Explaining Embodied Cognition Results

George Lakoff

Department of Linguistics, University of California

Received 15 February 2012; received in revised form 21 May 2012; accepted 23 May 2012

Abstract

From the late 1950s until 1975, cognition was understood mainly as disembodied symbol manipulation in cognitive psychology, linguistics, artificial intelligence, and the nascent field of Cognitive Science. The idea of embodied cognition entered the field of Cognitive Linguistics at its beginning in 1975. Since then, cognitive linguists, working with neuroscientists, computer scientists, and experimental psychologists, have been developing a neural theory of thought and language (NTTL).

Central to NTTL are the following ideas: (a) we think with our brains, that is, thought is physical and is carried out by functional neural circuitry; (b) what makes thought meaningful are the ways those neural circuits are connected to the body and characterize embodied experience; (c) so-called abstract ideas are embodied in this way as well, as is language. Experimental results in embodied cognition are seen not only as confirming NTTL but also explained via NTTL, mostly via the neural theory of conceptual metaphor.

Left behind more than three decades ago is the old idea that cognition uses the abstract manipulation of disembodied symbols that are meaningless in themselves but that somehow constitute internal “representations of external reality” without serious mediation by the body and brain. This article uniquely explains the connections between embodied cognition results since that time and results from cognitive linguistics, experimental psychology, computational modeling, and neuroscience.

Keywords: Embodiment; Embodied cognition; Metaphor; Cognitive linguistics; Lakoff

1. The old disembodied symbol manipulation theory

It may be hard to think back to a time before the idea of embodied cognition, but I was raised in that generation. Concepts were seen as characterized by inherently meaningless abstract symbols that got their meaning via the ability to “represent” external mind-free
reality, without the body or brain playing any significant role. The mind was seen as manipu-
ulating abstract symbols without regard to their meaning.

Language was seen through the Chomskyan metaphor: A sentence is a string of meaning-
less abstract symbols; a language is a set of such strings; and a grammar is an algorithmic
method of generating such sets of strings, independent of meaning or communication or any
aspect of embodiment. The brain and body were seen as irrelevant (Lakoff & Johnson,
1999, Chapter 22).

During the 1960s and early 1970s, generative semanticists like myself, James D. McCawley,
Haj Ross, and many others found hundreds of counterexamples both to the Chomskyan account
of language and to the very idea that meaning could be seen as internal representation of an
external, mind-free reality.

Take my ordinary 1963 baseball example, ‘‘Yastrzemski doubled to left.’’ ‘‘To left’’ is a
directional adverb that fits with a moving object. But the sentence, and even its ‘‘deep struc-
ture,’’ has no symbolic representation of such an object or motion. The moving object is the
ball, which is understood through the meaning of ‘‘double’’ in baseball. ‘‘Field’’ in ‘‘to left
field’’ can only be omitted in the context of a baseball game. The groundskeeper could be
‘‘mowing left field,’’ but not ‘‘*mowing left.’’ Symbol manipulation without regard to the
meaning of the symbols could not work (Lakoff, 1963).

And take my 1968 example, ‘‘If I were you, I’d hate me’’ versus ‘‘If I were you, I’d hate
myself.’’ In the if-clause, my consciousness is projected into your body. This cannot happen
in the real, external, mind-free world. It cannot be understood in terms of a ‘‘representation’’ of
external reality. Moreover, the ‘‘me’’ and the ‘‘I’’ in ‘‘I’d hate me’’ both do and do not refer to the same person. They refer to the same locus of consciousness, but different
persons. In ‘‘If I were you, I’d hate myself,’’ ‘‘I’’ and ‘‘myself’’ refer to the same person—but that person is you not me, whereas the ‘‘I’’ refers to my locus of consciousness.
This is not an account of reference that fits an ‘‘internal representation of external real-
ity’’ (Lakoff, 1968 ms).

If you grew up believing the symbol manipulation theory of mind and language, such
counterexamples were a big deal. We found them everywhere. But in the age of embodied
cognition, they seem quaint.

2. Embodiment in the emergence of cognitive linguistics: 1975

I was thoroughly shaken out of disembodied cognition in the summer of 1975. Four dra-
matic talks at Berkeley pointed to embodiment in the conceptual systems used by natural
languages.

1. Paul Kay, working with Chad McDaniel (Kay & McDaniel, 1978), observed that color
terms name colors, and that colors do not exist in the external world independent of
bodies. Colors have external requirements (e.g., wavelength), but they do not exist
without embodiment: Color cones in the retina and complex neural circuitry connected
to those color cones. Thus, the external world alone cannot account for ‘‘green.’’
2. Eleanor Rosch (Rosch & Mervis, 1981) showed that basic-level categories are defined by three kinds of embodiment: gestalt perception, mental imagery, and motor programs. These all require the body and brain. Her experiments revealed that basic-level categories were processed faster, recognized more easily, and learned earlier and more easily than non-basic-level categories. Moreover, you need the brain and body to understand the meaning of “chair” or “car.”

3. Len Talmy (1983, 1988, 2000a, 2000b) showed that spatial relation terms, across many languages, are based on widespread if not universal primitives that require reference to the body. Langacker (1987, 1990, 1991, 2008), working independently, made the same discovery. These are called “image schemas” or “cogs” (short for cognitive primitives) since they are components of visual perceptions, mental images, and other cognitive structures.

4. Fillmore (1976, 1977, 1982) argued that every word in every language is defined in terms of frame structures, which are not in the external world. Rather, they fit semantic roles concerning embodied notions such as agents, patients, recipients, sources, goals, beneficiaries, and so on.

For the most part, cognitive psychologists, philosophers, and linguists were unaware of, or ignored, such results in 1975 and for many years thereafter, until the publication of Women, Fire, and Dangerous Things (Lakoff, 1987) made them known. Meanwhile, embodiment studies proceeded in the field of Cognitive Linguistics, which began with those 1975 talks at Berkeley.

3. The embodiment of “abstract” concepts: Cogs

It should not be all that surprising that concepts for literal embodied actions, like running, actually draw upon physical experiences of running. That would be a literal approach: The concepts for what the physical body is and does are embodied.

What was surprising—and profound—was the discovery that so-called abstract concepts are embodied as well. There are two kinds: cognitive primitives and primary conceptual metaphors.

3.1. Cognitive primitives

Let us begin with “cogs” (cognitive primitives) or “image schemas” of the sort discovered by Talmy and Langacker. These are universal cognitive structures, either there at birth or developed very early. They structure visual perception, motor action, and mental images, and they are used in the semantics of natural language. The spatial relation cogs include structures like Source-Path-Goal, Containment (with Interiors, Exteriors, and Boundaries), Contact-Noncontact, Near-Far, Around, Along, Across, In Front Of, Behind, Beside, At, Toward, and so on. These words are English names for cross-linguistic cogs. Other cogs include Force Types (Talmy, 1988), for example, Directed Force, Resisting Force, Supporting

Cogs can also be Process Schemas; sequences of states or processes. For example, a Completed Purposeful Action Process has the following parts: a Precondition, a Starting Action Sequence, a Central Action Sequence, a Purpose Fulfillment Test, an optional return to the Central Action if the purpose is unfulfilled; a Finishing Action (if the purpose is fulfilled), and a Consequence of the Action. Other processes include actions that are not purposeful, actions without completions, events with no actor, ongoing states, and so on.

These process schemas characterize the semantics of what linguists call “aspect,” the structure of events and actions as characterized in every language that has been described well enough for us to tell. Via computational modeling with biologically plausible neural networks, Narayanan discovered that Process Schemas had two parts: a general process schema, and specific detail schemata. For example, iterated actions like bouncing a ball and tapping your foot have the same general iteration schema, but one instance that is neurally bound to circuitry for foot tapping and the other to circuitry for ball bouncing (Narayanan, 1997a,b). Narayanan’s process schemas characterize not only the structure but also the logic of aspect, carrying out inferences about events and actions of all types.

What is remarkable is that the same neural circuitry used to run our bodies physically also structures our reasoning processes about all events and actions, not just physical ones, but abstract actions and events as well, such as abstract discourse about international economics (Narayanan, 1997a,b).

4. Metaphorical thought and embodiment

By 1980, in Metaphors We Live By, Mark Johnson and I (Lakoff & Johnson, 1980) had shown that metaphors are conceptual in nature—frame-to-frame mappings across conceptual domains—and that linguistic metaphors are surface reflections of those conceptual mappings. We also showed that many basic metaphors arise from correlations between co-occurring embodied experiences; for example, Happy Is Up, Sad Is Down; More Is Up, Less Is Down; Affection Is Warmth.

These results have great philosophical significance, as discussed at length by Johnson (1987, 1993, 2007). Here are the basic results:

1. Conceptual metaphors are frame-to-frame mappings, with the roles of the source frame mapping to corresponding roles of the target frame. In conceptual metaphors, source and target frame mappings are not necessarily one to one. In some cases not all roles or role fillers are mapped, and in others metaphorical roles are added to the target domain (Lakoff & Johnson, 1980, 1999; Lakoff & Núñez, 2000).
2. Conceptual metaphors “preserve” the structure of cogs: paths map to paths, containers to containers, contact to contact, and so on (Lakoff & Johnson, 1980, 1999).
3. Conceptual metaphors project inferences from the source to the target frame. This is the source of metaphorical reasoning (Lakoff & Johnson, 1980, 1999).
4. Conceptual metaphors can be decomposed into combinations of simpler metaphors and ultimately to “primary” metaphors, which do not decompose further (Grady, 1997; Lakoff & Johnson, 1999).

5. Primary metaphors are motivated by embodied experiences coming together regularly. For example, when children are held affectionately by their parents, the experiences of affection and warmth correlate, yielding Affection Is Warmth (Lakoff & Johnson, 1980, 1999).

6. Primary metaphors tend to be cross-cultural, wherever appropriate experiences regularly co-occur (Grady, 1997; Lakoff & Johnson, 1999).

7. As children, we learn hundreds of primary metaphors, and they structure our systems of everyday thought (Lakoff & Johnson, 1999).

8. Emotion metaphors arise from physical correlates of emotion. For example, in anger, skin temperature and blood pressure rise. Thus, anger can be conceptualized as the heat of a fluid that releases pressure; for example, “his blood was boiling,” “he let off steam” (Lakoff, 1987; Case study 1; Kövecses, 2000).

9. Metaphorical idioms typically come with a conventional mental image. A conceptual metaphor maps knowledge about the image to the target frame, providing the meaning of the idiom (Lakoff, 1987, Case Study 2).

10. Metonymies are mappings from one role to another role within, not across, frames (Lakoff & Johnson, 1980).

11. Conceptual metaphors apply to concepts (e.g., frames and frame elements) not to words in a sentence (Lakoff & Johnson, 1980, 1999).


13. Fixed conceptual metaphors are “alive” in the minds of speakers. They are mostly used unconsciously and automatically in reasoning and word choice (Lakoff & Johnson, 1980).

14. Primary metaphors combine with each other and with frames, forming complex general conceptual metaphors that apply to many specific cases (Lakoff & Johnson, 1999).

15. The specific cases of conceptual metaphors form when specific frames fit the source structure of a general metaphor (Lakoff & Johnson, 1999).

16. Cogs that structure frames and are preserved by metaphor play a central structuring role in grammar (Lakoff, 1987; Case study 3; Dodge, 2010; Goldberg, 1995, 2006).


18. Basic concepts of morality cross-culturally (e.g., purity; uprightness; obedience to authority; nurturance; balancing accounts) arise from primary metaphors linking well-being and ill-being to other embodied experiences (Lakoff, 2008; Lakoff & Johnson, 1999).

19. Metaphors taken literally permeate our legal system (Winter, 2001). Consider Citizens United, in which the Corporations Are Persons metaphor was made into law, affecting political speech and money in political campaigns.

21. Mathematical ideas are built out of embodied cogs, frames, and conceptual metaphors (Lakoff & Núñez, 2000).

Overall, conceptual metaphors structure a huge amount of our mental lives. They are embodied in two ways: via embodied cognitive primitives that structure the frames in frame-to-frame mappings and via the hundreds of primary metaphors that ground human metaphor systems and more complex metaphors in embodied experience.

Via the embodiment of cognitive primitives and primary metaphors abstract (that is, non-physical) concepts become embodied. The embodied frames may characterize abstract ideas, and the embodied metaphors usually do.

With a clearer overview of the results of embodied cognitive linguistics, we can now see that cognitive primitives, frames, and conceptual metaphors are all characterized in NTTL, in terms of neural circuitry as described in the next section.

5. Neural embodiment in NTTL

Jerome Feldman came to the International Computer Science Institute (ICSI) as its founding director in 1988. We formed a research group to study the neural theory of language and thought. Feldman came with a number of basic ideas, which we fleshed out over the next decade and a half. Some ideas may seem similar to many neural and computational theories. But it is non-trivial to integrate them with results from neuroscience, cognitive linguistics, and psychology, and to examine their relationship to conceptual metaphor.

5.2. Functional neural circuits

The brain gets things done not neuron by neuron, but via groups of neurons, called “nodes.” A functional neural circuit contains node-to-node connections. Whereas each neuron either fires or not at a given time, a node may have a percentage of its neurons firing, with different strengths at a given time.

Like a neuron, a node has electro-chemical input and output. Because each neuron has thousands of inputs and outputs, a node has many times that. At each node, subcollections of connections may diverge or converge. Moreover, the nodes typically have some internal structure, with node-internal circuitry.

Our group adopted the ideas that this functional circuitry could, in principle, characterize conceptual structures of the types cognitive linguists had found; and that meaning lies in neural circuitry connected to the body (Regier, 1996). The following ideas illustrate how:

Mental imagery research has shown that much of the same neural circuitry is used in imagining action and perception as in actually acting and perceiving (cf Farah, 1988).

Research on systems of mirror and canonical neurons points to joint action-perception circuitry (cf Gallesse & Lakoff, 2005). Thus, meaning in natural language activates imaginative
circuitry that is the same as for perception and action. This would explain, on neural grounds, why natural language verb roots tend to be the same for first-, second-, and third-person experience. And these results would explain why motor programs, gestalt perception, and mental imagery form basic-level categories of the sort described by Eleanor Rosch (Rosch, 1975a, 1975b, 1977; Rosch et al., 1976; Rosch & Mervis, 1981).

Our group has hypothesized computational models for "binding circuits" that link independent properties as aspects of the same entity—as when shape and color are perceived together as a single entity with both shape and color. Or consider the restaurant frame. It activates together a business frame and a feeding frame, with business frame customer identified as the same as the food frame eater. The same type of binding happens when two primary metaphors form a complex metaphor. Although these examples are fixed bindings, there are dynamic bindings that happen on the fly during communication or imagination.

Research on cascades (DeHaene, 2009) and convergence/divergence zones (Damasio, 1989) indicates that (a) complex imagining, understanding, and acting are not localized in "modules," but rather require activation of an entire cascade; and (b) neural processes associated with experience are a two-way street, initiated both via bodily perception and action, and internally, without external bodily stimulation or with only linguistic stimulation.

Each cognitive primitive occasions inferences. For example, the container schema has an inference of the form: If A Is In B and B Is In C, then A Is In C. Because reasoning is a process like other processes, process schemas can carry out reasoning of this sort. In NTTL, the relation "A Is In B" is a neural binding between the Interior of a Container Schema (a node we are calling "B") and an Entity of any sort (a node we are calling "A").

The process schema is a sequence of three nodes. In honor of Aristotle, we call this process schema the "Syllogism Schema" and the sequences of nodes, "Major Premise," "Minor Premise," and "Conclusion." Each is bound to an instance of the Is In relation—Major Premise: A Is In B, Minor Premise: B Is In C, and Conclusion: A Is In C. Where we have used the same letter in this notation (e.g., the "B" in "A Is In B" and "B Is In C"), NTTL posits a neural binding between the two nodes notated by the letter "B."

In short, what have been seen in formal logic as axioms or postulates are understood in NTTL as arising from neural binding that links nodes in a process schema like that above. In general, syllogistic reasoning with cognitive primitives is understood in NTTL as making use of neural circuitry in this way. There are a few basic general types of syllogistic schemas, and hundreds of instances where specific cognitive primitives are neurally bound to the premise and conclusion nodes in those schemas. Syllogistic reasoning of this sort is real and carried out by neural circuitry. The hypothesized circuitry could carry out such reasoning, which is, of course, mostly unconscious, automatic, and so fast it often seems instantaneous.

Neural theory of thought and language assumes spreading activation and inhibition along existing neural pathways, and the concept of "best-fit." When there is more than one possibility for spreading activation or inhibition, the pathway requiring the least energy is used. Presumably, that will be the one with the greatest overall synaptic strengths. What is activated nearby also influences overall synaptic strengths, and thus also "guides" spreading activation.
These considerations are central to the idea of neural simulation. Fixed neural structures and their fixed inferences are activated or inhibited (say, by language), and that spreads according to existing pathways and best-fit. This constitutes neural simulation. Context-dependent inferences arise when context influences simulation.

Learning in NTTL comes from the following: (a) Hebbian learning. Neural firing results in synaptic strengthening. “Neurons that fire together wire together.” (b) Spike-time-dependent plasticity (STDP) (Song, Miller, & Abbott, 2000). Whichever of two synapsing neurons regularly spikes first has synaptic strength increased in its direction and synaptic strength decreased in the opposite direction. (c) Recruitment learning. Of the hundreds of trillions, or more, possible neural circuits, a circuit is turned into a functional circuit when its synapses are strengthened.

In NTTL, Gilles Fauconnier’s linguistic theory of mental spaces works by “partitioned simulations.” Each “mental space” is a simulation. Multiple spaces are multiple simulations. And identity connections across spaces are neural bindings.

Understanding in natural language is accomplished via mental simulations using embodied functional circuitry of the kind just described. Moreover, much of this involves topographic maps: neuronal groups that are linked together via circuitry that preserves closeness. Topographic maps are ultimately connected to the body.

This model also provides a basis for determining which circuit contributes meaning to another when two circuits bind. Conceptual structures include Schemas with Roles and Sequences. A Schema is a gestalt circuit with a Control Node and Constituent Nodes with the following constraints. Each Constituent node activates the Control node, and the Control node activates all the Constituent nodes. In a Sequence, each preceding node activates the following node, which then inhibits the preceding node. There are also simple circuits for comparison of action levels, branching, combining, and so on. Schema-to-schema mappings contain activation circuits, in which activation flows from the Roles in the Source Schema to corresponding roles in the Target Schema.

Primary metaphors arise in the following way. When two embodied brain regions are simultaneously activated, repeatedly, Hebbian learning guides synaptic strengthening, linking the two areas. At this point, STDP occurs strengthening the connections in one direction and weakening those in the other direction, thus providing an asymmetric activation pattern. The overall amount of coherent activation coming from the two connected regions determines the direction of strengthening, and thus, which is the Source versus Target domain (Narayanan, unpublished data). Thus, we get Affection Is Warmth, not Warmth Is Affection, because the brain is always computing temperature, but not always computing affection. This process starts before language is learned. It happens unconsciously, just by encountering phenomena that occur together.

Incidentally, NTTL contains a theory of Embodied Construction Grammar that uses the same mechanisms to characterize the grammar of words and constructions (Bryant, 2008; Chang, 2009; Dodge, 2010; Feldman, Dodge, & Bryant, 2009). But we will not take that up here.

The empirical research behind this theory of functional neural circuitry comes from three sources, each with an extensive technical literature in traditional areas: (a) Embodied
Cognitive Linguistics, which generalizes over a huge range of examples so as to characterize the cognitive structure and embodied meaning of linguistic expressions. (b) Embodied neural modeling, which suggests appropriate types of computational structure and processing in functional neural circuits (Feldman, 2008). (c) Experimental research, which tests hypotheses and uncovers phenomena to be explained via functional neural circuitry. Because of disciplinary boundaries, the relevant publications tend to be in cognitive linguistics, computer science, or experimental psychology. The study of functional neural circuitry is not yet a part of any existing disciplinary institution. This essay is a first attempt to describe the overall set of ideas, and what they can explain. It brings neuroscience, computer science, linguistics, and experimental psychology together via the study of functional neural circuitry.

6. Explaining experimental results on embodied cognition

Over the past couple of decades, experimental social psychologists have provided massive evidence not only for the existence of fixed metaphorical brain circuits, but also for their effects on producing and understanding social behavior. Many of the experiments cited in this volume are of this sort.

Experimental psychologists associated with cognitive linguistics began doing research on embodiment in this tradition in the early 1980s. The first major book was by Raymond W. Gibbs, Jr., *The poetics of mind* (Gibbs, 1984). Gibbs’s research showed experimentally that predictions made by metaphor analyses by Lakoff, Johnson, and Kövecses indeed held. See Gibbs (2006) for a broader survey. Gibbs’s student, Teenie Matlock, produced perhaps the most dramatic confirmation of embodiment in cognitive linguistics (Matlock, 2004). She demonstrated that subjects actually traced metaphorical fictive motion sentences (as in “The road runs along the cliffs above the ocean”) in real time via mental simulation.

Lera Boroditsky has been a central figure as well. For example, with early research regarding two opposing primary conceptual metaphors for time in terms of space, discussed in *Metaphors we live by* (Lakoff & Johnson, 1980): one in which times are entities moving toward you and passing you, the other in which times are locations along a path and you are moving toward them. As a result, sentences like “Let’s move the meeting ahead two days” are ambiguous depending on which metaphor you are thinking in terms of. If times move toward you, a meeting originally on Wednesday, would now meet Monday. If you are moving toward times, the Wednesday meeting now meets Friday. Subjects primed with the source domain of one of these metaphors move the meeting in the predicted direction, given the metaphorically ambiguous sentence (Boroditsky, 2000; Casasanto & Boroditsky, 2008).

This brings us to the body of experiments on embodied cognition discussed in this volume. Here are some typical examples.

A recent study in *Biological Psychology* showed that when subjects leaned forward the body posture activated desire. The primary metaphor is Achieving a Purpose (Desire) Is Reaching a Destination. Leaning forward activates motion to a destination, the source domain, which in turn activates the target domain of desire and purpose (Harmon-Jones, Gable, & Price, 2011).
At Yale, researchers found that subjects holding warm coffee in advance were more likely to evaluate an imaginary individual as warm and friendly than those holding cold coffee. This is predicted by the primary conceptual metaphor that Affection Is Warmth, as in She gave me a warm greeting (Williams and Bargh, 2008).

At Toronto, subjects were asked to remember a time when they were either socially accepted or socially snubbed. Those with warm memories of acceptance judged the room to be 5°C warmer on the average than those who remembered being coldly snubbed. Another effect of Affection Is Warmth (Zhong & Leonardelli, 2008).

Subjects asked to think about a moral transgression like adultery or cheating on a test were more likely to request an antiseptic cloth afterwards than those who had thought about good deeds. The conceptual metaphor Morality is Purity predicts this (Zhong & Liljenquist, 2006).

Students told that a book was important judged it to be physically heavier than one they were told was unimportant. The conceptual metaphor is Important is Heavy. In a parallel study, those with heavy versus light clipboards were more likely like to judge currency to be more valuable and their opinions and their leaders more important (Jostmann, Lakens, & Schubert, 2009).

7. Explanations

Such experimental results can readily be explained in NTTL. We have a myriad of conceptual metaphor circuits fixed in our brains. In each of the above examples, the source and target domain frames of a metaphor are activated in a context.

Why does this happen? Because primary conceptual metaphors are persistent (long-lasting or permanent) physical circuits in the brain. In each case, the metaphor circuit is activated by the context of the experiment, which in turn activates circuitry that can not only simulate the observed behavior but actually carry it out unconsciously. The reason is that they are two-way cascades. Hence, metaphor circuits can be activated by physical experience, and then the same circuitry can simulate and unconsciously control physical actions. Activating primary metaphor circuitry primes the circuits for carrying out the behavior, making the behavior more likely. In short, we really do live by metaphor—at least primary metaphor.

Primary metaphors differ from composed metaphors. Primary metaphors directly link two kinds of embodied experiences. That is why activating them can prime actual action performance. Composed metaphors bring together two or more primaries, each with different pairs of embodied experiences. With four or more regions activated, each is activated less, and thus less likely to serve as a prime.

The thing to remember about embodied cognition experiments is that, as beautiful as they are, they need to be explained in neural terms—and they can be. At present, NTTL is the only game in town providing neural explanations via direct links between conceptual metaphor, neural models, and experimental results.
References


