

Francis Hsu: "Plato as Software Designer"

"Plato's Ideal Types helps explain not only how our minds work, but perhaps also how computer software should work"

[About Francis Hsu: "I've worked with computers all my life. However, coming from a refugee family, I have an interest in politics, war, history and technology's impact on the human condition. These have caused me to think in Theory of Everything (TOE) terms of humanity, which is probably why I conflated Plato's Ideal Types to software and then to how human minds work. Essential facts: Birth: China. Education: Rutgers College, Political-Science-Economics (few credits shy of degree). Work: Data-Systems Architecture for large organizations, mainly Fortune 100. Now working for the U.S. State Department. Of course, this paper has little to do with my present work, but one thing I've noticed is that over human history, what were once thought of as very abstract things (Plato's ideal types, political equality, risk analyses, quantum theory, relativity theory, relational DBMS) are becoming essential to how we live." Francis Hsu: FH@ieee.org

Software Architecture and Plato's Ideal Types: How Minds Store and Use Symbols

Introduction

Human language is the first software. Software began when humans first used their minds and language to express themselves and communicate. This natural software, however, is still little understood. This essay establishes that the abstractions which human language allows enabled Plato to pose a question which still haunts us today.

When Plato spoke of Ideal Types [1] he was using language to try to understand the relations between the real (or concrete) and the ideal (or abstract). In doing so, he was exploring how our minds work, even if that was not his intention. His Ideal Types has perplexed thinkers over 2,500 years: they have argued about what Plato believed, what he was trying to do, and whether his notion is true or not. Few human conceptions have had such longevity --and that alone is sufficient for us to re-consider it in our computer-driven age. Recently, Plato was even blamed for fostering extremism in religions [2] because of this notion of Ideal Types.

I believe Plato's Ideal Types helps explain not only how our minds work, but perhaps also how computer software should work.

Quantity

How does the brain/mind know to store only one copy of any symbol? The answer to this question is critical to understanding how the brain/mind works. Putting the question in this way seems to confirm Plato's notion of ideal types --in that having one copy of any symbol is equivalent to an ideal type.

In fact this idea is even easy to test. In the English language (although this is not limited to English alone), studies show that texts have a consistent set of word patterns. In particular, word frequency is well known (especially as the number of words a text has becomes larger). Consider the novel "Moby Dick," reputed to be the longest novel in English, with some 1.5 million words. Table 1 shows the word frequency of the top 20 English words. What the table shows is that, given a large sample of text, 6.91% of the words will be the article 'the'. The verb 'is' occurs just under 1% (0.997) of the time. The first noun-type (actually a pronoun) on this list is 'he' --10th on the list. Most of these are

not content-words but connectors or structure-words. They help make sense or give meaning to the content words near them.

How does this confirm that our minds keep only one copy of a word or symbol?

If you read "Moby Dick" do you believe that your brain keeps 103,685 copies of the word 'the' in your mind? ($1,500,000 \times 0.0691232 = 103,685$). Or 54,021 copies of the word 'of'? ($1,500,000 \times 0.036014 = 54,021$). And all copies of the tens of thousands of words in English? (Or any other language for that matter.) The brain/mind clearly does not. Although it can not, of itself, tell us so. It is also, evolutionary-speaking, inefficient. And one thing that's very clear about evolutionary survival is that it's very efficient. It never wastes what it has and can readily use or re-use, nor does it adapt new devices for a function except when no other means is possible.

It is clear now, how Plato formed the notion of 'ideal types'. It was not only a philosophical issue, it was by its very nature an efficiency-survival issue. The brain/mind had already evolved (arrived at) the answer. When Plato posed the question, he was unaware (maybe he was aware!) that his ability to do so had already confirmed that that must be how the brain/mind works. His posing the question was in effect the casting which by its nature must fit the mold exactly.

Table 1 The Brown Corpus [3] of word counts in the English language.

Count	Word	%
70,008	the	.0691232
36,475	of	.0360140
28,937	and	.0285713
26,245	to	.0259133
23,530	a	.0232326
21,422	in	.0211512
10,597	that	.0104631
10,100	is	.0099723
9,816	was	.0096919
9,542	he	.0094214
9,500	for	.0093799
8,769	it	.0086582
7,291	with	.0071988
7,258	as	.0071663
7,000	his	.0069115
6,767	on	.0066815
6,391	be	.0063102
5,383	at	.0053150
5,347	by	.0052794
5,253	I	.0051866

Mathematicians describe numeric symbols in two ways, by their cardinality and/or their ordinality. Cardinality describes a number as a quantity. Thus every number has the cardinality of itself. For example, the number 1,000 denotes some quantity which includes all the values from 1 to 1,000. In this sense, cardinality is about totality and inclusion. Ordinality describes the order or sequence of a number or item. For example, in a 500page book, page 5 of the book is not

necessarily the fifth page of the book. The notion of fifth concerns order or sequence. Books have a title page, content page, introduction, sometimes a preface. These all push back the order which applies to page 5 of the book. It could well be that page 5 is the twelfth or seventieth page of the book. In this sense, ordinality is about 'before', 'after', and 'exclusion'.

This cardinality-and-ordinality contrast applies as well to non-numeric symbols. Every symbol has a cardinal value, mainly itself. This is the original copy, of any symbol of human artifacts, regardless of the medium in which it exists. As we grow up, our minds get imprinted with a copy of all these symbols. When we learn a symbol for the first time, that becomes our cardinal copy. Once learned, when we encounter the same symbol, our mind only has to make references to that cardinal copy.

Making references is an ordinal act. The referenced symbol already exists. We usually do not keep track of how often or the frequency with which we refer to cardinal copies. For example, all children learn the word 'mom' early in their lives. (By which I mean the linguistically specific form referring to mother, a person who gives birth to children, not necessarily the alphabets 'm', 'o', 'm'.) Yet no adults can remember on what day or where the word 'mom' was uttered by them for the 10th or 1,000th or 1,000,000th time in their lives. We can also be quite sure that no individual human has ever uttered the word 'mom' a billion times. This is because humans express themselves in context. And human life is not long enough to require uttering 'mom' that often. Such expressions are ordinal references to cardinal symbols.

The fact that humans do not keep track of the frequency of use for any specific symbol implies that, at least for how our minds work, it is not important for survival. It would confirm that in reading "Moby Dick" or anything else, when we come across a word or phrase for the 1,024th time, we do not know or care that it was the 1,024th time. Our brain/mind only keeps each symbol once --its cardinality. This is logical and sensible, for to keep the ordinality of any symbol is generally wasteful and useless.

Sidebar: In the 19th century, the mathematician Georg Cantor re-interpreted the notion of ideal types. He offered the definition: "A set is a Many which allows itself to be thought of as a One." Now that is a beautiful way to put ideal types. And it fits perfectly with the way humans use words and language. There are many ways to interpret his definition. One way is this: the One is the cardinality of a symbol, the Many is the references (ordinality) to which it is put. For example,

all languages, both living and dead, have the verb 'is', meaning an assertion of being or existence. Thus 'is-ness' is a One. When anyone anywhere uses the verb 'is', in any language, he or she is making a reference to the One. Each individual reference is a part, an instance of the Many.)

Referencing

The ordinal referencing of cardinal symbols shows another powerful aspect of our brain/minds. It is referencing. What happens when we reference something? We use symbols or words as substitutes for an entity or object. This is a not-so-subtle analogy of the mapping between the real to the ideal (Plato) and the many to the one (Cantor). It is powerful and efficient because symbols are arbitrary and without physical limits. Thus through symbols we can put heaven next to hell, marry Julius Caesar to Marilyn Monroe, compare apples to oranges, see the universe in a grain of sand, reflect on the sublime and the ridiculous, go from comedy to farce to tragedy, among other things.

This ability to unify specifics to the general, or vice-versa, is what gives humans the awareness of context. We easily flip-flop between the details to the overall, for any subject or topic and can do so between any of them. This is one source of humor: when a situation is put into a completely different context, we laugh because it is incongruous.

This referencing shows the immense flexibility, fluidity and adaptability of our minds. When we use symbols to refer to something else, our attention is on the something else, not the symbols themselves. The something else can be concrete (such as a ham sandwich, a shirt, a house) or can be abstract (such as goodness, an equation, governing). Let's illustrate this with the two-level diagram below:

|ENGLISH |BENGALI |CHINESE |HINDI |RUSSIAN ..

Symbols |MAN |Pur~S | 男人|cfbdL |ÌÓÆ◊ÈÍÀ |

The something else | a male human being |

Here a concrete something else is 'a male human being'. The symbols representing 'a male human being' in English, Bengali, Chinese, Hindi and Russian all look different. These symbols, if vocalized, would sound different. Yet of course they all refer to the same thing. When people talk using the above languages, they focus on the 'something

else' --not the symbol.

This is the reverse of one-to-many (cardinality-ordinality) mapping we saw before. Here many symbols in different languages map back to a single meaning. This is still within the Cantor's definition of a set, except now we are using it in reverse. This shows again how our mind can collapse and unify a class (the various language symbols) into an item (a male human being) without trouble. Of course, we do this without conscious effort.

The contrast with computers is stark. It is clear that computers do not use bit-patterns to refer to things outside itself. It is just an electronic recording device to keep symbols which we find meaningful. The following example shows how humans and computers differ. All modern computer software are strongly 'data-typed'. This prevents them from making mistakes no human ever would. Take, for example, telephone numbers, social security numbers, credit card numbers, or identification card numbers. Although we call them 'numbers', no human would ever think to add, subtract, multiply, divide or take the square root or the log of any of these. We know intuitively that, although called 'numbers', they are actually identifiers. In the computer's early years, 1950-60s, the lack of strong data-typing caused a lot of errors, because to computers any number is something it could do arithmetic operations on.

Quality

Once is enough: that's the brain/mind's storage algorithm. Yet, in what form does it store symbols? In short, what qualities must the symbols have? The brain/mind is efficient regarding its own capacity. It is most likely efficient also in its choice of symbol quality.

How do we identify the qualities of symbols in our brain/minds?

Symbols can come in many shapes, sizes, colors, media, sounds, intensities, clear or fuzzy, bold, italicized, orientation, order --the ordering here is semantic, not numeric --and other forms. Even before personal computers, many fonts for alphabets existed. (Although this author is not a font specialist, it seems that cultures with alphabetic languages have devoted much more artistic creativity to font design than others. Chinese culture prides itself on aesthetics of its language symbols. Yet until the modern era, say mid-19th century onward, there was precious little font design of Chinese symbols.) However one regards the Medieval monks, one must admire their dedication and artistry in transcribing the Bible and other ancient texts. The care and devotion they lavished on individual letters was clearly one source of font development. Now we are deluged by thousands of new fonts. Yet, when we see a font we've never seen before, we can read it without trouble. How is this possible?

This fact tells us something. The one copy of symbols in our minds is not sharp, clear, distinct, exact, or unambiguous. That symbol is not font-specific, it is amorphous, approximate, nearly-like. It has an 'about-ness', 'close-enough', or 'good-enough' quality to it. That symbol is font-free. This ambiguity makes it the opposite of symbols in digital computers. In computers all symbols are discrete, exact and un-ambiguous. In the real world in which humans live, things are analog, including the symbols in our minds. Analog things are by their nature, flexible, fluid, adaptable, useful in many contexts.

It is clear that patterns and shapes are etched on our minds. It seems equally clear that those patterns and shapes are not ideal types. That is, they are not perfect specimens or exact matches. Yet since there is only one for each symbol, we can treat them as if each were an ideal type. This is a tremendously powerful and efficient way for the mind to organize itself. After all, it's obvious that, given its small size (about 1,500 cc for the average adult), to have to encode every instance of experience (all the ordinal copies of every unique symbol) would exhaust its capacity quickly. By having just one shape of any symbol, any repetition of it is not a novel instance. In fact there is evidence that logographic languages such as Chinese (where there are tens of thousands of characters to remember) tax human memory. They are difficult to remember and easy to forget because there are too many of them. It is the brain/mind's ability to link these ambiguous unique symbols in any permutation (ordinality) that makes the mind efficient, effective and powerful in the extreme for survival.

Sidebar: Not only with word-frequency, but also in alphabetic languages, there is individual alphabetic letter frequency. It is because there are such anomalies in languages that cryptographers can break codes and read so-called secret writings. And of course, even at the level of each alphabetic letter, there seems to be no reason why our brain/minds would keep more than one copy of it.

Let's test to see if the 'sizes, colors, media, intensities, clear or fuzzy, bold, italicized' aspects of symbols are really inconsequential. This asserts that symbol meaning does not depend on those qualities. Symbol meaning depends solely on shape and to lesser degrees on orientation and order. Symbols have no inherent size. Their sizes do not denote that what they symbolize has greater or lesser meaning than symbols of another size. Consider the words 'big', and 'SMALL'. The fact that the latter is larger in printed form than the former in no way denotes that it is semantically larger. Take another example, the word 'Hollywood', has been emblazoned across a hilltop in southern

California for generations. It is probably the largest written name for a place in the world. Yet the city of Hollywood is not thereby larger in population than Bombay, India or larger in area than Rome, Italy. Finally, the bit pattern for lower case 'a' in ASCII is 01100001, In the 1950s when magnetic storage density was 200 BPI (bits per inch), lower case 'a' would have been visible to the naked eye. Today, when BPI is in the millions, no individual bit would be visible. So although the bit pattern 01100001 is orders of magnitude smaller, it still represents the letter lower case 'a'. So size is clearly inconsequential to a symbol's meaning. What about color? If the word 'Hollywood' appears in black, or gray, or pink, would that change its meaning? Again no. If the word is made up of solid rock in three dimensions, or made with letters with fuzzy edges, or made italicized, would that make any difference? No again. What about the medium it is displayed in? Suppose the word 'Hollywood' is spelled out with sand? or wood chips? or gold dust? or cosmic dark matter? Would any of that make a difference? No. What if the word blends in with the its background? That it is not in sharp relief and thus stand out, would that make any difference? In terms of its visibility, yes. Otherwise, no. With symbols, shape is primary with orientation and order having some significance.

Orientation is the direction in which we read symbols. Most human symbols are presented to be read or are on two-dimensional space that is flat and perpendicular to one's eyes. There are only a few possible ones: left-to-right, right-to-left, top-down, bottom-up, spiral-in, spiral-out, and various diagonal directions. The dominance of alphabetic languages means orientation has settled on left-to-right as the default. Chinese, which used to be read top-to-bottom, from right side of page to the left, has in many instances adopted the left-to-right convention. Of the major languages, only Arabic and Hebrew retains a different orientation: right-to-left.

Once adopted, orientation is nearly impossible to alter. Try reading this article from the last word, reading backward to the first. Orientation enforces a semantic order for words. Breaking this order would be disorienting and be as comfortable as sitting in a chair with one's belly on the seat and one's chest against the chair's back.

Finally, what significance does order have? Word order is crucial for languages lacking conjugation for verbs or declensions for nouns, adjectives, pronouns. For Latin or German, for example, which conjugate verbs and decline nouns, word order does not matter because their endings for verbs and nouns tells what role the word has in the sentence. In English and Chinese, on the other hand, word order is significant. Take the phrase: 'To be or not to be', just by shifting the words around to 'Be or to not be to' makes no sense at

all. In longer sentences, such shifting would make even less sense. Of course some people can read text upside-down. That implies that the mind knows to re-order symbols dynamically as the reading is in progress. It is less well known if anyone can read text from the opaque side of a page with ease. For example, if printed on clear sheets, then the text is visible from both sides. Whether it is equally readable is another matter.

In alphabetic languages, there is the additional need to have letter order within words. Yet letter order seems less significant than word order. Consider the following: 'Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttar in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be in the rghit pclae. The rset can be a total mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.' [4] Now most readers of English can read and understand those three sentences. That shows clearly that our brain/minds are not digital-only. Computers are now so pervasive that all users have experienced computers stopping for the simplest transposition errors of just two letters, much less entire series of them. It is remarkable that spelling errors are so marginal to our ability to comprehend. I, and probably many others, have experienced the strange sensation of seeing a misspelled word, yet even after staring at it for awhile I still can't provide the correct spelling; then, upon checking a dictionary, I see the correct spelling and it seems so obvious. Our ability to overcome confusion by misspelling (letter order) and get stopped dead by misplaced word order implies our minds process these differently.

Conclusion

When Plato proposed Ideal Types he became a software architect par excellence. It has taken a long time to appreciate his contribution to technology, but all great ideas have their gestation periods.

This essay identified three aspects of brain/mind symbols: quantity, referencing and quality. Quantitatively, our brain/minds store each symbol just once. This is storage-efficient and maximizes sharing of the similar. Referencing is a clever technique that exploits indirection and sharing. It permits comparison and contrast of things in imaginative ways in arbitrary realms. This is probably what provides us with context. Finally symbols must have two qualities: shape and orientation. A third quality, symbol order, depends on whether the language requires conjugation and declension or not. Symbol order is not important for conjugated and declined languages such as Latin or German, but it is important for non-conjugated and non-declined languages like Chinese or English. In addition, for the latter alphabetic

languages, letter order is not as important as word order.

Plato introduced Ideal Types to civilization. Why this concept has had such power and impact is now clear. He identified an inherent feature of our brain/minds. This feature provides stability, continuity and sharing among humans at an abstract level, just as our sharing DNA code with other living things provides continuity and sharing at a molecular (physical) level. This sharing at an abstract level is certainly a pillar of the foundation of knowledge and intelligence. After 2,500 years and a technological leap called digital computers, it has come full circle. It is clearer how our brain/mind works.

We started with the question: How does the brain/mind know to store only one copy of any symbol? While this essay has not answered how the brain/mind knows to do so, it has offered convincing evidence that it does indeed store just a single copy of symbols. In modern terms, Plato's Ideal Types means once is enough.

The challenge moving forward is: Can we duplicate what is within our skulls in a computer's address space?

[1] The Republic, Book VII by Plato

[2] The Dignity of Difference How to Avoid the Clash of Civilizations by Jonathan Sacks (London: Continuum, 2002) Chap 3.

[3] The Brown Corpus by WN Francis and H Kucera (Providence, RI: 1964) Brown University.

[4] The Times (London), 2003 Sept 23, p7.

Source: Ubiquity, Volume 6,
Issue 13 (April 19 -April 26,
2005)

<http://www.acm.org/ubiquity>