

Chapter 6

Ontology and Classification

Ontology is the study of what exists. This is closely related to metaphysics, the study of the nature of reality. In general, ontology deals with the identity of things in the world while metaphysics deals with existential causes such as God or the study of first principles, concepts that underlie all of reality. This will first addresses historic and modern ontologies starting with ancient views of existence and their transformation by modern science. Second, we will examine formal ontology and the role that the introduction of formal logic and machines play. Finally, a theory for modern interdisciplinary ontology and a specific ontology, the Quanta Generic Ontology (GO), is presented. Throughout this discussion, clarification will be made by visual example.

6.1. Ontology and Philosophy

Ontology, like epistemology, begins with experience: we see the world then attempt to understand it. What we first see are a myriad of objects with different visual qualities. Aristotle referred to the objects of the world as *primary substances*.

"Substance, in the truest and primary and most definite sense of the word, is that which is neither predicable of a subject nor present in a subject; for instance, the individual man or horse. But in a secondary sense those things are called substances within which, as species, the primary substances are included; also those which, as genera, include the species. For instance, the individual man is included in the species 'man', and the genus to which the species belongs is 'animal'; these, therefore - that is to say, the species 'man' and the genus 'animal' - are termed secondary substances." [6-1]

This is best illustrated by example. Consider the circles in Figure 6.1.



Figure 6.1. Objects with different properties

The primary substances are the two circles themselves, i.e. the actual circles. The secondary substance is the *concept* of circle, what Aristotle refers to as *species (kind)*. This idea is further illustrated if we introduce other kinds of objects (Figure 6.2).



Figure 6.2. Different species of objects with different properties

Here we have two species: circle and square. However, we notice that all of these objects share the property of having a bounded interior. The concept of

shape is in this case a genre to which circle and square belong. From the point of view of type, we might say that all that exists are shapes.

Plato emphasizes that both primary and secondary substances are to be considered equally real. In fact, in his view, the idea of circle is more real as it is the *form* of any circles we might draw.

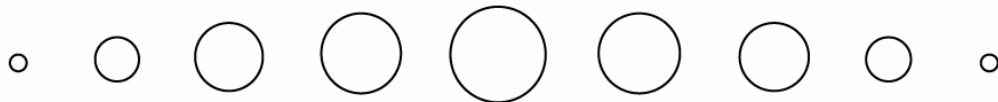


Figure 6.3. An object as something changing in time.

There are other ways one might reason about the existence of objects. A sequence unfolding in time is shown in Figure 6.3. The circle appears from nothing and then disappears. In the natural world, this is common as things are born, grow and die. Yet it suggests that circles and squares are impermanent things. If so, from what do they grow and die?

The explanation arrived at by Democritus is that the world must be composed of some other *matter* of which natural objects are composed, and it is through their composition and decomposition that impermanent things come into being.

"The material cause of all things that exist is the coming together of atoms and void. Atoms are too small to be perceived by the senses. They are eternal and have many different shapes, and they can cluster together to create things that are perceivable. Differences in shape, arrangement, and position of atoms produce different things. By aggregation they provide bulky objects that we can perceive with our sight and other senses." [6-2]



Figure 6.4. Objects composed of a more fundamental material.

How is this related to ontology? From Democritus' view, Figure 6.4, we see that it is incorrect to say only circles and squares exist. In fact, if anything is to be considered as fundamentally existing it is the atoms (dots) that make up these shapes. Democritus introduces *voids* as well, since the atoms must exist in some space otherwise they would all touch one another. Thus, at this point our ontology must be composed of:

atoms (dots)
voids
shapes
 circles
 squares

What of the circles and squares that are *filled* in? For Aristotle the *quality* of being empty or filled is also existent. Clearly the idea of being "filled" is not a shape itself but a property of a shape. The idea of being "filled" is also neither an atom nor a void, since the atoms themselves are not filled but the circle

and square are. Yet since one circle is filled, and another circle empty, it must be some real property of these particular objects. Thus the quality of being "filled" must be an existent thing as well (and also the quality of not being filled, i.e. empty). The taxonomy becomes:

- atoms (dots)
- voids
- shapes
 - circles
 - squares
- styles
 - empty
 - filled

What begins as a simple observation reveals that many things exist in addition to the observable object. To describe what we see as a single filled circle we have no choice but to introduce the concepts of matter, quality and shape. Such investigations led the ancient philosophers to consider the world as dependent on certain *universals*. Two examples of this are Aristotle's universal categories and Porphyry first tree of knowledge of existent things, shown in Figure 2.1.

Table 6.1. Ontologies according to a) Aristotle's Categories and b) Porphyry's Tree of Knowledge

Substance	Substance
Quantity	Body (material)
Quality	Living (animate)
Relation	Animal (sensitive)
Place	Human
Time	Beast
Position	Plant (insensitive)
State	Mineral (inanimate)
Action	Spirit (immaterial)
Passion.	

The *problem of universals* deals with the essential nature of reality. As we examine atoms more closely we see that they too are in a state of change. Perhaps some other substratum underlies their existence as well (i.e. sub-atomic particles). How would we classify the modern scientific notion of the atom which is quite different from Democritus's conception? Through vibration it is animate in some sense, but not living. If we require that animate means alive then the atom is not living but mineral, which is also untrue in modern terms. This issue of universals is a challenging one as the expansion of scientific knowledge, differences in belief, and differences in culture result in a dynamic, transformative concept. A system that deals with generic ontologies must be able to handle these continual changes and variations in belief.

Another issue has to do with the question of reality versus fantasy. Do dragons exist ¹ ? An important distinction is to be made between physical things, such as water and trees, and non-physical things, such as emotions. But what do we do with dragons? They are physical since they possess a body, legs and head. However they are different from water and trees in the sense of being non-actual. Consider Table 6.2.

Table 6.2. Two dimensions of reality: 1) physicality, 2) actuality

	Actual	Non-Actual
Physical	water	dragons
Non-physical	love	God (?)

With the goal of developing a truly generic ontology, it is necessary to be able to express things that have any of the following states:

- physical that which has a form, casts shadows, etc.
- abstract physical that which has a form, but is abstract (circles)
- non-physical that which has no form (love)
- actual that which is seen in the world (water)
- non-actual that which is never seen in the world (dragons)

¹ I am referring here to the mythical fire-breathing sort that flies. The Komodo Dragon is actually lizard from parts of Indonesia.

Which of these is selected as a primary means of classification is somewhat arbitrary. Physical and non-physical are natural divisions, but not always clear. We have a natural grasp of that which is physical - capable of having volume, surface, or form - versus non-physical. What is actual or non-actual, however, is even more dependent on philosophical outlook. Quine introduced the idea of possibilism in [6-3], in which he deals with statements such as "dragons possibly exist". The issue is captured nicely by Nico Cincharella, who compares possibilism to actualism [6-4]. The possibilist says everything exists, including things that possibly exist such as dragons. The actualist says that only actual things exist.

For the actualist (materialist), dragons do not exist since they only *possibly* exist. For the artists all things exist - even dragons, they just have the status of not being actual. What actually exists is a matter of perspective. Due to the subjective nature of belief regarding reality, divisions of real and non-real are not suitable as a top-level ontological categories. However, the property of being physical (able to exist in space-time) versus immaterial may be.

The complexity of the world demonstrates the challenge of classification. Perception, belief, differences in areas of study, and differences in culture all suggest different ways in which one might partition the universe. The history of human civilization may be *defined* as the history of these conceptual

divisions and their change in time. Ontology is therefore not an objective science for, like anthropology, it depends greatly on our participation, perception and influence.

The issues described thus far represent a very brief introduction to the history of the study of ontology, and a necessary precursor to the following discussion. The remainder of this chapter will explore modern principles of ontology. For further reading, an excellent summary of ontological solutions through history can be found in *Sophie's World* by Stephen Gardner, a non-fictional account of philosophy from the point of view of a young woman [6-17].

6.2. Modern Ontology

When developing systems that capture knowledge, we find that machines do not care if dragons actually exist. It is enough that the word "dragon" exists in the system. Similarly, in science we might classify minerals without regard to those we do not yet know of. The artist who conceives of imaginary worlds is content (hopefully) that those worlds do not *actually* exist, although they may just as real as numbers, people and stars. Thus the philosophical discussion of existence, while essential to understand the nature of reality, is not

essential to classification. What is more important in practice is that we can fully *express* any statement about reality.

This is not to say that philosophical ontology is superfluous. Rather, it is critical to determine categories for a personal or collective ontology: it defines belief and relationship of ones self to the world. However, once we adopt a suitable starting point the challenges of classification can proceed without concern over existence. It is perhaps not purely coincidental that machines were developed simultaneously with such pragmatic philosophies, like those of Charles Pierce:

"This paper is based upon the theory already established, that the function of conceptions is to reduce the manifold of sensuous impressions to unity, and that the validity of a conception consists in the impossibility of reducing the content of consciousness to unity without the introduction of it." [6-5].

It is instructive to compare this to a more modern definition of ontology. Gruber is often cited as a foundation for a working definition of machine-based ontologies:

"Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly." [6-6]

While Pierce attempts to define conceptualization, Gruber does not go into what a conceptualization is in any detail. Suffice to say for now that when building machine ontologies we should not forget that conceptualization,

according to Pierce, is always a reduction from some sense-perception of a larger universe. In this sense an ontology is never complete. It must be allowed to grow, merge and change as our views as a civilization change. However, once a conceptualization is made in relation to some social entity, it can be used in place of the original experience. This introduces the social aspect as well, for a *specification* will always take place in some socio-intellectual context.

To build a more generic knowledge system, therefore, we must allow multiple specifications and their subtle variants to be present in a system simultaneously. The three possible approaches suggested by Wache *et al.* include: 1) global ontologies, 2) multiple ontologies, and 3) a hybrid approach [6-7]. Global ontologies provide a single vocabulary for all users, multiple ontologies define terms by dividing the users into subgroups, and the hybrid approach includes some concepts for everyone, and some for local groups.

To illuminate this discussion further, let us examine another example in Figure 6.5. Here we find six shapes classified in three different ways: 1) by shape, 2) by border style and 3) by shading style.

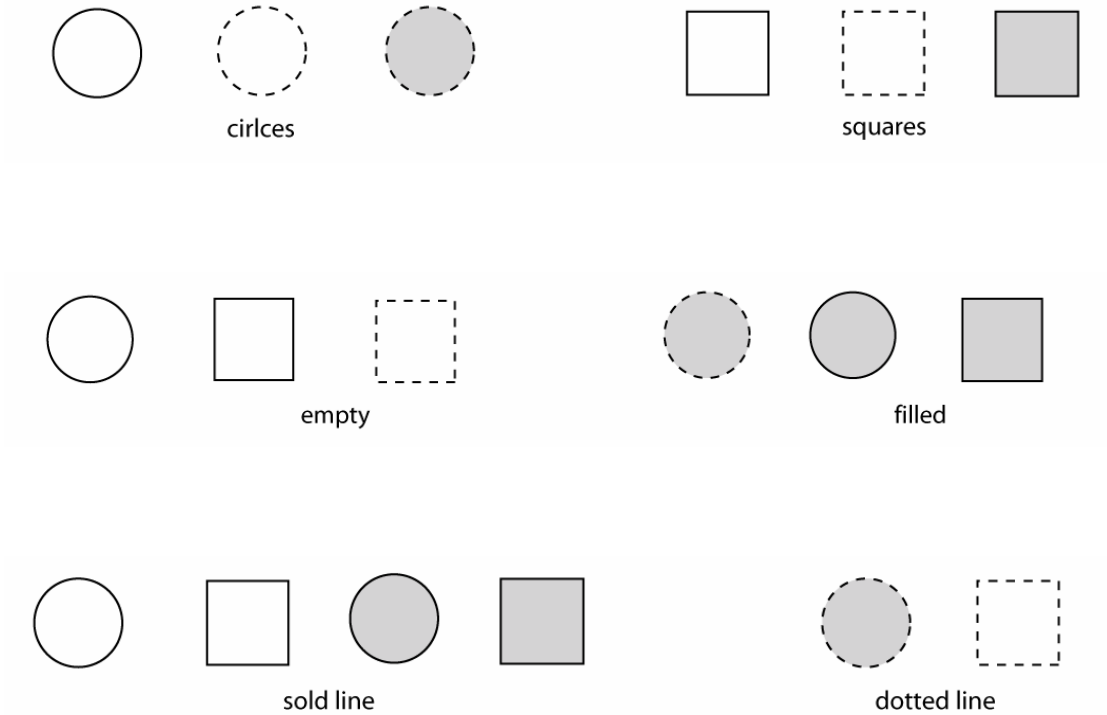


Figure 6.5. Six objects classified by a) shape, b) fill style, and c) shading style.

We observe that these classifications give rise to identifying terms such as *circles*, *squares*, *empty* and *filled*. Each set of terms defines a new taxonomy for a particular quality of the object. Thus, via its properties, a set of objects will have multiple overlapping taxonomies. Barry Smith writes:

"... it is an unrealizable ideal to suppose that ontology would consist in a single taxonomy comprehending all of reality and satisfying the rules for well-formedness we have mentioned above. The features listed are not simultaneously realizable. Above all, ontology must consist not in one tree but in a family of trees, each reflecting specific views (facets or factors) of the targeted domain – for example (microscopic, mesoscopic, macroscopic) views effected at different granularities." [6-8]

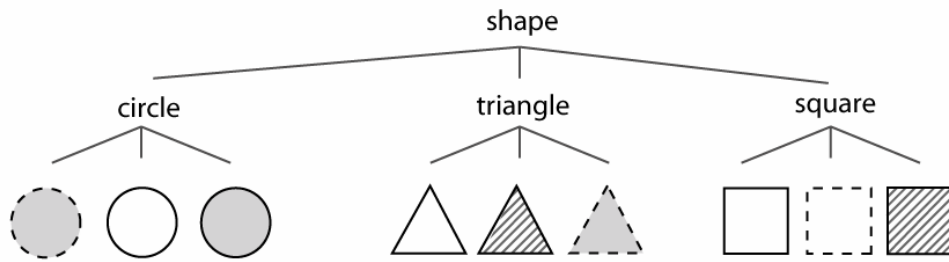
This idea of multiple taxonomies for classification is found in information management for library systems as well. S.R. Ranganathan developed a faceted system for library classification which allows written works to be classified according to multiple facets, or aspects. The vocabulary of these aspects may themselves fall into various taxonomies [6-9]. A modern database, the Prometheus system supports multiple classifications for taxonomic work in biology [6-11].

When dealing with even single objects, faceted classification is necessary. Figure 6.6 makes explicit the taxonomies present in the above example. For each object (a), a hierarchy is created for a particular quality including b) shape, c) line style, and d) fill style. A classification taxonomy for each quality provides a faceted classification of a set of objects.

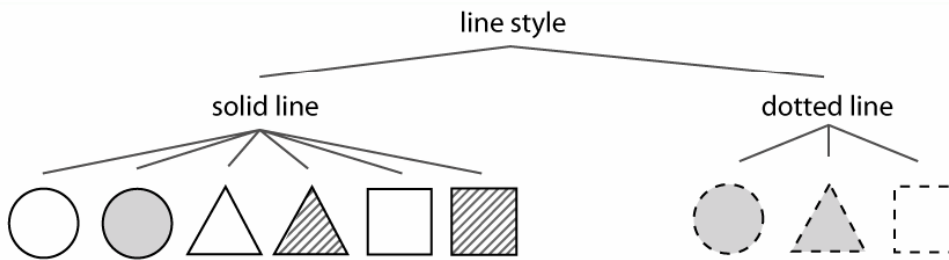
Figure 6.7 shows, however, that multiple taxonomies are possible even for a *single quality*. In the latter example, 6.7b, the word "filled" becomes an abstract term including "hatched" and "solid". There is no difference in the result, i.e. the objects themselves. Rather, the difference lies in how one chooses to define the word "filled". Thus, while the classes in Figure 6.6 can be understood as differences in the *object*, Figure 6.7. may be understood as differences in the *observer*.



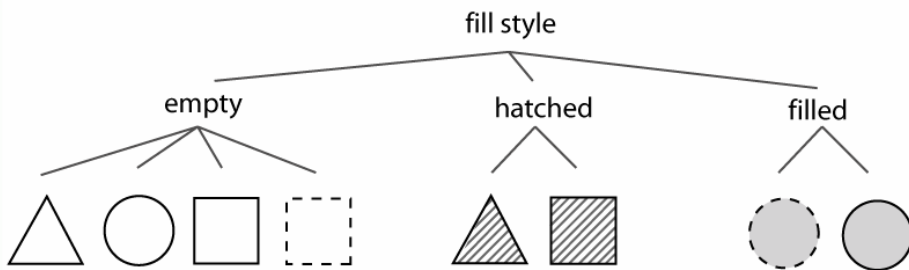
a. Objects to be classified



b. Classification by shape

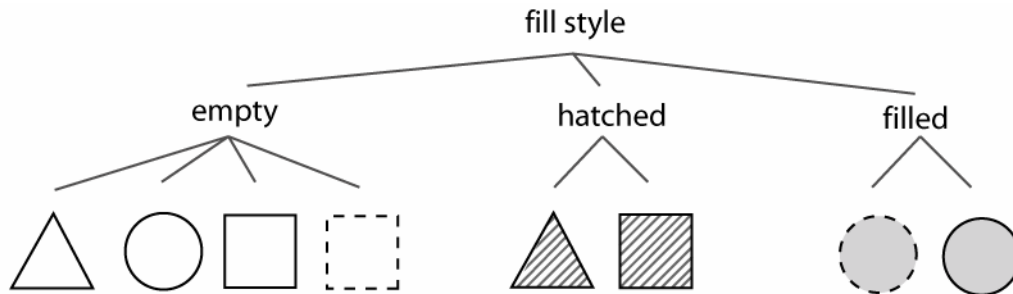


c. Classification by line style

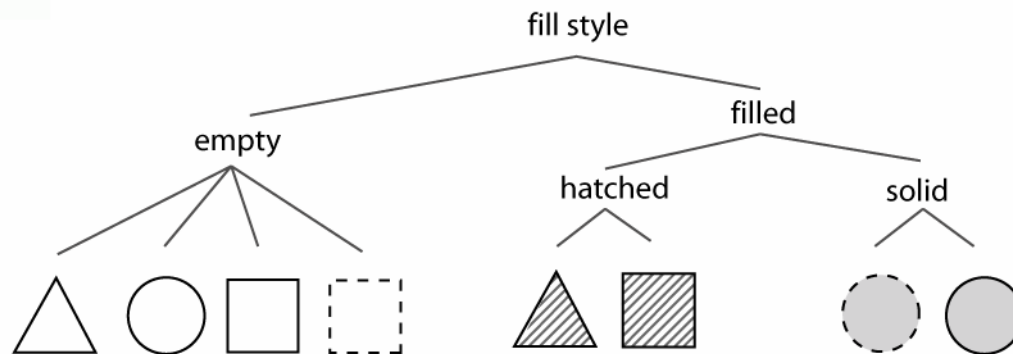


d. Classification by fill style

Figure 6.6. Explicit specification of the taxonomies introduced by the qualities of six objects.



a. Taxonomy for fill style



b. Alternate taxonomy for fill style

Figure 6.7. Multiple taxonomies are possible even for a single quality.

The taxonomies in Figure 6.7. are in fact two *ontologies* in that they provide two different ways to classify the same set of objects over the same qualities. A critical problem in generic ontology construction is communicating differences in belief among users. This can be reduced to the question: When one casually says a "filled" circle, does one mean a solidly *filled* circle (6.7a), or the class of ways in which we might *fill* it (6.7b)?

A concern with modern ontological approaches in information systems is that current approaches continue to focus on the object of the ontology, as a representation, rather than the tasks and processes necessary to resolve conflicts. The questions above may need to be resolved through proper action of a user, for structure alone cannot clarify intent or belief. Rather, we should be investigating further the functions of an ontology. The majority of ontological engineering is still focused primarily on representational challenges and not on the processes and tasks they should support [6-11].

While the tasks will involve careful user interface design, representation of multiple ontologies like those above is trivial in Quanta.. With a *layered grammatic* structure, one may provide sufficiently more expressibility than a simple tree or semantic network supports. In a single sentence (hypergraph edge), the first noun identifies the object in question, a verb selects a mode of speech, another noun gives descriptive quality, and adjectives (or classifier-nouns) select the descriptive term in a qualia-taxonomy.

Square | has | Fill Style [1] | Hatched

This may be interpreted as, "This square has the Fill Style [1] property of being Hatched." Notice this is more explicit than a casual description such as "This square is hatched." because Fill Style here refers to a single taxonomy among many. In effect it defines *whose* conception of Fill Style.

With sufficient grammatic structure, the challenge is not representational but functional. In the previous example, a useful interface would allow the user to select which conceptual view of Fill Style is used. A default ontology would provide a basis from which alternatives could be chosen. While not especially interesting in this example, this would clarify the identification of biological organisms according to individual perspective: For example:

Platypus | has | Linnean classification [1] | Ornithorhynchus

Here, the "Linnean classification [1]" is a specific ontology distinguished among the authors of other modern taxonomies for living things. A necessary part of the design of any ontological system is therefore not only the relational structures it employs but a consideration of the specific actions performed by the user to determine what set of beliefs are in operation.

6.3. Primary Classifications and Quanta

The concept of identity differs from its properties, yet both existence and quality produce taxonomies. In language it is often natural to interchange the existential and qualitative taxonomies. We say, for example, that "John is a mathematician." when in fact John is a person who has the occupation of being a mathematician.

It can be difficult to determine the existential taxonomy of a given topic. Consider devices such as the radio, telephone and microphone. These are all aural devices, so perhaps the root of the existential taxonomy is aural device. Yet a television is similar to a radio and a camera is similar to a microphone. Thus aurality cannot be an existential quality of a device.

Perhaps microphone and camera should be existentially classed together as *input devices* as opposed to *transmission devices*. Yet some cameras are also a recording devices while a microphone is not. In the end, it is often better to reduce any existential taxonomy in a domain and call all of them simply *devices*. In this way, any number of other classifications may be used to qualify particular instances according to their properties.

The Quanta Generic Ontology (GO) is the ontology designed for this thesis. It is a minimal top-level ontology designed to assume as little as possible by having abstract classes with a large number of entries, i.e. shallow bushy trees, rather than larger deep trees. For example, in this ontology all devices are specified as "X is a device". Secondary taxonomies are used to specify what particular type of device it is, since no single classification will do. The Generic Ontology provides a taxonomic reference for basic users of the Quanta concept network.

6.4. Dependent Classification

Due to the complexity of biology, John Ray held that no characteristics were more important than any others [6-12]. The current structure of the Linnean taxonomy is based on evolutionary phylogeny, which is linked to natural selection as it has occurred over millenia.

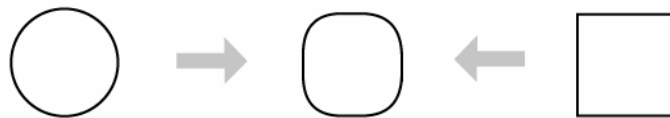


Figure 6.8. A circle and square combine to form an ellipsoid.

Things give rise to others which are a combination of the first. In Figure 6.8, a circle and square combine to create an *ellipsoid*. In modern ontological discourse we can observe that both John Ray and Linneaus were correct. There can be no universal set of classifiers, but there are some more useful ones - such as functional or structural *dependence*.

The concept of dependence can be generalized to many different disciplines. Ontological dependence exists when one thing must precede the first, as in the property of a thing being "empty" or "filled" depending on their being a thing in the first place. Physical dependence occurs when something arises

after something else. The idea is also found as a principle in other cultures. Usually applied to personal practice, the Salingtamba Sutra in Mahayana Buddhism says that:

"Because this is present, that will arise, and because that was born this is being born." [6-13].

The classification of man-made entities, for example, is particularly suited to dependent analysis. In Table 6.3, we see the main subclasses of man-made entities as expressed in the Quanta Generic Ontology (GO). In this system the earliest example of each is used to determine the order of entries. Regardless of the fact that many modern tools were created after some written works, the *earliest* tool predates the *earliest* written work and therefore the category of Tools precede Written Works in this construction. This principle is applied throughout the ontology, and to subcategories of tools and works so that Stone Tools precede Power Tools.

Table 6.3. Quanta base ontology for man-made entities. Date of earliest example is used to order categories. Source: Wikipedia (various articles)

Quanta Category	Earliest Example	Date	Period
Language	Utterances, speech	2,000,000 - 40,000 BC (?)	Paleolithic - Homo habilis
Tools	Knives, axes, scrapers	2,000,000 BC	Paleolithic - Homo habilis
Civil Objects	Basic shelters	1,500,000 BC	Paleolithic - Homo habilis
Organizational Units	Villages	40,000 - 20,000 BC	Mesolithic
Materials	Stone (as writing material)	30,000 BC	Mesolithic
Visual Works	Cave painting	30,000 BC	Mesolithic (Grotte Chauvet)
Semiotic Units	Symbols, glyphs	7000 BC	Neolithic (Vinca Script)
Written Works	Sumerian tablets	4000 BC	Early Bronze Age
Dance	Rituals, performance	3300 BC (First depiction)	

Admittedly the distinctions are somewhat subjective especially considering that these earliest dates are not known exactly. There is no single ordering that would be perfect but some orderings, such as dependent arisal, are more *natural* than others. Like the species taxonomy, this base ontology for man-made objects is one among many that may be simultaneously present in Quanta.

While lexical systems such as WordNet favor the elimination of hierachies in concept organization, additional hierarchical ontologies can be a useful supplement to provide a ground for understanding by individual users [6-23]

6.5. Multiple Inheritance

There is another way to look at the ellipsoid in the previous example. Rather than introduce a new term, we might say that the new shape is both a circle and a square. In this way, we would say that the ellipsoid inherits the properties of both shapes.

Multiple inheritance is not the same as a substance-quality relationship. That is, to say that "a filled circle is both a circle and is filled" is not the same as saying that "an ellipsoid is both a circle and a square". The nature of the ellipsoid depends on both equally, while the filled circle may just as well have been empty. In addition, a new object often introduces properties that are present in neither of its parents.

A shape may have a fill style (empty or solid), circles a radius and squares the length of its sides. A simple union of these, the ellipsoid, has the same properties: style, radius, and side length. But a nubbed-ellipsoid, Figure 6.9, may also include the size or number of nubs as a property - something not present in shapes, circles or squares. Thus we might attach new properties to the inherited concept.

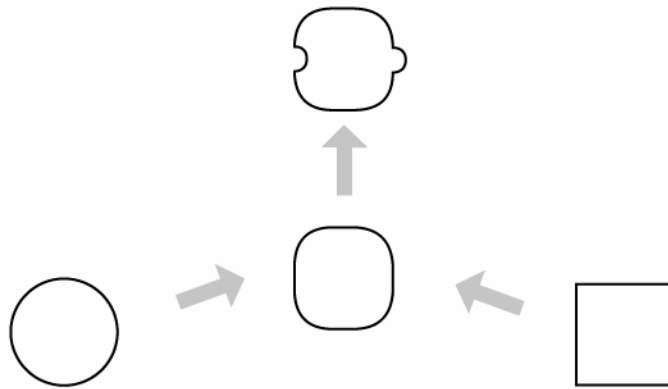


Figure 6.9. New properties can arise in inherited classes.

Multiple inheritance is useful when we have something that should be classified both ways. A dragon is both a mythological creature and an organism (presumably it has bones, tissues, muscles and cells like any other). A spork is both a spoon and a fork.

Multiple inheritance can also lead to problems, for example when properties of an object's parents conflict. A classic example is Richard Nixon, who was both a Quaker and a Republican. While Quakers are considered peaceful, Republicans generally favor war. Thus to be both is a contradiction. However, this can usually be corrected by being more explicit. Nixon was a Quaker early in life, but a Republican later. To be a Quaker or Republican may not imply that its members are *always* peaceful or favor war. Finally, while morally questionable it may not be a contradiction at all to live peacefully (oneself) but to tolerate or advocate war (in distant lands).

Multiple inheritance is present in ontology engineering and programming but not always used in the latter case [6-14]. Multiple inheritance is supported in Quanta simply by providing multiple existential statements that describe each object. Two statements may both define a base class, as in the following example which lists spork as both a spoon and a fork:

```
spork | is a | fork
```

```
spork | is a | spoon
```

The hypergraph structure of Quanta makes it trivial to represent multiple inheritance. Unlike programming languages in which multiple inheritance is an integral part of the language, providing a database structure that naturally expresses such relationships simplifies implementation. The functional extensions to enable operations on multiply inherited classes can be introduced as needed at a later time. Some inference procedures and visualizations in the Quanta interface are currently designed to support multiple inheritance, while others may be extended in this way. In both cases, the implementation of the semantic database does not change to support this feature.

6.6. Continuous Properties

Another common problem in classification appears when we attempt to organize continuous properties. Consider Figure 6.10.

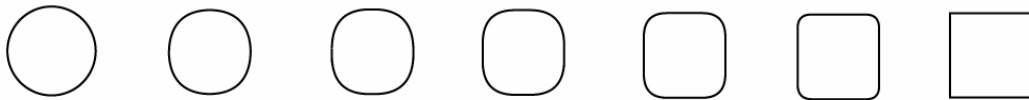


Figure 6.10. A circle becomes a square. How many classes of objects are there?

How many categories are there? When a property becomes continuous, our distinctions of class are arbitrary. Color is a good example. What specific wavelengths (or RGB values) constitute the color "bright red"? Divisions across continuous properties are always subjective and thus require that we clearly specify the artificial boundaries of our schema. Hayes, mentioned by Sowa [6-15], shows that fluids have many continuous properties - making the many forms of fluids particularly difficult to classify [6-16].

As mentioned earlier, Quanta allows users to tag statements based on their source. In this way we can identify the particular classification scheme being used for any classification.

6.7. Formal Ontology & Logic

One of the most significant challenges to philosophical ontology is the *problem of universals*. As we delve deeper into the nature of reality, or to the top of an ontology, it becomes more difficult to say what exists. Plato believed that ideal Forms were universal. Aristotle believed substances, qualities, quantities, space and time were universal. Democritus believed that atoms were universal. Plotinus believed that only one universal was needed, the One. Various religions hold that God or deities are universal [6-17].

Around 560 BC, both western and far eastern cultures introduced a new possibility: that our own impermanence prevents us from grasping universals. [6-18] [6-19]. Further interpretations of this are that only *change* is universal (a process), the only universal is language, or that there are no objective universals.

Nominalism, the idea that language must be the root of existence, was greatly advanced by modern philosophers such as Frege and Wittgenstein. In 1918, Wittgenstein developed *Tractatus*, a work in which existential philosophy was expressed in the logic of mathematics [6-20]. It had a profound influence in the development of modern ontology. Specifically, it suggested that *logic*

should be the operating mode of ontology as what exists can only be expressed through a clear and specific analysis of what we say.²

While predicate logic was first initiated by Frege, first-order logic was more completely defined by David Hilbert and Wilhelm Ackerman in 1928 [6-22]. To give a very brief overview as it relates to semantics, first-order logic allows the use of *universal* and *existential* quantifiers to make existential statements:

In some places, for many people, survival is difficult.

Second-order logic allows for statements *about* other statements, such as:

Jane believes that in some places, for many people, survival is difficult.

First and second order logic will not be described here except in as much as they were necessary to the design of Quanta. Predicate logic is often used in knowledge representation and artificial intelligence. One example is Cyc, initiated by Douglas Lenat [6-22]. This is a project to collect common sense knowledge, and has been successfully applied to understanding certain phrases in natural language. Since Quanta is a knowledge database, rather

² As an author's note, I believe that formal logic is necessary any time we wish to express existence *visa vi* language, or in some system. Thus it is necessary to building ontological frameworks. However, I do not hold that existence is defined by logic. Our first mode of knowledge is experience, not expression. Thus the problem of universals is not solved by logic. Language is universal to *our* existence, but as to the universal nature of reality - it remains a mystery.

than an experiment in artificial intelligence, there are instances in which certain constructs in first and second order logic are unnecessary. The current ontology is an incomplete second-order logic, but the layered grammatic structure of the underlying system allows for future expansion in this direction.

An example of where first-order logic is unnecessary is in the use of logical disjunctives (OR):

Mary likes apples or Jane likes oranges

Expressions like these are important to language understanding and artificial intelligence, but unnecessary in database design. This is because the statement above represents incomplete knowledge. We do not know for certain which person likes which fruit, but when we do the other is predetermined. In Quanta, the above statements are represented as:

Mary | may | like | apples

Jane | may | like | oranges

Notice that the disjunctive relationship is gone. When the user knows that Mary likes apples, they must explicitly state that Jane does not:

Mary | likes | apples

Jane | not | like | oranges

One future goal of Quanta is to implement full second-order logic by introducing *rules* as grammatic structure. This would not affect the underlying hypergraph of the system and would allow certain things to be done automatically. As Quanta is a database, however, this is only needed to augment knowledge generation.

The relationship of Quanta to predicate logic is best explained by examining the formal definitions of objects in first and second-order logic and how they are applied to Quanta. *Appendix A* gives a summary of Quanta from the perspective of predicate logic. It provides an analysis of some of the other areas where first and second-order logic principles were not used.

6.8. Disambiguation

We have seen how Quanta uses modal logic to allow for statements such as "Mary | may | eat | apples". Another important type of ambiguity arises from multiple definitions of the same word as in 1) painting, the study of the technique and 2) painting, the particular object that is created. Word ambiguity is the reason that language is statistical. If all words had only one meaning, it is immediately obvious how to parse a sentence. For example:

Cats | fear | water

Ambiguous words are very common. Even when there are few variant definitions there can be ambiguity. The meaning of the above example is fairly unambiguous. Yet "cat" may be an abbreviation for catamaran. Fear can mean to be physically afraid, or it can mean to revere as in "to fear the wrath of God".

Quanta resolves ambiguity explicitly by using brackets [] to disambiguate individual words. Any time there is an ambiguity, this is separated into distinct uses:

- cat [1] A feline animal with four legs and fur
- cat [2] An abbreviation for catamaran, a boat with two hulls
- cat [3] An abbreviation for catfish, a fish with whiskers

This introduces an issue of user interface design. How do we encourage the user to select the correct definition among those available? Careful interface design should provide feedback as to the meaning that will be selected.

The syntax of brackets is beneficial from a performance perspective. The bracket is included directly with the hypergraph node for that noun. It is not a separate node. That way, if the user searches for the word "cat", the system is able to simply search for all nodes that contain the word "cat" while ignoring

any brackets. This produces a disambiguation list from which the user can select a specific definition.

6.9. Quanta: Ontology Design

The Quanta semantic database is a network of connected concepts. The Generic Ontology (GO) is a shallow hierarchy of top-level categories that provides structure and organization for these concepts. In a lexicon, such as WordNet, the use of an existential hierarchy is eliminated to allow full expressibility among terms [6-23]. In others, such as the Suggested Upper Merged Ontology, a detailed analysis is undertaken to carefully decide the top-level categories and divisions [6-24]. The GO ontology presented here takes a hybrid approach by providing a shallow, but fully connected upper ontology and a set of arbitrarily deep classification taxa.

The GO ontology is divided into an existential ontology and a set of classification taxonomies. The classifications are themselves a subset of the first but remain conceptually distinct, as shown in Figure 6.11. Classifications may include, for example, subject areas, styles of art, occupations and types of software. The highest depth of a classification in the current data set is eighteen for the Linnean taxonomy. However, the maximum depth of the existential taxonomy is only five (not including classifications).

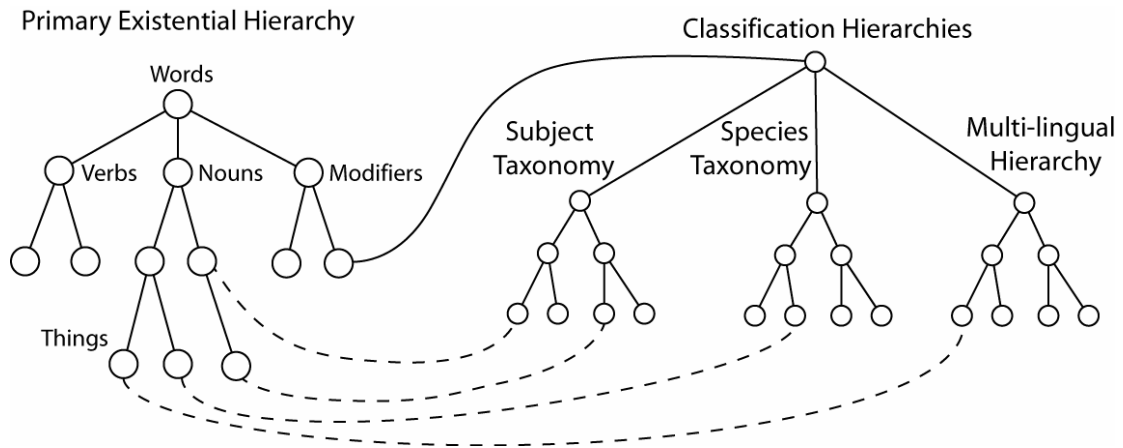


Figure 6.11. Existential and classification hierarchies in the Quanta Generic Ontology (GO).

There is no distinction between a class and an instance, except that the latter have no children in the existential hierarchy. For example, all specific living things are listed as immediate children of the concept "organism". Organisms are listed as a natural entity (Noun -> Entity -> Natural entity -> Organism). The species taxonomy provides a *classification* for each organism in the same way that an individual, such as John, is a "person" first and additionally classified with the occupation of being a mathematician.

To summarize thus far, many factors were used in developing the Quanta ontology. Considerations of philosophy, logic, programming, and language have resulted in the design principles in Table 6.4.

Table 6.4. Quanta: Ontology Design Principles

Objects & Properties	Things versus the quality of those things
Multiple Classification	Trees for the properties of objects
Multiple Taxonomies	Trees for the interpretations of a single property
Multiple Ontologies	Trees for different existential belief systems
Classes & Instances	Concepts not tied to any functional limitations
Word Disambiguation	Use of a special syntax to disambiguate
Modal Logic	Used to express probabilistic statements
First-Order Logic	Used to express most relationships implicitly
Second-Order Logic	Used to express views of particular people (users)

In addition to these constructs, the Quanta Ontology is developed as a *layered ontology*. Ontological principles resulting from the above ideas are divided into coherent layers which allow for a simpler implementation.

The various layers of the Quanta Ontology are shown in Table 6.5. A formal description of these layers is presented in *Appendix B*. Programmatically, there is a corresponding functional level of Quanta for every ontological level described above. The lowest level is the implementation of a semantic

hypergraph (L1). The second layer introduces deductive programs that can process language on this hypergraph (L2). The third provides tools to extract, navigate and organize trees (L3).

Table 6.5. Quanta: Ontological Layers

Ontology Layer	Definition
<i>Layer 1: Formalized Ontology</i>	This layer represents ideas at the most basic level of a hypergraph. Nodes and relationships are introduced.
<i>Layer 2: Linguistic Ontology</i>	Language is introduced, including parts of speech such as nouns, verbs, propositions, adverbs and adjectives.
<i>Layer 3: Structural Ontology</i>	Organization structures are introduced, including classes, instances, classification trees, taxonomies, ontologies
<i>Layer 4: Generic Ontology</i>	Expresses general content including the top-levels of the ontology, middle-levels, and specific objects.

Layers one through three are the ontological principles used to develop Quanta. Layer four, the Quanta Generic Ontology, represents the base content of the database. An excerpt of the top-level ontology is presented in *Appendix C*. Additional data sets from a variety of disciplines, built using this ontology, are described in chapter nine (Quanta Prototype and Future Directions).