Three Kinds of Meanings

Patrick Suppes

Abstract: It is argued that there are at least three distinct kinds of meaning that have wide currency across many different kinds of language use. The first kind consists of formal definitions of terms in mathematics and science. These definitions are usually clearly distinguished, as such, in the discourse context in which they occur. The second kind consists of dictionary definitions, familiar to all of us. The third kind, that of associative meanings, is not as widely recognized as the first two, but associative meanings are at the center of our cognitive and emotional experience. Baldly stated, the thesis defended is that associations provide the computational method of computing meaning as we speak, listen, read or write about our thoughts and feelings. This claim is supported by a variety of research in psychology and neuroscience. For much of the use of this third kind of meaning, the familiar analytic-synthetic philosophical distinction is artificial and awkward.

1. Meaning given by formal definition

I first consider definitions formalized within a theory in the ordinary mathematical sense. This means the language itself is not described formally but that the primitive concepts of the theory are given and any use of mathematics other than that of intuitive logical inference is explicitly stated. For simplicity here, I shall mainly refer only to theories that in principle could be formalized within first-order logic, but that degree of formalization is not itself considered. This means that the only formal terms in the theory to begin with are the primitive symbols of the theory—no other mathematical notation being used. Within this framework, we then say this about the formal definitions within a given theory.

The first definition of the theory is a sentence of a certain form which establishes the meaning of a new term of the theory using only the primitive terms of the theory. The second definition in a theory is a sentence of a certain form which establishes the meaning of a second new term of the theory by using only the primitive terms and the first defined term. And similarly for subsequent definitions. The point to be noted is that the definitions in a theory are introduced one at a time in some fixed sequence. Because of this fixed sequence we may always speak meaningfully of *preceding* definitions. Sometimes it is convenient to adopt the viewpoint that any defined term must be defined by using only the primitive terms. In this case there is no need to introduce definitions in some fixed sequence. However, the common mathematical practice is to use previously defined terms in defining new terms; and to give an exact account of this practice, a fixed sequence of definitions is needed. So, in summary, the meaning of new terms in a mathematical theory is entirely derived from the primitive terms. And the formal meaning of these primitive terms is determined by the axioms of the theory. For example, in affine geometry, typical axioms for the termary relation B of betweeness assert that B has the following symmetry property for any three points a, b, and c:

If $B(a, b, c)$ then $B(c, b, a)$.

From the standpoint of the logic of inference a definition in a theory is simply regarded as a new axiom or premise. But it is not intended that a definition shall strengthen the theory in any substantive way. The point of introducing a new term is to facilitate deductive investigation of the theory, but not to add new content to it. Two criteria which make more specific these intuitive ideas about the character of definitions are that (i) a defined term should always be eliminable from any formula of the theory, and (ii) a new definition does not permit the proof of relationships among the old terms which were previously unprovable; that is, it does not function as a creative axiom. I shall not be concerned here to elaborate on the criteria of eliminability and noncreativity except that the question of creativity will be discussed a little later.

The important philosophical point about meaning within such a framework is that once an extensive theory like axiomatic set theory is given, then almost all, even if not all, of the ordinary concepts of mathematics can be defined in this formal way within set theory. The problem of formal meaning has thereby been reduced to the meaning of the primitive terms, which in the case of set theory is essentially reduced to the meaning of set membership as the only essential primitive (Suppes, 1960/1972).

There can certainly be objections to what I am saying about the character of the formal theory of meaning given by definitions in a formal theory. But there is no question that there is also something satisfactory and permanent about the kind of result that was obtained over the twentieth century in showing how almost all ordinary mathematical concepts could be defined within set theory, with little, if anything, left out. Surely, a surprising and historically important outcome.

Still in ordinary mathematical practice the standards of definition just characterized are a little bit too strong. Many ordinary definitions in mathematics, and even more in the ordinary use of a language, are conditional in form. The definition is really only aimed at situations in which the conditions formulated explicitly in the antecedent of the conditional definition are satisfied. A good example is the conditional definition of division in the theory of real numbers. Such a definition is used in order to avoid division by zero.

A natural definition using just the usual notation for real numbers without introducing any explicit predicate for real numbers is the following:

If $y \neq 0$ then $x/y = z$ if and only if $x = y \cdot z$.

The objection to this definition is that we cannot decide on the truth or falsity of such a simple assertion as:

$$\frac{1}{0} = \frac{2}{0}.$$

By introducing the unary predicate $R(x)$ meaning x *is a real number*, we can then give a more satisfactory conditional definition:

If $R(x)$ & $R(y)$ & $R(z)$ & $y \neq 0$, then $x/y = z$ if and only if $x = y \cdot z$.

Given this definition, we cannot prove:

$\frac{x}{0}$ is a real number,

and we cannot prove:

$\frac{x}{0}$ is not a real number,

but we are not faced with the counterintuitive situation of being forced to make $x/0$ a real number.

The intricacies of these matters we will not explore further. The point is to make it clear that even in ordinary mathematical practice the casual use of conditional definitions can lead to some problems related to those of meaning often discussed in philosophy. I have in mind, for example, such questions as what really are sets or natural numbers.

One approach to this kind of question is to use creative definitions of identity, which are forms of abstraction that, in the simple cases, eliminate all traces of other entities being referred to in the definitions. A good example is the definition of ordered pairs, famous in the literature, going back to the early results of Wiener (1914) and Kuratowski (1921). Instead of using the Kuratowski formulation, which is standard,

DEFINITION. $(x, y) = \{\{x\},\{x, y\}\}$,

we can use the creative definition of identity:

CREATIVE DEFINITION. $(x, y) = (u, v)$ *if and only if* $x = u$ & $y = v$.

I call this last definition *creative* because under rather weak assumptions we can show that something new can be proved using the definition. More exactly, the *criterion* for a definition not to have this property is formulated as follows.

CRITERION OF NON-CREATIVITY. A formula S introducing a new term of a theory satisfies the *criterion of non-creativity* if and only if: there is no formula T in which the new term does not occur such that *If S then T* is derivable from the axioms and preceding definitions of the theory but T is not so derivable.

Here is a still simpler example of a definition that we can show is creative. Let the context be just the theory having as axioms those for a weak ordering relation $\succcurlyeq$.

Axiom 1. *If* $x \succcurlyeq y$ *and* $y \succcurlyeq z$ *then* $x \succcurlyeq z$
Axiom 2. *Either* $x \succcurlyeq y$ *or* $y \succcurlyeq x$.

And we have as the formula S introducing the new term, the conjunction of (1) and (2):

(1) $\{x\} = \{y\}$ if and only if $x = y$.
(2) $\{x\} \neq x$.

From (2) we infer at once; a formula corresponding to T in the Criterion:

(3) $(\exists y)(y \neq x)$,

which cannot be derived from Axioms 1 and 2. Notice that for complete explicitness, S includes not just the creative identity of (1), but (2), settling an ambiguous matter of identity.

These rather technical points are not of central concern in what I am discussing here, so I will move on. It is important, however, to note that there is much to be said for cutting off the speculation about meaning that we find in various places, even for ordered pairs, by insisting always on just the criterion of identity as a definition. Then, to the question, "What are ordered pairs? ", we can reply that there is no appropriate answer except to say that they are abstract entities with just the defining property stated in the definiens. It is important that we can have, within its limitations, a satisfactory and definite theory of meaning in standard mathematics and in many mathematical developments in science. I emphasize again, of course, that this theory, which I call the formal theory of meaning, does not help in speculation about how mathematicians think up new theorems and new concepts. The psychological apparatus, or ultimately the neural, apparatus required for that purpose is not considered at all in the formal theory. (For an elementary exposition more detailed than what is given here, see Suppes (1957/1999, Ch. 8.)

2. Dictionary meanings

If you ask the man in the street, so to speak, "What is the meaning of a word?", he will likely say, "Why don't you look it up in the dictionary? " In other words, in ordinary talk, meanings are to be found in dictionaries. This is not a bad idea. In fact, all of us, almost without exception, appeal to the dictionary for a sense of meaning at some time or other. Yet from a philosophical standpoint, dictionary meanings are not easy to handle. For example, I discuss later the noun *capital* in the sense of the capital of a country, province, or state. If you look up *capital* in the OED (Oxford English Dictionary) you will find a bewildering variety of meanings, which clearly indicate the polysemous character of this widely used word form, as the linguists would say. It does not take much reflection to realize why mathematicians quickly fled this jungle of meanings to a crisp and clear ice palace of exactly one meaning for each word.

If we think of a particular OED meaning for a word like *capital*, we need to find the context of use. For such a word it is simply fantasy to talk about *the* meaning. Indeed, the most important contrast between the formal definitions of mathematics and those of a dictionary is the move in the formal case of mathematics to a unique meaning in all standard contexts. Exactly the opposite is true for a dictionary, and especially one as extensive as the OED. We expect to find, and do find, for almost any word we look up a variety of meanings. This pluralism is very much in the spirit of the context-dependent approach to language, so characteristic of much recent work.

But in my view both kinds of results are needed, i.e., the univocal meanings of mathematics and the polysemous ones of almost all ordinary use. The world would be too rigid a place, intellectually, if all definitions had to be as univocal as standard mathematical ones. On the other hand, mathematics and science would suffer from anything like the decidedly polysemous character of many ordinary words, if they had no other linguistic possibility. So, as for many things systematic and otherwise, it is desirable to have a plurality of approaches. In this spirit, I now turn to psychological and neural senses of meaning.

3. Meanings as associations

To illustrate the point of why I now turn to psychological and neural senses of meaning, let me quote first the excellent dictionary definition of *chair* in the OED.

> **1. a.** A seat for one person (always implying more or less of comfort and ease); now the common name for the movable four-legged seat with a rest for the back, which constitutes, in many forms of rudeness or elegance, an ordinary article of household furniture, and is also used in gardens or wherever it is usual to sit. *to take a chair*: to take a seat, be seated. (OED, Vol. II, p. 248)

This is about as systematic and complete a definition one could ever expect to give. So now consider someone asking me the question, "Where is that red chair that used to be in your study?" My answer: "It got old and worn, and now is in another part of the house." What is important about this example is what happens to me when I hear the question. I immediately associate the phrase "that red chair in my study" to a vivid mental/neural image of that chair. It comes up within 500 milliseconds I would say, and I can imagine the shape and the color, as well as have a sense of the texture of the cloth. I do not associate to anything like the OED definition of *chair*. I have quick access by association to the image even though the image of that red chair in my brain has not been recently retrieved. That chair was visible to me many days over many years and it is not easy to forget. If a further question comes, "What kind of material covered the chair?" I quickly answer, not by exploring any OED meanings of words, but by mentally or neurally scanning the image of the chair, I reply, "The cover was a heavy rough velvet."

Now there are both philosophers and psychologists who really do not believe in such image theories of mental representations or mental computations. I think their views are wrong, but will not try to set out an argument in detail here. My own personal experience, and a great wealth of experiments that can be cited from many different kinds of attacks on this sort of question, persuade me that images are real enough. Even more to the point, I have recently been conducting with others, experiments observing the neural activity of the brain. In some cases we can make reasonable claims for isomorphism of images generated by stimuli and images generated by imagination, as in the case of the red chair. Theses about structural isomorphism between mental or neural images and objects in the world will not be examined here, because the subject is complicated and still controversial. I could not say much that would be useful in a short space (For details, see Suppes et al., 2009). I will just stipulate that I think in terms of such images and very much believe they are the "meaning" we often properly attach to a given word or phrase. The point is that if I am asked a question and I need to compute an answer, I often do so, not by using some semantic computations that depend upon verbal associations, but rather, by associative links to mental images of a visual or auditory kind, nonverbal in character. Now, of course, other kinds of questions will depend upon associations to other words, and I will give some examples in a moment. But what is fundamental about meanings as associations is that for much computation required for answering a great variety of questions about what is going on in the world, what is in the world, what acts I have committed, etc. it is nonverbal images that I use; visual images, auditory images, haptic images, and sometimes tactile images that are associated quite directly to the words or phrases used. Such mental "models", so to speak, play an important role in thought, rather neglected by many philosophers such as Quine, but what I have written about dictionaries here is rather close to what he has to say about lexicography (1992, pp. 56–59). Hume, unlike Quine, has much to say about imagination and the role of images in thought. Here is a brief quote to that effect:

> . . . The imagination has the command over all its ideas, and can join, and mix, and vary them in all the ways possible. It may conceive objects with all the circumstance of place and time. It may set them, in a manner, before our eyes in their true colours, just as they might have existed. (Hume, *Treatise*, p. 629)

Associative networks for computing truth. Before discussing meaning in the context of associative networks, it may be helpful to show, in brief outline, how associative networks can be used as a model of how the brain computes the truth of ordinary sentences—sentences whose truth or falsity should be easily available to a large number of people. The network used for computing such truth is, I believe, a good example of a small-world network, a topic about which I will say more later. It has the virtue of providing something that is meant to be very operational and will illustrate how meanings as associations can be used.

The basic idea is that the computations are made by an associative network with brain representations of words being the nodes of the network and the links between the nodes being the associations. More generally, auditory, visual, and other kinds of brain images can also be nodes. There is a reasonable body of evidence to support the hypothesis that the nodes of the network are collections of synchronized neurons.

In the initial state, not all nodes are linked, and there are, in this simple formulation, just two states, *quiescent* and *active*. No learning or forgetting is considered. It is assumed that, after a given sentence is responded to as being either true or false, all the activated states return to quiescent. The axioms, which are not stated here, are formulated just for the evaluation of a single sentence (for details, see Suppes & Béziau, 2004). The way to think about the networks introduced is that a person is asked to say whether sentences about familiar phenomena are true or false.

The sentence input comes from outside the associative network in the brain. I will consider only spoken words forming a sentence, although what is said also applies to visual presentation, as well. So, as the sentence is spoken, the sound-pressure image of each word that comes to the ear is drastically transformed by a sequence of auditory computations leading to the auditory nerve fibers, which send electrical signals to the cortex. In previous work, I have been much concerned with seeing if we can identify such brain signals as brain representations of words. Some references are Suppes, Lu, and Han (1997) and Suppes, Han, Epelboim, and Lu (1999a, 1999b).

The brain activates quiescent states by using the signal brought into the cortex as the brain representation of the verbal stimulus input. With the activation of the brain representation of words by external stimuli, the associations between activated brain representations are also activated by using this same signal.

Moreover, it is assumed in the theory that activation can be passed along from one associated node to another by a phenomenon characterized some decades ago in psychological research as *spreading activation.* For example, in a sentence about a city like Rome or Paris, some familiar properties are closely associated with these cities and the brain representation of these properties may well be activated shortly after the activation of the brain representations of these words, even though the names of these properties, or verbal descriptions of them, did not occur in any current utterance. This is what goes under the heading of *spreading activation.* Some form of it is essential to activate the nodes and links needed in judging truth, for, often, we must depend upon a search for properties, which means, in terms of processing, a search for brain representations of properties, to settle a question of truth or falsity. A good instance of this, to be seen in the one example considered here, is the one-one property, characteristic of being a capital: *x is capital of y*, where *x* is ordinarily a city and *y* a country. There are some exceptions to this being one-one, but they are quite rare and, in ordinary discourse, the one-one property is automatically assumed.

One other notion, introduced in the axioms of Suppes and Béziau (2004) for computing truth, is the concept of the *associative core* of a sentence, in our notation, $c(S)$ of a sentence S. For example, in the kinds of geography sentences given in the experiments referenced above, where similar syntactic forms are given, persons apparently quickly learn to focus mainly on the key reference words. So, for example, the associative core of the sentence *Berlin is the capital of Germany* is a string of brain representations of the three words *Berlin*, *capital* and *Germany*, for which I use the notation BERLIN/CAPITAL/GERMANY, with, obviously, the capitalized words being used to denote the brain representations. A more complicated concept is needed for more general use.

In the initial state of the network, associations are all quiescent, e.g., PARIS ~ CAPITAL, and, after activation, we use the notation PARIS ≈ CAPITAL. In the example itself, we show only the activated associations and the activated nodes of the network, which are brain representations of words, visual or auditory images, and so forth. The steps of the associative computation are numbered in temporal steps t1, etc., which are meant to include some parallel processing.

Example. Berlin is the capital of France.

t_1. BERLIN, CAPITAL, FRANCE	Activation
t_2. PARIS, 1–1 Property	Spreading activation
t_3. BERLIN ≈ CAPITAL, CAPITAL ≈ 1–1 Property CAPITAL ≈ FRANCE, PARIS ≈ CAPITAL PARIS ≈ FRANCE	Activation
t_4. GERMANY	Spreading activation
t_5. PARIS/CAPITAL/FRANCE BERLIN/CAPITAL/GERMANY	Activation
t_6. TRUE ≈ PARIS/CAPITAL/FRANCE TRUE ≈ BERLIN/CAPITAL/GERMANY	Spreading activation
t_7. FALSE ≈ BERLIN/CAPITAL/FRANCE	Spreading activation

This sketch of an example, without stating the axioms and providing other technical details, is meant only to provide a limited intuitive sense of how a plausible associative theory can be developed for computing the truth of simple empirical sentences.

Notice, of course, that in this discussion of truth no direct mention was made of meaning, but meanings are implicit. I want now to turn to such considerations. Let us begin with the OED and the "meaning" it assigns to the proper name *Paris*:

> *Paris*, the name of the capital of France, used in various collocations: e.g. in names of materials or articles made in Paris as *Paris crisp* . . . (OED, Vol. VII, p. 478)

Then a large number of this latter use of *Paris* is given, for instance, "*Paris candle*, a kind of large wax candle" and then additionally, as is characteristic of the OED, many examples of use. An amusing one is from 1599, Shakespeare's *Henry the Fifth*, "to that end I did present him with the Paris-Balls." It is also easy to accept the dictionary "meaning" of Paris as the natural answer to the question, after all, "What is Paris? " to which the response is given, "the capital of France."

In the associative network context, meanings are of course given by strong associations. It is also important to point out that my earlier red-chair example shows that it is natural to think of meanings being given by visual or auditory images. In previous work (Suppes, 1980) I have emphasized the private nature of the detailed procedures or associations connected even to proper names. So, for example, although it is not ordinarily given a procedural or associational characterization, in my own mind (or brain), with Paris I associate a number of procedures including scanning of visual images and sometimes images of other modality in connection with questions asked about Paris. An even better example is the procedures associated with the name of a particular person. For someone that I know well enough to recognize, procedures for recognizing him are associated with his name. It is not customary to think of such recognition procedures as part of the meaning of the proper name, but it seems to me that just as set-theoretical definitions of terms do not work well for ordinary language, so the ordinary use of language also requires a more extended notion of meaning than that provided by dictionaries. The important point is to insist here on this wider conception of meaning once we move to much more detailed questions, such as the one I just considered in the computation of truth. Such computations are frequent, and often do not involve sentences, but rather, situations of action, which require continuous associations and with these, comes meaning.

It is to me obvious that cognitively we mainly use neither of the first two kinds of meaning, i.e., formal definitions within a theory or dictionary meaning. We, and other animals, depend upon rapid associations of one brain image to another to cognitively evaluate, and then take action, in situations of every possible kind.

Only very small networks were introduced in my discussion of computing the truth of a sentence by the way we do familiar ones, that is, by association. But for each of us, such small networks are embedded in the much larger associative network representing much of our past experience. We can, for simple problems, restrict ourselves to a small set of associations that give us quickly a definite answer, or at least a highly probable one.

A mathematician or physicist working on a hard problem, or a composer, stalled in completing his current work, may associate in a much larger network in search of new connections and suggestive ideas. Such is the varied associative concept of meaning, the kind we use continually in our daily lives. Moreover, it is, of the three kinds I have introduced, the only one that, in its detailed elaboration, is private, due to the unique personal history of each person, the residue of which is stored in his or her associative network. For an elaboration of the place of small networks, especially semantic ones in the framework of much larger associative ones, see Steyvers and Tenenbaum (2005).

Empirical character of meanings. As I have moved to an ever wider circle of claims about meanings, it will be natural for those who hold to the analytic-synthetic distinction to believe that I am moving too far from any analytic conception of meaning. Such is the case. I am happy to use the word *meaning* but not to draw a sharp line between the meaning of a phrase and some empirical property usually associated with it. Perhaps the most common associate of the brain image of the word *Paris* is the brain image of the phrase *the capital of France*, and so we make it the dictionary meaning. But in terms of computing truth, or many other kinds of use of language or thought, that association does not really have a special status. There is no principled analyticity in that dictionary listing, and it could be easily changed by a move of the capital of France to Lyon, let us say, as happens occasionally to some capitals. The important point is that what we think of as the dictionary meaning need not have the status of being the most immediate or the most important association in many situations. These other associations may be the ones most recent, relevant and vivid at the moment.

The outcome of this kind of analysis that moves very far from the austerity of the set-theoretical structures of modern mathematics is meant to support the claim that there is a genuine plurality of kinds of meanings. As in many other things, there are different practices and different devices, including devices of language, for different occasions and purposes. It is not surprising that there is no one all-encompassing conception of meaning that is meant to be universally satisfactory. Moreover, the variety is certainly not exhausted by the three kinds of meanings considered here.

References

Hume, D. (1739/1951). *A treatise of human nature.* L. A. Selby-Bigge (Ed.). London: Oxford University Press.

Kuratowski, C. (1921). Sur la notion de l'ordre dans la théorie des ensembles. *Fundamenta Mathematica*, 2, 161–171.

Quine, W. V. (1992). *Pursuit of Truth.* Revised edition. Cambridge: Harvard University Press.

Steyvers, M. and J. B. Tenenbaum (2005). The large-scale structures of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41–78.

Suppes, P. (1957). *Introduction to Logic.* New York: Van Nostrand. Reprinted in 1999 by Dover, New York.

Suppes, P. (1960). *Axiomatic Set Theory.* New York: Van Nostrand. Slightly revised edition published by Dover, New York, 1972.

Suppes, P. (1980). Procedural semantics. In R. Haller & W. Grassl (Eds.), *Language, Logic, and Philosophy*, Proceedings of the 4th International Wittgenstein Symposium, Kirchberg am Wechsel, Austria 1979. Vienna: Hölder-Pichler-Tempsky, 27–35.

Suppes, P. and J.-Y. Béziau (2004). Semantic computations of truth, based on associations already learned. *Journal of Applied Logic*, 2, 457–467.

Suppes, P., B. Han, and Z.-L. Lu (1997). Brain wave recognition of words. *Proceedings National Academy of Sciences USA*, 94, 14965–14969.

Suppes, P., J. Epelboim, B. Han, and Z.-L. Lu (1999a). Invariance between subjects of brain wave representations of language. *Proceedings National Academy of Sciences USA*, 96, 12953–12958.

Suppes, P., J. Epelboim, B. Han, and Z.-L. Lu (1999b). Invariance of brain-wave representations of simple visual images and their names. *Proceedings National Academy of Sciences USA*, 96, 14658–14663.

Suppes, P., M. Perreau-Guimaraes, and D. K. Wong (2009). Partial orders of similarity differences invariant between EEG-recorded brain and perceptual representations of language, *Neural Computation*, 21, 3228–3269

The Oxford English Dictionary (1933/1970). Oxford: Oxford University Press.

Wiener, N. (1914). A simplification of the logic of relations. *Proc. of the Cambridge Philosophical Soc.*, 17, 387–390.