The Domestication of Meaning:

Predictable Trends in Semantic Evolution

Matthew Zellman

matt@zellmanlab.com

Zellman Lab Inc

Author Notes

https://orcid.org/0000-0003-1770-5579

Declarations of Interest: none

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Word Count: 8,022

Abstract

In this paper, I have provided justifications for situating meanings in the complex interactions between individuals, linguistic communities, and the physical world, and have explored some implications of such an understanding. Some linguistic communities apply selective pressures to certain kinds of meanings, and when those selective pressures are combined with the natural variations in meanings introduced by individuals through the process of definition, the result is an evolving system. Through analogy with domestication, I have argued that certain linguistic communities produce trends in the evolution of meanings that are consistent and predictable, based on sets of values that are characteristic of those linguistic communities. I have provided some examples of such evolution in several domains that show how the values of those different linguistic communities act on certain meanings in predictable ways. Finally, I have explored a few ways in which this understanding of meaning can or should impact our approach to some interesting problems in philosophy, education, artificial intelligence, and cognitive science; important implications for public and academic discourse; avenues for further investigation; and possible tests of this theory.

Keywords: Semantics, Evolving Systems, Meaning, Linguistic Community, Cognition

The Domestication of Meaning:

Predictable Trends in Semantic Evolution

1 The Meaning of "Meaning"

The meaning of "meaning" is, historically, a very sticky question. In that historical context, "meaning" is the subject of a dispute between semantics, semiotics, and pragmatics. Each of these disciplines has a plausible claim to sovereignty over the term and concept, but the reality of the situation is that the separation of these fields with respect to "meaning" introduces clarity only at the expense of accuracy.

> "I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics." (Lewis, 1970)

David Lewis is incorrect here. While mixing of these topics increases the complexity of investigation, it does not necessarily bring about confusion. In fact, it is impossible to discuss the meanings of symbols in any manner that does justice to our natural understandings of the terms without accounting for the complex interactions between symbols, the linguistic communities in which those symbols are used, the individual minds that comprise those communities, and the world in which all of these reside.

In normal usage, "meaning" is a property of a symbol, like an utterance, a word, or a gesture. It entails impressions of "intent" and "significance", and is commonly understood to suggest some synthesis of the speaker's understanding of their own intent and the listener's understanding of the significance of the symbolic representation. I consider this normal usage relevant because we use words primarily within our natural language communities, and the meanings of words as they are used and

understood within those communities are our starting points for understanding more technical definitions in our more technical dialects (more on that later, see 2.2.2). It is interesting to note that even though we generally ascribe a meaning to the symbol itself—usually in relation to something in the world—we understand that meaning in terms of relationships between the mind of a speaker, the mind of a listener, and that something in the world.

1.1 Two Perspectives, One Meaning

We can construct two distinct but similar theories of meaning from this understanding of the different factors that we incorporate into our ideas of meaning. From here, I will describe aspects of the world that participate in actual interactions with individuals or communities as their "environment", and those interactions as "environmental phenomena". Similarly, for the sake of normal understanding, I'm using "a communication" in a manner roughly synonymous with "a symbol".

The first theory is constructed from the perspective of the speaker, which we could call a theory of implication. A theory of implication must account for the interactions between: the environmental context of both the production and potential reception of the communication, the intended and potential unintended linguistic communities that make up the audience of a communication, the mind of the speaker, the ways the speaker has previously interacted with and participated in various other linguistic communities, and the speaker's history of interactions with their environment.

The second theory is constructed from the perspective of the listener, which we could call a theory of inference. A theory of inference must account for the interaction between: the environmental context of both the reception and imagined production of the communication, the known and potential unknown linguistic communities within which the speaker is participating, the mind of the listener, the ways the listener has previously interacted with and participated in various other linguistic communities, and the listener's history of interactions with their environment.

There are clear and intentional parallels between these two theories, and both perspectives can be generalized into a broader theory of meaning that attributes a meaning to a symbol and situates that meaning within a complex collection of interactions between individual minds, linguistic communities, and environmental phenomena.

2 The Development of Meaning

This is a strong claim, and in order to make it palatable, it will be necessary to provide support to this theory of meaning, facilitate understanding of the kinds of complex interactions that make up these processes, and clarify why it is better to just embrace the complexity than to try to reduce it. In that pursuit, let's examine ways that the three identified components contribute to the development of meaning.

2.1 Individual Minds

Let's start by establishing a definition for "concept": a concept is an individual's mental representation of a thing. Concepts, as best we can tell within our historical context, probably manifest in humans as some combination of individual neurons or networks of neurons (engrams) (Handjaras et al., 2016). Precisely how they manifest is not particularly important, as long as what they are manifest in is capable of making persistent associations with other concepts or collections of concepts.

Within neuropsychology, there is a maxim that could probably be considered its "central doctrine": "neurons that fire together, wire together" (Hebb, 1949). The reality of neuronal behavior is a bit more complex than that, but we can accommodate this with a less catchy, but more generally applicable maxim: "all neural processes are associative, such that the activation of neural entities in concert alters the strength or character of associations between the activated entities".

2.1.1 Association, Analogy, Schema, Abstraction

In this section, when I say "an x is a y", this should be understood in a sense like "an x is, at minimum, a y", or perhaps, "y is a necessary condition for the existence of x". Note, in particular, that an

abstraction is a schema is an analogy is an association is a concept, such that each level is not always well-defined, but that these words can be used to refer to mostly consistent levels of association, analogy, schema, and abstraction in many different contexts.

Associative behavior is extensible from single neurons to networks of neurons, to networks of networks, etc., and is not strictly confined to each level, i.e. a single neuron can be associated with a network of networks, and the connections between them will be influenced even though there is a disparity in their complexity. Associations can also be thought of as relationships, and their strengths and characters can vary in similar ways. (Handjaras et al., 2016)

An analogy is an association of associations. This is pretty straightforward; an analogy is basically saying that the relationship between these things is related to the relationship between those things. The form of this kind of structure is clearly represented in the standard formulation of analogies that you might see on a standardized test: “humans : houses :: birds : nests”. In other words, the relationship between humans and houses is related to the relationship between birds and nests.

A schema is an association of analogies. For example, If I have a schema like “animal family life”, that schema likely consists of associations of analogies like the above with analogies like “birds : nests :: foxes : dens”, “humans : children :: birds : chicks :: foxes : kits”, “children : corn dogs :: chicks : worms :: kits : rabbits”, etc.

An abstraction is an association of schemata or an analogy of analogies. A typical abstraction like “2+3=5” is constructed from analogies like “2 objects more than 3 objects is the same as 5 objects” (which would be an association of the relationships between particular object cases of the same operation), along with analogies like “2 units of liquid more than 3 units of liquid is the same as 5 units of liquid” and “2 units of distance further than 3 units of distance is the same as 5 units of distance”. Support for the idea that abstractions require analogies of analogies can be drawn from studies showing

that presenting subjects with two or more analogical scenarios (or associated schemata) promotes the construction of abstract representations of problems and facilitates transfer (Gick & Holyoak, 1983).

2.1.2 First Words

Our construction of understanding and meaning is clearly demonstrated during our acquisition of our first language. As we learn to interact with both our environments and our linguistic communities, we are participating in the construction of meaning, and building our own understanding as part of our normal developmental process.

As we learn new words, those words rapidly subdivide and refine conceptual space. Our first word generally starts as a performance—an action that we repeat because it gets us a positive reaction. In general, this first word is a single syllable (possibly repeated) with an easy consonant and a neutral vowel (ma, da, ha, ba, na), which some member of the language community interprets as a word appropriate to the context (mama, more, dada, that, yeah, dog, doll, hi, hot, ball, bottle, bye, nana, no, night, etc.). Its meaning is refined as new object labels are acquired, and as social feedback marks particular usages as appropriate or inappropriate—either by explicit rejection, or by simple relabeling. Additional words are given likely meanings based on associations in their linguistic context and in their environmental/experiential context. With repetition, we come to associate words with what is common to the experiential contexts in which we encounter them.

As there is less novelty in both contexts, we make associations with other groups of words based on similar usage in the linguistic context. By analogy with the referents of those words, we identify aspects of the experiential context that are similar and are lacking clear associations with an existing label. For example, if a child who understands the sentence “the cup is hot” is presented with the new word “blue” in the sentence “the cup is blue”, they will make an association between “hot” and “blue” based on the common usage. Based on that association, the child will look for a meaning for “blue” that is closely associated with the cup, ruling out possible meanings from unrelated aspects of

their environment. These kinds of associations form the basis of our implicit understanding of parts of speech.

Additional exposures combine with social interactions to perpetually modify the collection of associations that constitute an individual's concept of the meaning of a word. In a sense, our concepts of the meanings of words are distillations of the similarities present in community-curated collections of our experiences of our environments. (Clark, 2016)

2.1.3 Illusory Buckets

These "distillations of community-curated collections" that we associate with particular symbols (as labels) often appear to us to constitute sharply-defined categories, despite the varied strengths and characters of the associations that constitute them. In practice, we frequently conceive of categories as buckets or other containers, into which we sort various objects, aspects, phenomena, and events from our experience.

Most common words represent concepts as they were experienced centuries or even millennia ago. In the past, when populations were smaller and there was less opportunity to interact with as many diverse cultures and environments, the distinctions between categories would consequently appear to be far sharper. With fewer data points, it is much more likely that people would simply not encounter examples that proved problematic to their category concepts (or would be able to deny the significance of whatever exceptions they did encounter), reinforcing the idea that such categories reflect a natural division of the universe into distinct kinds of things.

The dominant metaphor of categories as buckets into which concepts are sorted is misleading enough to be actively discouraged. A better metaphor might be something like piles of sand, because the edges of a pile are often fuzzy and unclear, piles can overflow into each other, and the grains of sand can be disturbed and move around in the pile. In actuality, a category is a kind of schema: a collection of associated analogous concepts all tied to a common label (Hofstadter & Sander, 2013). Conceptually,

this manifests as an association between the label and a sort of constructed prototypical member of the category—we basically condense the entire schema into a prototype concept.

Some research has suggested that our mental representations of categories are defined by familial or similarity relationships between concepts or entities and a prototype entity or idea. This characterization is basically accurate, provided we remember that the prototype is (necessarily) both constructed and dynamic (Rosch & Mervis, 1975).

2.2 Linguistic Communities

Linguistic communities are, simply considered, groups of individuals who all communicate with other members of the group. I will be characterizing these communities by the nature of their communication, which can be loosely collected into three groups: natural languages and dialects, technical dialects, and formal languages. These three groups differ primarily in the extent of their dedication to precision and the specific value systems they enforce among their speakers. My use of "linguistic" should not be construed to mean that all such communities communicate through spoken or written languages, but should be understood to include any kind of symbolic signaling system.

An individual is likely to be a member of multiple different linguistic communities at any point in time, and their participation in those communities is likely to change over time. The interactions between individuals and their various communities affect the individuals' understandings of concepts and the spread of values between communities.

2.2.1 Natural Languages and Dialects

Natural languages are those spoken by communities of speakers that are not organized with respect to particular values, attitudes, or topics. This usage is, I believe, largely consistent with the way the term "natural language" is used in linguistics-oriented communities, although it may accommodate some constructed languages, as well.

While it is clear that natural languages are often spoken by communities that exhibit clear value structures, these features can generally be attributed to associated technical dialects that become extremely common or influential within the natural language community, or even become nearly coextensive with it, for some period of time. During such periods, a natural language community could be observed to exhibit a value system, but we should not consider a clearly temporally-defined subset of a linguistic community to be equivalent to the natural language community *as such*.

Natural dialects can be observed among natural languages with large numbers of speakers, or with relatively stable local communities. These account for regional variation within a natural language, and are also generally value-free.

2.2.2 Technical Dialects

A technical dialect arises whenever a community of speakers attempts to strictly define some subset of their natural language. Generally, this occurs in order to facilitate precision and clarity of communication with respect to some topic. Technical dialects also incorporate value systems that guide the usage of language among their members. Trivially, all technical dialects include some inclination to clarity and precision among their value systems.

Any kind of intellectual, cultural, religious, academic, political, professional, or interest-based association of people could be modeled by considering them to be a linguistic community speaking a sort of technical dialect. There is significant flow of words and meanings between technical linguistic communities and their broader natural language communities, at rates that generally reflect the influence of the members of the technical community among the natural language community.

2.2.3 Formal Languages

Formal languages are languages (in the broad symbolic systems sense) wherein all symbols and usage are precisely defined. There is a lot of discussion to be had on this topic, but it is not for this paper.

2.3 Environmental Phenomena

The third contributor to our theory of meaning is the universe we inhabit. My contention here is that while it exists prior to us and independent of us, we experience it entirely in ways that we construct in the context of our linguistic communities.

2.3.1 Perception, Association, Memory

The mechanism through which our environments interact with our minds is perception. Perceptual systems can be generalized with a definition like, “anything for which environmental stimuli result in some kind of change of state or structure.” For ease of reference, let’s call that “some kind of change of state or structure” *an activation of a thing*. In animals, that would be activation of a neuron or neural network.

There are two additional properties of our perceptual systems that determine how we perceive our environments: association and memory (Murray & Richmond, 2001). A similarly generalized definition of associative processes might be, “processes wherein activations of things that occur in close spatial, temporal, or conceptual proximity to one another result in further changes such that activation of one thing alters the activation of associated things”. I established above that all neural processes are associative, and perception is an excellent example of this. Nevertheless, it is possible to imagine perceptual apparatus that lacks this feature, such as an individual sensory device, so the importance of the associative process is worth reiterating.

Memory is conceptually a simpler feature than either perception or associative processes. It means simply that associative processes are persistent; associations between things are maintained and exist outside the immediate context of their activation.

2.3.2 Objects

An object is a thing that is represented in terms of sensory experience, is coherent, is persistent, and is categorizable (perhaps even actually categorized). This is, I think, largely consistent with the natural usage of the word.

People generally understand objects to be those things in the world that they are perceiving, but this understanding is erroneous. Every quality that makes an object an object—sensory representation, coherence, persistence, categorizability—is a quality that it has *only* by virtue of the way our minds are organized. Objects require perceptual apparatus in an associative system with memory to become objects. Perception is necessary for a sensory representation. Associations between individual perception events are what bind them together into coherent things. Memory of those coherent things is what allows us to conceive of things that persist through time. The categories that we have constructed are a consequence of our particular linguistic communities and histories of experiences.

As I stated previously, this is not to say that the world does not exist outside our minds. It is also not to say that the universe is homogeneous or unarticulated (differences do not inherently entail distinctions). It just does not exist *as objects*.

2.3.3 Classification

This claim that the world does not exist as objects is akin to a claim that there is no natural partition of the universe. Proposing that the universe is composed of different things is a very different claim than proposing that the universe is composed of different *kinds* of things, let alone that it is composed of kinds of things that are accurately approximated by our constructed categories.

An easy candidate for a natural partition is provided by the standard model of particle physics. We might claim, based on this theory of fundamental particles, that different categories of particles correspond to different kinds of things. To do so would be to overlook that even these categories are constructed according to a paradigm that humans have created, based on particular properties of

particles that we have decided are more important than others. Quantum phenomena may even be easier to digest when we recognize that even identifying something as an electron involves imposing the concept of a particle or wave onto some phenomenon that behaves to some extent like our constructed prototypes of both of these concepts, and that we simply use whichever concept is most useful in our current context.

For example, we have classified a set of physical entities as atoms, categorizing them by their number of protons as elements, because we place great importance on their chemical properties and this categorization schema provides us a reliable way to predict the chemical properties and behavior of any atomic entity. We could conceive of scenarios where we might place more importance on atomic mass or number of electrons, and classifications according to these criteria would constitute conceptions of the universe that are no less correct or accurate or true than our current model. A more fanciful inquiry into the classification of animals is well-known from Borges (1952).

2.3.4 The Real World

Even as we concede that objects qua objects do not exist outside of minds, we must also consider that minds are not creating objects out of whole cloth. The mind is creating objects based on memories and associations of perceptions of an external environment. Even though the mind is, in one sense, constructing our experience of reality, it does not follow that our experience of reality is equal to reality, or that reality exists as a consequence of our constructed experience. Minds exist as discrete dynamic entities within a real universe, and interact in complex ways with the universe and with other minds (within linguistic communities) to construct representations of their common environment that are largely consistent among the members of their linguistic communities. To put it concisely, I guess, reality is external to and prior to the mind, and our experience of reality is constructed out of the interactions between our minds, our environment, and our linguistic communities—out of our understandings of meanings.

We might consider the world to be a sort of shadow or wire mesh over which we superimpose cutouts of objects, partitioning our experiences of our environment into coherent and comprehensible objects and structures that generally resemble the world without necessarily capturing its fine structure. In the same vein, we could consider our mental representations of the universe to be roughly analogous to the manner in which a stick figure resembles a human—there is a clear structural similarity on a non-trivial level, but it does not do justice to the fine structure of any human organism.

3 What This Means

Meaning resides in complex interactions between the world, individual minds, and linguistic communities. Peculiar features of each of these contributors and the manners in which they interact determine how our understandings of concepts develop, how meanings change, and how those changes in meaning propagate through different linguistic communities.

3.1 Things Change

The world changes. Individuals encounter different environments. Individuals change. Individuals interact with different linguistic communities. Linguistic communities change. Against this backdrop, we must always consider that meanings of symbols are constantly in flux.

As a result, we must consider meaning to be dynamic on time scales similar to the scales of the changes experienced by the systems that entail it. We can only engage with meaning outside our immediate context insofar as we can identify and keep track of whatever relevant changes have taken place in its substrate.

3.2 The Unity of Mind

One of the major problems in Cognitive Science involves accounting for the apparent unity of the conscious mind (Bayne & Chalmers, 2003). With the understanding of “object” that has been given here, the answer to this question becomes clear. The conscious mind appears as a single indivisible entity because it is constructed out of associated perceptual experiences; that is, our brains associate

impulses for action with our perceptions of those actions being carried out by limbs we can see, motions we can feel, etc. In addition to this, our brains co-opt sensory systems for many internal processes, including our inner monologues and imaginations, which are also associated (sometimes by decently long chains of association) with impulses to action, and with each other. As these internal processes continue to associate with each other and produce effects that are either internally or externally perceptible, we build our experiences of them into a concept of self, and our experience of self essentially constitutes our conscious mind.

3.3 Too Good for Our Own Good

Linguistic communities are composed of individual minds that generally work in similar ways. As those communities are composed of human minds, they will be subject to systematic errors.

A particularly salient example of such systematic error involves the human predisposition to pattern recognition. Humans are extraordinarily good at recognizing patterns, to the extent that we often impose patterns on phenomena that are unrelated except for the associations we have made between them. (Meschiari, 2009)

Clinically, this tendency is labelled “apophenia”, but many examples of erroneous pattern recognition are subclinical and shared by large numbers of people. A good label for these community-wide errors of overzealous pattern recognition might be “constellations”. This calls back to an old and familiar example of this principle: identifying patterns in essentially random distributions of stars in the night sky. I’ve compiled a short list of examples of constellations of meaning, from a few different domains.

3.3.1 Constellations: Pareidolia

One of the most basic constellations is pareidolia—seeing faces in randomness. Our brains are so strongly wired to pick faces out of our surroundings that we identify face-like patterns as faces regardless of whether they are attached to a human (or other animal) body. Examples that could be

considered particularly meaningful include the man in the moon, the old man of the mountain, or an image of Jesus Christ on a piece of toast.

3.3.2 Constellations: Luck

Examples of words that are constellations are “luckiness” and “unluckiness”. We see illusory trends in random events and consider them to be attributable to some trait of those individuals who experience them, some token they have acquired, or some behavior they have adopted.

3.3.3 Constellations: Stereotypes

Race and gender prejudices are frequently a result of identifying patterns between unrelated phenomena, or of over-application of an observed pattern to individual cases.

3.3.4 Constellations: Personification

Ascribing emotional characteristics to weather events is an example of a constellation. Describing a storm as “angry” reveals an association between aspects of a meteorological phenomenon and aspects of human emotional expression. The pattern identified by the word “angry” is not a pattern that exists in the meteorological phenomenon itself.

3.3.5 Constellations: Pre-science and Pseudoscience

Various concepts from prescientific and pseudoscientific belief systems are examples of constellations. Dodging falsifiability criteria and appealing to confirmation bias are effective ways to reinforce ideas that are based on uncritical pattern recognition. Examples include: humours theory, astrology, personality types, new age crystals, homeopathy, etc.

3.3.6 Constellations: “Literal”

The word “literal” is a particularly notable constellation. It identifies a pattern that suggests that ease and clarity of association between symbols and perceived objects or states-of-affairs implies natural one-to-one mappings between them. As examples are presented suggesting that factors other than the words themselves contribute to the meaning of a communication, operations like

"contextualizing functions" are introduced to move data points that don't fit the constellation into conformation. Actually defining these operations and applying them to particular usages turns out to be more complicated and opaque than performing similar operations on explicitly metaphorical usages. (Israel, 2005)

3.4 Ghosts and Echoes

Linguistic communities are continuous entities that are capable of indefinitely outlasting any of their members. This continuity beyond generations and lifespans suggests that the etymologies of words, as far as we can determine them, can provide significant insight into the development of meanings over time and into nuances of contemporary meanings. Etymology does not determine meaning by itself, but the ghosts and echoes of long-past usages still inform our understanding of words now, through the traditions of appropriate usages that are maintained, propagated, and perpetuated by linguistic communities.

3.5 Definitions

Definitions of words serve two purposes. Within technical dialects, a definition will frequently be used to prescribe usage of a word, either to constrain or expand its appropriate usage. Within natural languages, definitions are necessarily descriptive, because there are no universal value systems in place to enforce a prescriptive definition.

Definitions in natural language, then, constitute attempts to define rules against which the usage of a word can be tested in order to determine whether it is likely to be considered appropriate or inappropriate by some significant portion of the natural language community.

Definitions and values from technical dialects frequently leak into natural language communities, which can lead usage within significant technical communities (notably educational communities) to conform more closely to both technically prescribed and naturally described

definitions. This trend then carries over into the natural language communities that surround those technical communities.

3.6 Records

Written and recorded language allows for extraordinarily complex situations regarding meaning and usage. Recorded language enables the divorce of an utterance from its speaker, giving it the illusion of timelessness and enticing us to ascribe meaning to the recorded statement itself.

Meaning in recorded works is functionally determined by the mind, environment, and linguistic communities of the recipient (listener, reader). In certain contexts, these variables can be situated in such a way as to attempt to simulate the conditions wherein the original work was recorded, but there are fundamental limitations on the certainty we can ascribe to the accuracy or fidelity of such simulations. This implies that recorded work, or, in fact, any non-negotiated communication (or perhaps at least any such communication which is received outside the speaker's narrow linguistic community or developmental/environmental contexts), is perhaps rightly considered under the "death of the author" or reception theory paradigms.

What an author intended when recording a communication can at times be approached, but it cannot be reliably ascertained from an original recording outside the author's own environmental and linguistic community contexts. Whether the author's intent is an important consideration is a different question altogether (and may also depend on context).

Recall, also, that these meanings are not inherently deficient or erroneous (although within a particular context attitudes may exist that impact its respective meaning in relevant ways). These are what recorded works *actually mean*. These meanings vary across different contexts and change over time, and are at all points rightly considered to be what a work actually means.

4 Evolving Meanings

I've established that meanings change over time and suggested that those changes can be directed by the values of technical linguistic communities, based on the influence they exert within their natural language communities. In this section, I will explore some ways in which the evolution of meanings is analogous to biological evolution, describe what kinds of fitness functions might be entailed by the value systems of technical communities, look at the mechanisms that drive selection within semantic spaces, and provide several examples of words whose meanings have evolved in ways consistent with these principles.

4.1 Biological Analogies

Evolutionary mechanisms imply trends towards greater fitness within a static or slowly changing niche or environment. In order to understand how evolutionary mechanisms could act on the meanings of words, we need to understand what selective pressures might apply to meanings and what fitness might mean with respect to meaning in the context of linguistic communities.

These mechanisms are perhaps best examined in the context of another biological analogy. Thomas Kuhn's paradigm shifts are analogous to biological cataclysms. Changes in the meanings of certain significant terms cascade through the rest of the semantic space, changing the shapes of existing niches, opening new niches, and altering the factors that define fitness. As a result, similar changes in meaning take place in large numbers of related terms.

Fitness of meanings in scientific communities would then be a function of a combination of traits of a particular representation, interacting with values like simplicity, aesthetic appeal, usefulness, vulnerability to falsification, and predictive accuracy, among others. (Kuhn, 1970)

4.2 Fitness Functions

Fitness functions can operate based on any number of values present in linguistic communities. Values can act on meanings in several different ways. These functions can act expansively, tending to

increase the number of appropriate usages, or they can act restrictively, tending to decrease the number of appropriate usages. Fitness functions can also act directionally, simultaneously encouraging some new usages while discouraging some existing usages.

Fitness functions often behave differently based on the emotional connotations of the words or meanings they are acting on. Some sets of values may entail fitness functions that tend to act expansively on words with positive connotations and restrictively on words with negative connotations, while others may do just the opposite: acting restrictively on words with positive connotations and expansively on words with negative connotations.

Some noteworthy and easily observed fitness functions are created by value systems that emphasize one particular contributor to meaning. “Traditionalist” value systems create fitness functions that evaluate meanings based on the extent to which they resemble meanings as they are presumed to have been according to our understanding of some earlier linguistic community. “Physicalist” value systems create fitness functions that evaluate meanings based on the extent to which they resemble our understanding of the observable universe.

4.3 Selection Processes and Biases

Concepts undergo selection based on their conformity to both perception of environmental contexts and appropriate usage within a community of speakers. These selection processes are rooted in cognitive dissonance and social pressure, respectively. Experiences of environmental contexts/phenomena that are incongruent with existing concepts weakens the coherence of those concepts and causes discomfort to individuals. Similarly, social pressure to use a word in ways that conform to its current meaning tends to actively discourage usages that the linguistic community considers inappropriate.

Defense mechanisms exist that tend to reinforce existing concepts in preference to altering those concepts to accommodate new information. The existence of these mechanisms implies that, in

the absence of correction mechanisms in the social/cultural context, concepts do not intrinsically tend towards better representations of environmental phenomena. (Kahneman, 2011)

These defense mechanisms exist because the accuracy of mental representation is not the only, or even the primary, evolutionary driver of mental processes. For example, we associate certainty with confidence, confidence with strength, and uncertainty with weakness. Responding to new information by showing uncertainty, while beneficial from the perspective of developing accurate concepts and meanings, has tended to be detrimental to social standing and other factors affecting individual fitness. Historically, it has been better to be wrong and strong than weak and right, at least as far as those affect individual survivability and reproductive desirability. Most things that people can be wrong about are not overwhelming existential threats such that they cannot be overcome (or at least ignored) with sufficient physical or social strength or resourcefulness. (Kreps et al., 2017)

On the scale of societies and communities, similar conflicting forces are at play. The values that affect the development of meaning within a linguistic community depend on where that community identifies its threats, how it seeks to address those threats, and its tendency and strategy for the propagation of its memes.

4.4 Definition and Domestication

One of the principal difficulties with drawing on evolutionary biology for our primary analogy is the unpredictability of evolutionary trends in practice. There are, however, some special situations where evolutionary processes can not only be predicted, but controlled—most notably, those involving domestication.

Technical linguistic communities are essentially environments promoting the domestication of some set of meanings, using definition as their primary mechanism for artificial selection. Just as crop species evolve in ways that are dictated by the needs and values of agricultural communities (greater yields, disease/pest/weather resistance, easier harvest, etc.) (Gepts, 2010), and pet species evolve in

ways that are dictated by human social needs, economic needs, and values (less aggression, more expressive faces, compulsive working behaviors, etc.) (Wilkins et al., 2014), meanings evolve in ways that are dictated by the needs and values of the technical linguistic communities that domesticate them. Insofar as we can identify the needs and values of particular technical linguistic communities, we can predict evolutionary trends for some aspects of words and meanings that are being defined by those communities.

4.5 Examples

I have identified a few clear examples of different ways that fitness functions have been observed to operate on the meanings of specific words, based on values that can be attributed to specific technical communities.

4.5.1 Creative

The word “creative” is frequently employed in educational and artistic communities where we tend to find values associated with encouraging self-determination, self-esteem, humanism, and empowerment. As creativity is a concept with a strong positive connotation, members of these communities will frequently be reluctant to consider the use of the term inappropriate when members use it to describe their own work or the work of others. Its use as a common and/or generic positive assessment of children’s artistic output is a testament to this tendency. The effect of these values on the meaning of the word “creative” over time is clear. While the word was originally reserved for divine acts of creation, its range of appropriate usages has gradually expanded to include human acts of unprecedented artistic talent and innovation, to acts of more mundane artistic talent and innovation, to acts of talent and innovation in non-artistic domains, to general aptitude for divergent thinking, etc. (Hanchett Hanson, 2015)

4.5.2 Intelligent

The meaning of the word “intelligent” follows a path similar to the meaning of the word “creative”. It is used in similar communities, notably educational communities, and is similarly endowed with strong positive connotations. The inclination not to deny a label of “intelligent” to anyone has led, in part, to theories of multiple intelligences, which helped to expand the meaning of the word to include proficiency in non-intellectual domains, for instance.

4.5.3 Fish

The word “fish” used to refer to any animal that lived entirely in the water (cf. shellfish, starfish). Because of the influence of biological scientific dialects, and their associated values of enabling clear classification (under Linnaeus), then an emphasis on family resemblance (under Darwin), and now an emphasis on familial relationships (also under Darwin, but with additional assistance from genetics to inform cladistic relationships), the meaning has changed: first to refer to aquatic vertebrates with gills, fins, and scales; and now to animals that share a common ancestor with all other fishes, (usually) excepting members of the nested clades of amphibians, reptiles, birds, and mammals.

4.5.4 Dinosaur

Similar to “fish”, and under the same kinds of influences, “dinosaur” once referred only to certain kinds of extinct reptiles known only through fossil evidence. The current meaning of the word includes (or is starting to include) living birds, as well. Usages like “non-bird dinosaurs,” can be observed being used to refer to dinosaurs as they would have been understood under earlier meanings of the word.

4.5.5 Adult

The word “adult” has been affected by legal, political, educational, religious, biological, and psychological linguistic communities. The values involved have included precision in classification, fairness, biological accuracy, psychological accuracy, etc. Over time, “adult” has meant, variously and in

no particular order, “a person who is married”, “a person who is sexually mature”, “a person who is psychologically mature”, “a person who doesn’t live with their parents”, “a person who has a child”, “a person who has a job”, “a person who is over 15”, “a person who is over 16”, “a person who is over 18”, “a person who is over 21”, or “a person who is over 25”.

The different degrees of influence that the wide variety of different linguistic communities have exerted on their natural languages over time have made the evolutionary path of this meaning particularly convoluted, and in the broad view, it may appear sort of random. When considered alongside the specifics of those dialects over time, however, we do see consistent evolutionary mechanisms at work. Trends in the meaning of “adult” follow the values of those linguistic communities that are most influential at any given point in time.

This example illustrates the kind of behavior we should expect to observe when environments change more rapidly with respect to the changes in meaning.

4.5.6 The Euphemism Treadmill

The euphemism treadmill is an interesting case where there exists a certain “niche” for a word where the niche itself is associated with a generally taboo subject with a significant negative emotional component, including disability, sex, excrement, religion, etc. The niche tends to render any word that is associated with it distasteful under some influential value system. As a word is adopted to fill the niche, its association with the underlying negative emotional component—which derives from its experiential associations—gradually renders it similarly distasteful, until it is replaced with another new word. This can be observed with words like “crippled” > “differently abled”, “retarded” > “developmentally disabled”, and to a lesser extent “arse” > “bottom”. Contrast these with examples like “shit” > “crap”, “damn” > “darn”, etc., which are euphemisms without as much of the treadmill effect. (Pinker, 2003)

5 Conclusions

This theory has widespread implications across many disciplines, provides a broad context for several existing theories, and suggests avenues for scientific investigation of a number of topics that have historically resisted such investigation.

5.1 Interesting Implications

The existence of active evolutionary mechanisms in semantic spaces suggests that the existence of certain kinds of value structures in a linguistic community means that trends in meanings within that community will be predictable, and meanings that have developed according to that value structure will be likely to exhibit greater fitness with respect to it. In other words, if we're looking for truth (for example), we should figure out what kinds of values would result in meanings that approach whatever it is that we consider truth to be, and focus on communities concerned with truth that exhibit those values, instead of evaluating investigations of truth without consideration of the values of the communities within which those investigations take place. Being able to judge how competing meanings from different technical communities are likely to apply to a particular topic of discussion may allow us to resolve differences that previously seemed intractable, such as whose definition of a key term should drive a discussion, or be considered preferential in some context.

Because meanings are subject to develop around constellations, we need to be aware of the possibility that terms that we use, even common or meaningful terms, may refer to associations of phenomena that are not connected with each other in reality.

Academic investigation into the meaning of some word or concept is likely to influence the actual meaning of that concept in predictable ways. Most forums for academic investigations encourage precise definitions of the ideas being investigated, and the ways academics define their terms are likely to significantly impact broader understanding of the ideas they are investigating.

The technical community of educators may be the technical community with the most influence on the development on the natural English language community, if not natural language communities generally, because nearly all English speakers receive some of the most impactful feedback on their usage during their formative periods from members of the community of educators. While we generally have a sense for how important educators are for the education of the populace, I believe understanding the ways in which educators are major contributors to concept formation and the development of meaning in the natural language community at large suggests that their importance with respect to the development and maintenance of cultural values has been significantly underestimated.

Being aware of the values that determine the development of meaning provides a number of ways for people to try to influence the values of groups or to attempt to alter meanings within influential groups to take advantage of particular sets of values. It also means that we need to be aware of the values that govern the technical communities we participate in and identify potential abuses of our communities. To draw on a biological analogy, we should probably be wary (or at least aware), of invasive species.

This theory of meaning has significant implications for the development of strong artificial intelligence and other artificial intelligence applications that incorporate natural language processing. An artificial intelligence that aims to communicate meaningfully with humans would benefit from consideration of the principles laid out here. An artificial intelligence must be constructed to perceive, associate, and remember. In order to communicate meaningfully, it needs to perceive the world somehow; it needs to be able to create analogies, schemata, and abstractions to construct objects and events; and it needs to be an active participant in at least one linguistic community.

5.2 Further Investigation

There are numerous significant opportunities for expansion of this theory, and quite a few opportunities to test its predictions.

5.2.1 Expansions

I have not attempted to provide a detailed model of the kinds of interactions that take place between the environment, individual minds, and linguistic communities. The development of such a model would significantly improve our understanding of language, cognition, and perception.

I have not attempted a thorough investigation of the kinds of values we might find in technical linguistic communities or specific descriptions of their fitness functions or interactions with particular features of symbols. I have also not attempted to describe or categorize particular technical linguistic communities, beyond a cursory exposition of “traditionalist” and “physicalist” orientations.

I have only traced the evolutionary paths of a few particularly salient words. Additional investigation into additional words or even larger semantic entities (phrases, sentences, essays, stories) would provide a rich understanding of the intricacies of the cultures they have passed through and the selective processes they have been subject to.

5.2.2 Tests

Since Karl Popper, one important criterion of a scientific theory is its falsifiability. I believe that this theory could be falsified and have identified some investigations that could potentially disprove it.

If a word could be traced through a technical dialect that could be expected to subject its meaning to some predictable fitness function, and the meaning of the word develops in a manner contrary to what we would expect from the values of the technical community, this theory would be false.

If technical dialects could be shown to frequently exercise inconsistent values over short time spans in unpredictable ways, this theory may not be proven false, but such a result would suggest that

its application could be too complicated to be useful, or that its usefulness is limited to applications involving select linguistic communities that behave amenably.

6 Acknowledgements

Two individuals have had especially significant impacts on the development of this theory, and deserve to be recognized for their influence and contributions: Michael Hanchett Hanson of Teachers College and Joshua Kerr, formerly of the University of Oregon. There is also one major work that should be considered an important precursor to this, and which covers much of the same ground, but which I was not familiar with when I wrote this: Mark Richard's *Meanings as Species* (2019).

7 References

Bayne, T., & Chalmers, D. J. (2003). What is the Unity of Consciousness? In *The Unity of Consciousness: Binding, Integration, Dissociation* (p. 41). Oxford. http://consc.net/papers/unity.html

Borges, J. L. (1952). El idioma analítico de John Wilkins. *Otras Inquisiciones*, 158–159.

Clark, E. V. (2016). *First Language Acquisition*. Cambridge University Press.

Gepts, P. (2010). Crop Domestication as a Long-Term Selection Experiment. In J. Janick (Ed.), *Plant Breeding Reviews* (pp. 1–44). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470650288.ch1

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Hanchett Hanson, M. (2015). *Worldmaking: Psychology and the Ideology of Creativity*. Palgrave Macmillan Limited. http://ebookcentral.proquest.com/lib/teacherscollege-ebooks/detail.action?docID=4082226

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., & Pietrini, P. (2016). How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage*, *135*, 232–242. https://doi.org/10.1016/j.neuroimage.2016.04.063

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hofstadter, D., & Sander, E. (2013). *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books.

Israel, M. (2005). Common Sense and "Literal Meaning." *The Literal and Nonliteral in Language and Thought*, 31.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology*, *113*(5), 730–752. https://doi.org/10.1037/pspi0000103

Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago press.

Lewis, D. (1970). General semantics. *Synthese*, *22*(1–2), 18–67.

Meschiari, M. (2009). Roots of the Savage Mind. *Quaderni Di Semantica*, *30*, 183–222.

Murray, E. A., & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current Opinion in Neurobiology*, *11*(2), 188–193. https://doi.org/10.1016/S0959-4388(00)00195-1

Pinker, S. (2003). *The Blank Slate: The Modern Denial of Human Nature* (Reprint edition). Penguin Books.

Richard, M. (2019). *Meanings as Species*. Oxford University Press. https://doi.org/10.1093/oso/9780198842811.001.0001

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The “Domestication Syndrome” in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics*, *197*(3), 795–808. https://doi.org/10.1534/genetics.114.165423