syntactic construct as defined by the appropriate W3C recommendations. The term “R-URI,” which is a shorthand for “referential-URI” or “representational-URI,” will be used to refer to the notion of using URIs to explicitly *refer* to some resource, whether that resource be on the web, off the web, abstract or concrete.

2.1 Technical Analysis

To understand why using URIs as general identifiers of resources, i.e., as R-URIs, is a problem, recall that the dominant use of URIs is in conjunction with the HTTP protocol. In HTTP, URIs (in the guise of URLs) are used to provide addresses of “locations on the Web.” A URI (URL) can be used to retrieve a web page if that web page happens to be at the location addressed by the URI. This way of using URIs is no more difficult to understand than it is to understand how a postal address can be used to find a particular geographical location.

Now suppose, in analogy to the situation with R-URIs, that we wanted to use postal addresses to *identify things*. To make this discussion easier let’s say that **A** is a basic syntactically well-

formed US address like “123 A Street, Any Town, Any State, 12345-3333.” Then we might say that any base address **A** denotes the “abstract mailbox” to which things addressed with **A** will be delivered. We say the “*abstract* mailbox” because the post office will deliver the mail to whatever actual concrete mailbox is at the address, but that may vary from one day to the next. If the mailbox that is addressed by **A** at the current time is stolen by an envious neighbor and put in the location with address **B**, the post office would not start delivering packages labeled **A** to the new location of the stolen mailbox.

Continuing with the analogy, we might extend the syntax to allow for additional uses of addresses. For example if `<name>` is a placeholder for a proper name, we could allow **A.<name>** to denote an individual with that name. If someone sends a piece of mail addressed **A.<name>** the post office will know that the mail should be delivered to a person of that name at the given location.

Extending this idea again, we could let

A.house.kitchen.refrigerator.thermostat (1)

denote the *abstract* thermostat in the *abstract* refrigerator in the *abstract* kitchen of the *abstract* house at the location to which things addressed to **A** will be delivered. Again we say the “abstract” such-and-such because we want to *address* whatever concrete such-and-such (house, kitchen, etc) might actually exist at the specified location at that time.

Now there are two problems with this scheme. First, the post office is unlikely to deliver mail to the thermostat in your kitchen refrigerator. On the other hand, in the future, appliance service companies that monitor devices remotely might have use for such an addressing scheme. The second problem, however, is not so easily dealt with. Because of the nature of the addressing scheme employed, the fact is that we haven’t really succeeded (yet) in identifying any *thing* in the way that, for example, a serial number for a part, such as a thermostat, uniquely identifies that part. As we have seen, address (1) can be used to identify *any* thermostat that happens to be in the “right place at the right time,” so to speak.

This is not to say that we could not extend the scheme to identify the *actual* thermostat that resides at **A** at *this* particular time. We could use a time-stamped version of (1) as a *name* for that particular thing. But without some such explicit extension of the scheme if we say that (1) as it stands can *unambiguously identify* a particular thing, then we might as well say that **A** unambiguously identifies a particular mailbox (and when you move you had better take your mailbox with you if you want to get your mail).

Bringing the discussion back to URIs, it is clear that there is a tension between the normal use of an address belonging to an established addressing scheme, namely to “locate” the thing that happens to be the current occupant of the addressed location, *and* using *that same address* to essentially *name* the individual so located. If that individual “moves” to a different address, it doesn’t take the name with it. If you really want to label that individual using the address you will constantly need to qualify the use of the label, e.g., “the person who used to live at **A** in 1999.” Otherwise how do you know which of the individuals that was at some time addressable by **A** is the intended referent? If you are content to use **A** as an identifier for whatever happens to be at that location at the time you use the address, that is a

coherent double use of the address, but it is not what people usually have in mind when they talk about “identifiers.”

Using the same syntactic entity to do two entirely different jobs leads to difficulties that can easily be avoided. Suppose we stipulate that when we use URIs as addresses and when we use URIs as R-URIs we are talking about two independent uses of the term “URI.” Suppose we have some way of making sure that every system and piece of software that uses URIs is able to distinguish those uses in every possible circumstance. For example, an address appearing as in (1), would be interpreted as an address in a context. But to use an address to identify or refer to the object located at (1) we would write:

[A.house.kitchen.refrigerator.thermostat] (2).

So if we wanted to say something about the thermostat itself, wherever it might be, we would use form (2). Would this dispel the URI-Identity crisis?

The answer depends on how one interprets the way in which this scheme is supposed to operate. If we want to give this particular thermostat a name or id so that we can always refer to it, then the fact that we used the address in (2) to do *that* is completely accidental. We could use any term we want as a name of that thermostat as long as that intention is clear to us and anyone else we want to understand our use of the term. We could just have easily written any of the following

[thisThermostat], [T-135], [x], [George] (3)

to function as a name of the thermostat. As long as we and others *know* that, the scheme would work.

So this brings us to the crux of the matter and what seems to be the real issue at the heart of the URI-Identity crisis, namely, how does any person or machine *know* what it is that an identifier refers to? To use philosophical terminology, the key issue is an *epistemological* one. Looking at the proposed naming scheme from an epistemological point of view, it is clear that the use of an address as a constituent of the identifier, as in (2), is *not* accidental. The idea is that the address is supposed to help us figure out, *to know*, what the name is referring to.

This scheme could get around the problem caused by the fact that the content of an address can change by providing for variables that are persistent and have global scope. For example, suppose the following assignment statement is executed at time *t*

O = [A.house.kitchen.refrigerator.thermostat] (4).

At time *t*, the given address is dereferenced and whatever object is returned is assigned to the variable *O*. That object will be the referent of *O* for as long as *O* is in scope (and is not reassigned) regardless of what might come to occupy the location given by the address in the future.

Bringing this back again to the case of R-URIs, when we dereference a URI (*qua* web address), we get back whatever is at that location at the time. It would be a miracle indeed if that ever turned out to something like a thermostat. Whether we call what we get back a “web page” or a “representation of an information object,” the point is that it can never be a thermostat *per se*.

But all is not lost. Something on the Web may not *be* a thermostat, but it could be a *representation* of a concept of a thermostat. Indeed it could even be a representation of a particular thermostat. And isn’t a representation of a particular thermostat something that can be used to identify that object? So if, for the moment, we imagine that the address in (1) is really the

address of a web page, then what is dereferenced in (2) is the web page at that address, which gives us something that can in turn, in some sense, be “dereferenced” to give us the thermostat. So instead of (4) we need to write something like

O= <[A.house.kitchen.refrigerator.thermostat]> (5)

where the angle-brackets indicate the kind of “dereferencing” appropriate to representations of information objects as opposed to addresses.

The format given in (5) is indeed an abstract representation of the positions taken by [4] and [8] in two recent papers. The specific details of the two proposals need not concern us in this paper. Both of these also seem consistent with the overall vision for R-URIs taken in [1]. Both of the proposals use a syntactic mechanism to determine whether a URI is being used purely as a web address or as an R-URI. But these approaches go beyond providing a syntactic disambiguation mechanism. They also propose a mechanism for establishing the reference of an R-URI that hinge upon using the URI part of the R-URI as an address. In both cases the mechanism involves 1) first using the URI part of the R-URI as a web address, 2) retrieving the contents of the address, and 3) using those contents to determine what the referent of the R-URI is. This is more or less the same scheme as is represented in the discussion that culminated in (5).

2.2 Philosophical Analysis

As we saw in the previous section, the URI-Identity crisis really boils down to an epistemological issue: how do we know or figure out what an R-URI refers to? Saying that the URI component of an R-URI always addresses a retrieval-able representation of an information object that can in turn be used to refer to some *resource* (no matter whether “resources” are thought of as concepts or concrete objects), does not solve *that* problem. Rather, it shifts the burden of the analysis to explaining how an information object can be used to determine the intended referent.

Let’s step back for a minute and ask *why* we should be concerned with figuring out what the reference of an R-URI is? Recall that the point of RDF is to allow machines to interpret and make statements about resources that would be analogous to human-generated statements about the same resources. Clearly in order to know whether or not a given RDF statement has a human-interpretable meaning, and what that meaning is, we need to know what the RDF statement means. Therefore we need to know what the constituent terms in the RDF statement refer to.

This should sound familiar. It is basically an application of the correspondence view of language to RDF statements. The formalism used in (5) basically mirrors the idea that in order for a term to have meaning it has to be about some non-linguistic reality. In the case of RDF, the information object represented by a web page is used to establish this correspondence.

The last sentence in the preceding paragraph is, upon reflection, a rather strange thing to say. How can we “establish” a correspondence between a term and some non-linguistic object? Certainly we can *stipulate* for some limited purpose that a term “stands for” something, such as we contemplated in the discussion leading up to (2) and (3) above. For example, I can stipulate that for the remainder of this paper the phrase “my reading glasses” will refer to the thermostat in a certain refrigerator. The statement “My reading glasses is set to a temperature of 40 degrees F,” would then “mean” that said thermostat has been set to 40 degrees F. But have I thereby “established” that the term “my reading glasses” corresponds to that thermostat?

Not at all. Language is a rule-governed activity. The rules of a language are not established by a person or a committee or a single community of users at some point of time. The root of the URI-Identity crisis lies in the idea that we can somehow *engineer* what the terms of a language refer to, both in general and in some particular instance.

Why do people believe this? The reasoning must be something as follows. RDF is an artificial formal language. We can give the statements of such a language meaning by using the methodology of formal semantics, which show how every statement in a formal language can be given a meaning by assigning the vocabulary of the language various objects. Often such interpretations use abstract domains such as numbers and sets of numbers, *but when we make assignments to terms in RDF, the assignments need to match the same assignments that are given to corresponding natural language terms*. For example, if we intend to use the R-URI

[<http://www.ontology.net/types.owl#c10507>]

to designate a resource that corresponds to the resource designated by the natural language term “employee,” then we need to make sure that the page at this web address represents the same information object that the term “employee” represents in natural language.

While I believe that we can engineer, i.e., stipulate, what the terms of a formal language refer to, I do not feel that the case of RDF fits into that realm. The reason is that the intended interpretation of RDF terms is supposed to match the interpretation of some corresponding natural language term(s), and I do not believe that *we* can stipulate what the later mean. The best we can do is to say that we intend to use a certain RDF term in a manner similar to a certain use of a corresponding natural language term. When possible, we can use formal tools such as ontology languages to *explicate* our understanding of those natural language usages. OWL representations of natural language statements such as S1 above can be used to do *that*, but they cannot be used to *establish* the meaning of natural language terms, because nothing can be used to do that.

As my remarks in the preceding paragraph indicate, one’s attitude to the question of establishing reference for RDF terms is a *matter of choice*. One can take the correspondence view that natural language terms have meaning by virtue of correspondence to non-linguistic reality, and that the meaning of abstract terms such as “employee,” “concept,” “humanity,” “justice” can be determined by a dialectical process, rational reflection, or direct intuition of a realm of concepts or Platonic forms, or some other method. On that view, the idea of *engineering* the meaning of RDF terms makes sense regardless of the connection to natural language: one establishes the meaning of the RDF term by linking it to an information object that represents the corresponding concept.

However, this choice has consequences for the representational power of Semantic Web languages such as RDF, a topic we explore in the next section.

3. TOWARDS A MEANING-AS-USE MODEL

We have seen that the URI-Identity crisis boils down to an epistemological issue, namely, how does any person or machine *know* what a given R-URI refers to? If representations of information objects can be used to establish reference, and such representations can be processed by machines and humans, then

the answer lies in implementing those representations and processing them in the right way. That is at least part of the message of [1][4] and [8]. Let us call that view, “the received view of R-URIs.”

I have suggested that the received view of R-URIs is based on a naïve correspondence view of language. This view can lead one to think of “establishing” the reference of terms in a language like RDF as something that can be done by simply figuring out what terms in a natural language refer to and making the RDF terms refer to the same things.

3.1 Intended Usage Records

But if we say that language does not work *that way* - that terms like “employee” do not get their meaning by correspondence to some determinate concept, but that the meaning of terms is to be found in the way they are *used in a complex language-belief system* - then what becomes of the epistemological issue? Does it just go away? How does viewing meaning “as use” change anything?

First of all, viewing meaning as use necessarily changes our interpretation of an R-URI works. Instead of using the URI component of an R-URI as an address for a representation that establish reference, we use it as an address for a *record that establishes an intention to use an RDF term according to some, presumably established, usage*. The R-URI itself is just a syntactic mechanism that lets us, and machines, know that the URI addresses such a record.

But isn’t there a similar epistemological quandary for this view? Given an R-URI, how does a machine or human know which *usage* is intended? This is, however, not the same issue as arises for the received view, and it is by no means an epistemological quandary or a crisis. Usages of terms can be identified and tagged as distinct usages without being formally defined. Lexicographers have been doing that for centuries. The intent to use a term in accordance with a usage can be established by creating a record containing the tag for the intended usage. The same record can contain information that explicates or gives examples of the intended usage. This information can be processed by a human or machine to better understand what the intended usage is.

In order to make this clearer, figure 1 shows a schematic of what an *Intended Usage Record* (IUR) might look like. Every IUR would contain a *Usage-ID* field that would be a unique identifier of some presumably established usage of the term(s) listed in the *Language-Term* field. As shown, the latter contains pairs of <language, term> identifiers. This means that each of these terms have usages in the given language that correspond to the intended usage. A key requirement on the administration of such IURs would be to make sure that a given Usage-ID is always linked to the same value for the *Language-Term* field in every IUR. Letting Usage-ID 101.1, for example, occur in one IUR with Language-term field <English,employee> and in some other IUR with Language-term field <English,ice> would obviously cause problems. Therefore, even though these two fields are listed separately in the schematic, it should be understood that, in practice, someone creating an IUR would not be able to arbitrarily assign Usage-IDs to some Language-Term value. Some sort of dictionary of usages would be maintained and made accessible by a standards body. The connection between a Usage-ID and the appropriate Language-Term value would be given by the dictionary. Someone creating an IUR would look up the desired

Usage-ID: 101.1		
Language-Term:	ENGLISH	EMPLOYEE
	SPANISH	EMPLEADO
Explicated-By: http://www.ontologies.net/myontology.owl		

Figure 1. Schematic of an Intended Usage Record

Language-Term value in the dictionary, locate the desired usage, since terms can (and usually do) have more than one, and then use the associated Usage-ID in the IUR they are constructing. Mechanisms and procedures for creating new Usage-IDs would be required, but that is not an insurmountable obstacle to implementation of such a scheme.

So now an R-URI is interpreted by looking up the IUR it addresses and getting the Usage-ID value. That would enable a machine (or human) to know that two or more occurrences of a term in various RDF statements are intended to be used in the same way. Looking at the associated Language-Term field value in the dictionary would be helpful for a human wanting to get more information as to the intended usage, but is not likely to help a machine. That is where the *Explicated-By* field of the IUR comes in. This is an (optional) field whose value is machine-readable information useful in understanding the intended usage. In terms of present day technology, that information would be represented in a language such as RDF or OWL.

3.2 Is Pluto a Planet?

One of the key benefits of adopting the meaning-as-use point of view is that *real disagreements about things* can be represented and reasoned about in a coherent fashion. A machine should be able to represent and understand a situation in which contradictory competing statements using the same term are entertained as the result of conflicting theories or belief-systems. That is something that, I, for one, am unable to understand how to accommodate in the received view of R-URIs.

A contemporary example can be used to illustrate this point. Astronomers have debated what a planet is for some time. One school of thought contends that Pluto is too small to be a planet. This is a debate recently rekindled by the discovery of a solar system object called “2003 UB₃₁₃”, an object bigger than Pluto, but still relatively small compared to other planets. The details of the discovery and the current scientific debate are related in [5]. For our purposes the following statements quoted from [5] will frame the problem:

Alan Boss, a planet-formation theorist at the Carnegie Institution of Washington, called the discovery "a major step." But Boss would not call it a planet at all. Instead, he said Pluto and other small objects beyond Neptune should be called, at best, "Kuiper Belt planets."

We may postulate two scientific schools of thought concerning planets, one which accords both 2003 UB₃₁₃ and Pluto planetary status, and one which accords neither of them such status. We may suppose that there are two OWL ontologies that explicate these competing theories, astro-1.owl and astro-2.owl, such that the statement “Pluto is a planet” is true in the former but false in the latter.

While the holistic vision would analyze this situation as one in which there is no truth to the matter as to whether these schools of

thought are arguing over language or planets, a holist would still see this as a *substantive disagreement*. How the term ‘planet’ is used in our belief system has important consequences in our lives. On the correspondence vision the statement “Pluto is a planet” is either true or not depending upon whether or not the corresponding truth-conditions obtain. On this view, these two schools of thought disagree over what reality is like.

Ironically, however, using the received view of R-URIs, it turns out that this disagreement can not be represented as a disagreement at all. To see why, we need to further characterize the situation. Let ‘qn1’ and ‘qn2’ be abbreviations for

‘<http://www.ontologies.net/astro-1.owl>’

and

‘<http://www.ontologies.net/astro-2.owl>’

respectively. Let ‘qn1:planet’ be the R-URI for planet in the first theory and let ‘qn2:planet’ be the R-URI for planet in the second theory. (We may assume that in all theories about the solar system there is agreement about how the term ‘Pluto’ should be interpreted, so we will simply let that term stand for an R-URI by itself.)

Now consider the following statements:

Pluto is a qn1:planet (P1)

Pluto is not a qn2:planet (P2)

When interpreted using [astro-1.owl](#), P1 will be true. When interpreted using [astro-2.owl](#), P2 will be true². The important point, however, is that the R-URIs `qn1:planet` and `qn2:planet` *cannot possibly be referring to the same resource*. Under the received view the resources identified by `qn1:planet/qn2:planet` are determined by the information encoded in [astro-1.owl/astro-2.owl](#). But, by hypothesis, [astro-1.owl](#) represents reality in such a way that Pluto is in the extension of `qn1:planet`, while [astro-2.owl](#) represents reality in such a way that Pluto is *not* in the extension `qn2:planet`. If these two R-URIs referred to the same resource, in this case, the same concept, then Pluto would both be and not be an instance of that concept, which is impossible. This is just another way of saying that there is no model for the theory formed by merging [astro-1.owl](#) with [astro-2.owl](#) if the two R-URIs are assigned the same resource. The bottom line is that, since the two R-URIs must refer to different resources, statements P1 and P2 really do not represent a disagreement about anything.

The problem here is a kind of “semantic catch-22”: the only way the R-URI for a concept like *planet* can have meaning on the received view is for it to correspond to a concept defined by the information object addressed by its URI component, namely the [astro-1.owl](#) ontology. If, according to the definition of *planet* in [astro-1.owl](#), Pluto has what it takes to be a `qn1:planet`, then the statement that Pluto is *not* a `qn1:planet` is basically self-contradictory. A wedge needs to be driven between the way a term is intended to be used and a particular representation that can be used to fix the reference of a term. The received view simply offers no way to do that.

² As there is no explicit propositional negation operator in OWL or RDF it would be more accurate to use “Pluto is a `qn2:planet`” and say that when using [astro-2.owl](#) P2 will be false. That does not alter the point of this example.

Under the meaning-as-use view of R-URIs, however, there are more representational “degrees of freedom,” so to speak, so disagreements like this one can easily be represented as substantive disagreements. In order to see how that can be done, refer to figure 2. Since R-URIs are not seen as “resource identifiers” their associated URIs do not need to address representations that identify those resources. The URI component of a R-URI addresses a IUR. In turn the *Explicated-By* field of an IUR can be used to provide access to ontologies such as [astro-1.owl](#) and [astro-2.owl](#).

R-URI:		
http://www.FirstSchool.astro/planet		
Usage-ID: 503.1		
Language-Term:	ENGLISH	PLANET
Explicated-By: http://www.ontologies.net/astroOnt-1.owl		
R-URI:		
http://www.SecondSchool.astro/planet		
Usage-ID: 503.1		
Language-Term:	ENGLISH	PLANET
Explicated-By: http://www.ontologies.net/astroOnt-2.owl		

Figure 2. Intended Usage Records for Case About Planets

So in this situation we create two R-URIs for ‘planet’, with URI components

<http://www.FirstSchool.astro/planet.iur> (IUR1)

<http://www.SecondSchool.astro/planet.iur> (IUR2)

to provide addresses for the IURs (where the “.iur” extension is used to indicate the address of an IUR).

If we use ‘qn1’ and ‘qn2’ to abbreviate IUR1 and IUR2 respectively, then the two statements P1 and P2 can be re-used to represent the disagreement over whether or not Pluto is a planet. But this time, this is a situation that both machines and humans can easily appreciate. The reason is that the two R-URIs resolve to IURs having the same Usage-ID (503.1). That makes it easy to see that the usage of those terms in P1 and P2 is intended to be the same. Access to the respective IURs also shows that the disagreement represented by P1 and P2 can possibly be understood by looking at the associated ontologies in the *Explicated-By* fields. In the case of P1 and P2 a machine could easily see that both come out true using the given theories and that there is therefore a conflict between the two theories.

3.3 Avoiding Ontological Commitments

The planetary case raises another interesting question regarding the way in which Semantic Web representation languages function. This is the issue of how to avoid unintentionally committing oneself to RDF statements, or their consequences, that have been asserted somewhere on the web simply by using certain

vocabulary. (This is similar to, if not the same as, the issue of “social meaning” [3]).

For example, suppose an astronomer, A1, wanted to have a Semantic Web site that made assertions concerning planets, but that A1 did not particularly want to become embroiled in the raging debate concerning what is and is not a planet. Assume that astro-1.owl and astro-2.owl are the only ontologies that any astronomer would want to use as authoritative. Currently A1 would either have to link occurrences of “planet” with one ontology or the other, thus appearing to take a position on the issue. Alternatively, A1 could decide to develop a neutral ontology. The problem with doing the latter is that no one else would know about the terms in A1’s ontology. For all intents and purposes, A1’s use of the term “planet” would be a totally different concept from those recognized as being authoritative.

The meaning-as-use approach offers a nice way out of this dilemma. To see how, consider figure 3.

R-URI:		
http://www.UndecidedSchool.astro/planet		
Usage-ID: 503.1		
Language-Term:	ENGLISH	PLANET
Partially-Explicating-Competitors:		
http://www.ontologies.net/astroOnt-1.owl		
http://www.ontologies.net/astroOnt-2.owl		

Figure 3. Intended Usage Records for Avoiding Ontological Commitments Example

Here is how A1 gets to use the term “planets” in a way that everyone else will see refers to the same usage, but that avoids taking sides on the theoretical squabble. Figure 3 shows the IUR associated with the R-URI

http://www.UndecidedSchool.astro/planet.iur (IUR3)

which is what A1 uses for ‘planet.’ What is different about IUR3 is that it does not have an *Explicated-By* field, but rather a *Partially-Explicating-Competitors* field that lists the addresses of the two competing theories. Intuitively, by using IUR3 A1 is signaling an intent to use the term ‘planet’ with the same usage (503.1), but also signaling that the term is not fully explicated in either of these two competing theories. The question is, how are statements using R-URI

http://www.UndecidedSchool.astro/planet

to be evaluated for truth-value?

The answer is to use a semantics that assigns truth or falsity only to statements that are assigned truth/falsity by both of the partially-explicating-competitors. Any statement, such as, “Pluto is a planet,” that is assigned true by one and false by the other, is not assigned a truth-value. This interpretation can clearly be extended to cases involving any number of competing theories.

4. DISCUSSION

The discussion and examples in this paper have focused on general terms such as “employee” and “planet.” Does looking at the enterprise of the Semantic Web from a holistic standpoint

have implications for R-URIs that correspond to so-called “singular terms.” such as proper names?

It seems to me that the same issues arise. For example, consider the proper name ‘George Washington.’ Suppose we want to create an R-URI that refers to George Washington (the first president of the USA, etc.). On the received view this R-URI would address some information object that would be used to figure out the referent of this term. Presumably that information object would imply statements such as “George Washington was the first president of the USA.” Now as long as the information object is taken as being definitive of the referent of the term, the statements entailed by the information object would amount to necessary or analytic truths, exactly as in the planet example. Anyone who wanted to use the R-URI for George Washington to put forth a theory that conflicted with some entailment of the corresponding information object, e.g., a George-Washington-impostor was actually the first president, would be saying something necessarily false.

If the key issue of the URI-Identity crisis is *how to know* what a given term refers to, then whether the referent is thought to be a concept (as in the case of general terms) or an individual (as in the case of proper names), the philosophical argument is the same. On the correspondence view, reality is determinate, concepts are determinate and the latter can be used to definitively mirror the former in an appropriate language. Figuring out the reference of terms in such a language is a matter of linking them to the right piece of reality. For general terms this linkage might be justified by means of formal definitions of concepts. For proper names, which are generally thought of as being devoid of meaning in a technical sense, the linkage might be justified by a causal theory of reference [6].

On the holistic view, the reality-belief-language “web” is indeterminate. Figuring out the meaning/reference of terms in a language is a matter for empirical investigation [10]. To know what a term means or refers to is to have an adequate theory of its use(s) in a given linguistic community. As with any theory, a theory of meaning and reference for a particular linguistic community may be underdetermined by all possible evidence. In other words, two or more conflicting hypotheses concerning the meaning/reference of terms in a language may account equally well for all observable behavior. In such a case, a holist would say that there is simply no truth to the matter as to which of these conflicting hypotheses is correct.

Whatever one may think of the holistic vision as a philosophical theory, the notions of linguistic indeterminacy and meaning-as-use, offer valuable insights into language. The notion of IURs presented here is one example of an application of these ideas to computational knowledge representation. Allowing R-URIs to address IURs is entirely different than the proposal to allow the meaning of R-URIs in a Semantic Web context to be determined by local context and explicitly referenced documents [7]. The latter proposal falls within the received view and is philosophically linked to the correspondence vision precisely because it is a proposal about how Semantic Web meanings ought to be *engineered*. The IUR approach does not offer “yet another way” to assign meanings to Semantic Web terms. To assign an R-URI an IUR is not to assign that term a meaning. It is to record an intention to use a term according to a certain usage: it commits one to following certain rules of use that guide the members of a particular linguistic community. Conceived in this light, usages, by their very nature, are not determinate. For example, referring

back to the discussion in section 1.3, to say that one intends to use the term ‘employee’ according to the dominant usage in current American English, does not, in itself, determine whether one will assent to the statement that “A dog can be an employee.”

5. CONCLUSION

Old philosophical issues never die, and only unimportant ones fade away. The fact that questions concerning the nature of linguistic meaning, reference, truth, etc. should arise in the context of the Semantic Web is neither surprising nor unwelcome. And while some may take *that* as a sign of a lack of fundamental scientific or technological progress, the fact is that most, if not all, great scientific theories tend to butt up against fundamental issues that go beyond the narrower confines of a particular field.

In this paper I have shown that there is a link between the two alternative philosophical visions described and alternative approaches to the role of R-URIs in a Semantic Web representation language. The holistic vision has, up until now, had little influence on the community of interest. As I have shown in this paper, taking its tenets seriously can lead to valuable insights that result in potentially useful approaches to difficult issues. The ideas presented in section 3 of this paper need to be fleshed out and more precisely formalized. But even in their current form, I think that they offer an approach that is very much in line with the vision of the Semantic Web as a cooperative venture between the peoples of the world and their machines.

6. ACKNOWLEDGMENTS

I thank Leo Obrst, Neal Rothleder and Marnie Salisbury for discussions, comments, and corrections. I also thank the anonymous reviewers for their critiques.

7. REFERENCES

- [1] Berners-Lee, T., What do HTTP URIs Identify?. <http://www.w3.org/DesignIssues/HTTP-URI.html>.
- [2] Clark, K.G., Identity Crisis. <http://www.xml.com/pub/a/2002/09/11/deviant.html>.
- [3] Clark, K.G., The Social Meaning of RDF. <http://www.xml.com/pub/a/2003/03/05/social.html>
- [4] Halpin, H. and Thompson, H., Web Proper Names. <http://www.webpropernames.org/>
- [5] http://www.space.com/scienceastronomy/050729_new_plane.html
- [6] Kripke, S. *Naming and Necessity*. Harvard University Press, 1972.
- [7] Parsia, B. and Patel-Schneider, P., Meaning and The Semantic Web. Proceedings of WWW2004. pp. 306-307.
- [8] Pepper, S.. Curing the Web's Identity Crisis. <http://www.ontopia.net/topicmaps/materials/identitycrisis.html>
- [9] Quine, W.V.O., Two Dogmas Of Empiricism. *The Philosophical Review* 60 (1951): 20-43. (<http://www.ditext.com/quine/quine.html>)
- [10] Quine, W.V.O., Ontological Relativity. *Ontological Relativity and Other Essays*. Columbia University Press, 1969.
- [11] RFC 2396: Uniform Resource Identifiers (URI): Generic Syntax. <http://www.ietf.org/rfc/rfc2396.txt>.
- [12] Wittgenstein, L., *Philosophical Investigations*. New York: The Macmillan Company, 1963.