Report: January, 2009. An earlier version appeared in:
R. Gibbs. 2008 *The Metaphor Handbook*, Cambridge University Press.

The Neural Theory of Metaphor

George Lakoff

University of California at Berkeley

The neural revolution is changing our understanding of the brain and the mind in radical ways, and that is no less true in the theory of metaphor. It is 30 years since Mark Johnson and I wrote *Metaphors We Live By* in 1979. Though the fundamental outlines of what we discovered remain as valid today as they were then, developments in brain science and neural computation have vastly enriched our understanding of how conceptual metaphor works. This is an intermediate report, as of January 2009.

You may well ask why anyone interested in metaphor should care about the brain and neural computation. The reason is that that what we have learned about the brain explains an awful lot about the properties of metaphor. For example, have you ever asked why conceptual metaphor exists at all, why we should think metaphorically, why metaphors should take the form of cross-domain mappings? Have you thought about how our metaphor system is grounded in experience, or about why certain conceptual metaphors are widespread around the world or even universal? Have you wondered about how complex poetic metaphors are built up out of simpler metaphors? Have you wondered about how whole systems of philosophical or mathematical thought can be built up out of conceptual metaphors. The neural theory explains all this.

It explains more as well: Why metaphorical language should take no longer to process than non-metaphorical language. Why some sentences of the form *X is Y* make sense as metaphors and why others fail. How conceptual metaphors can play a role in abstract concepts. These and other wondrous properties of conceptual metaphors fall out once one considers metaphor theory from the perspective of the brain.

In 1988, Jerome Feldman came to Berkeley as director of the International Computer Science Institute, and he and I formed the NTL (Neural Theory of Language) group. Feldman is one of the founders of the theory of neural computation, and we have been working together since then. Feldman's landmark book, *From Molecule to Metaphor* surveys much of the work of our group, and is a must-read for metaphor theorists. As a background both to reading that book and to our discussion of metaphor, I offer a brief and overly simple introduction to the neural theory of language (NTL).

A Brief Introduction to NTL

Every action our body performs is controlled by our brains, and every input from the external world is made sense of by our brains. We think with our brains. There is no other choice. Thought is physical. Ideas and the concepts that make them up are physically "computed" by brain structures. Reasoning is the activation of certain neuronal groups in the brain given prior activation of other neuronal groups. Everything we know, we know by virtue of our brains. Our physical brains make possible our concepts and ideas, everything we can possibly think is made possible and greatly limited by the nature of our brains. There is still a great deal to be learned about how the brain computes the mind. NTL combines what is known scientifically with linking hypotheses based on neural computation.

The Shaping of the Brain

We are born with an enormously complex brain with hundreds of precisely and beautifully structured regions and highly specific connectivity from every region to many other regions.

Each neuron has connections to between 1,000 and 10,000 other neurons. Between birth and the age of five, roughly half of the neural connections we are born with die off. The ones that are used stay; the others die. That is how are brain is shaped, and such a shaping is necessary if the brain is to learn to do the huge number of things it does.

The flow of neural activity is a flow of ions that occurs across synapses — tiny gaps between neurons. Those synapses where there is a lot of activity are "strengthened" — both the transmitting and receiving side of active synapses become more efficient.

Flow across the synapses is relatively slow compared to the speed of computers: about 5 one-thousandths of a second (5 milliseconds) per synapse. A word recognition task — Is the following word a word of English? — takes about half a second (500 milliseconds). This means that word recognition must be done in about 100 sequential steps. Since so much goes into word recognition, it is clear that much of the brain's processing must be in parallel, not in sequence. This timing result also shows that well-learned tasks are carried out by direct connections.

Neuronal Groups

Jerome Feldman and colleagues, in the 1970's, developed an account of "structured connectionism" — *not* PDP connectionism! In PDP connectionism, where all computation is distributed over an entire network and nothing is "localized" — that is, no meaning or function can be assigned to any single neuron or any small collection of neurons in the network. Only very restricted parts of the brain work that way.

On the other hand, structured connectionism takes into account the local structure that exists in the brain. Neuronal groups (of size, say, between, 10 and 100 neurons) are modeled as "nodes" which are meaningful and which enter into neural computation. Since each neuron can have between 1,000 and 10,000 neural connections, nodes can "overlap," That is, the same neuron can be functioning in different neuronal groups, or "nodes." The firing of that neuron contributes to the activation of the each node it is functioning in. Though single neurons either fire or not, neuronal groups contain neurons that fire at different times, making the group active to a degree depending on the proportion firing at a given time.

The modeling of neural computation is done over networks with nodes, connections, degrees of synaptic strength, and time lapses at synapses.

Embodiment and Simulation Semantics

The link between body and brain is central to the concept of semantics-as-simulation in NTL. Suppose you imagine, remember, or dream of performing certain movements. Many of the same neurons are firing as when you actually perform that movement. And suppose you imagine, remember, or dream of seeing or hearing something. Many of the same neurons are firing as when you actually see or hear that thing.

Mirror neurons occur in fiber bundles connecting pre-motor/SMA cortex (which choreographs actions) with the parietal cortex (which integrates perceptions). The same mirror neurons fire when you perform an action or you see someone else performing that action. The mirror neurons are thus "multi-modal," that is, they are active not only when acting or perceiving the same action, but when imagining that you are perceiving or performing an action. Now a word like, "grasp," applies both to performing and perceiving grasping — that is, it is multimodal.

Simulation semantics is based on a simple observation of Feldman's: if you cannot imagine someone picking up a glass, you can't understand the meaning of "Someone picked up a glass." Feldman argues that, for meanings of physical concepts, *meaning is mental simulation* — that is, the activation of the neurons needed to imagine perceiving or performing an action. Thus, all mental simulation is embodied, since it uses the same neural substrate used for action, perception, emotion, etc.

One thing we know is that not all imagination or memory is conscious, and so not all mental simulations are. That is why we typically have no conscious awareness of most such simulations.

A *Meaningful Node* is a node that, when activated, results in the activation of a whole neural simulation, and that when inhibited inhibits that simulation. *Inferences* occur when the activation of one meaningful node, or more, results in the activation of another meaningful node.

NTL, following the theory of simulation semantics, suggests that the neural circuitry characterizing the meaning of "grasp" is the neural circuitry in the mirror neurons that are activated when imagining either performing or perceiving grasping.

The meaning of concrete concepts is directly embodied in this manner. There is now considerable evidence that perceiving language activates corresponding motor or perceptual areas. For example, *He kicked the ball* activates the foot area of the primary motor cortex.

Activation and Inhibition

A flow of ions across a synapse may either contribute to the firing of the post-synaptic neuron, or may help to inhibit such firing — depending on whether the charges of the ions are positive or negative. The activation of neural simulations constitutes meaningful thought.

We obviously don't think all possible thoughts at once. Indeed, most possible thoughts are either unactivated or *positively inhibited* most of the time.

Mutual Inhibition

Two neuronal groups can be connected so that each inhibits the activation of the other when there is an active flow of ions of the opposite charge. This is called *mutual inhibition.* This occurs, for example, when there are two inconsistent, but equally available, ways of looking at a situation.

This is common in politics, where a strict conservative worldview is typically inconsistent with a nurturant progressive worldview. That is, they are mutually inhibitory. But many people have both worldviews active in different areas of their lives, and can think of a given situation first from one worldview and then from the other. When one is activated, the other is inhibited.

Spreading activation: Neurons that fire together wire together

Spreading activation at the behavioral level has been the mainstay of psycholinguistics for decades - NTL models link this behavior to neural structure. When two neuronal groups, A and B, fire at the same time, activation spreads outward along the network links connecting them, which we experience as a chain of thought.

During learning, spreading activation strengthens synapses along the way. When the activation spreading from A meets the activation spreading from B, a circuit may be formed. The more A and B fire together the stronger the synapses forming links in circuit. As we shall see, this the basic mechanism by which primary metaphors are formed.

Neural maps

We are born with neural circuitry that effectively activates a "map" of one part of the brain in another part of the brain. For example, the 100 million neurons coming out of the retina grow connections before birth from the retina to other areas including the primary visual cortex at the back of the brain. These connections form a "topographic map" of the retina in V1. That is, the connections preserve topology (relative nearness), though not absolute orientation or absolute distance. When neurons next to each other coming from the retina fire, the corresponding neurons fire in V1 and are next to each other in V1.

Len Talmy (2000) has observed that spatial relations in human languages preserve topology as well. For example, containers remain containers no matter how their boundaries are stretched or contracted, and paths remain paths, no matter how they wind around. Terry Regier (1997) has constructed computational neural models of topographical maps of the visual field that can compute image schemas with topological properties, and accurately learn the words for a nontrivial range of spatial relations in a variety of languages.

Neural Binding

Imagine a blue square. We know that color and shape are not computed in the same place in the brain: they are computed in quite different areas. Yet the blue square appears to us as a single whole — not as separate squareness and blueness. The name given to this phenomenon is "neural binding." Neural binding is responsible for two or more different conceptual or perceptual entities being considered a single entity.

There are three types of neural bindings:

(1) Permanent obligatory bindings, e.g., in your stored mental image of a parrot, the feathers are green. There is a permanent obligatory binding in the neural representation for the parrot image, between the neuronal groups that characterize feather-shapes and those, elsewhere in the brain, that characterize the green color.
(2) Permanently-ready-but-conditional bindings, like the bindings in the neural structure for an election-night map on which any given state can be either red or blue depending on the outcome of the vote.
(3) Nonce bindings that occur on the fly as they happen to arise in context.

It is not known just how neural binding operates in the brain. One hypothesis is that neural binding is the synchronous firing of nodes. Lokendra Shastri has modeled the computational structure necessary to carry out binding in such a theory.

Neural Choreography

In general, the premotor cortex and supplementary motor area (SMA) choreograph specific actions, like grasping. Grasping has a neural structure of its own. There are, in addition, neural connections between the pre-motor/SMA and the primary motor cortex – M1. M1 is laid out topographically according to the neurons as they are connected to the body. For example, neurons connected to the hand are in the same region of M1, with neurons connected to the index finger next to neurons connected to the middle finger. The whole body is topographically connected to the neurons in M1.

Each M1 neuronal group can perform only a simple action, like opening the elbow or pointing the index finger. In order to pick up a bottle, those simple M1 actions must be sequenced and choreographed. The premotor cortex/SMA does the choreography, having learned neural circuits that fire in complex sequential patterns. As each premotor/SMA neuron fires, a connection to M1 makes the right M1 neurons fire, which in turn moves certain muscle groups in the body. Picking up a bottle is like an exquisite ballet with choreographic instructions being carried by the connections to the neurons in M1, which individually control each little movement.

When the bindings are in place, the Premotor/SMA Circuitry + bindings + PrimaryMotor Circuitry acts seamlessly like a single simple circuit.

Circuit Types

NTL modeling assumes that, as our neural circuitry is being shaped by experience, certain relatively simple basic types of neural circuits emerge, as follows. The research includes ways in which circuits with these properties can be formed.

What is important for the study of thought is not the study of precise neural circuitry, but rather the study of *the kinds of computations* that neural circuitry can carry out. An important topic in the neural theory of language is exactly what kinds of circuit types are necessary for human thought — for frames, image-schemas, conceptual metaphor, lexical items, grammatical constructions, and so on.

Neural bindings play a crucial role, forming complex circuits by binding nodes in one circuit type to nodes in another circuit type.

The winner-take-all circuit:

- Two or more subcircuits, say A and B, with mutually inhibiting connections between them.
- When A is firing B cannot fire, and conversely.

Winner take all circuits apply, for example, to high-level "worldview" circuits that make sense in a single way of a wide range of experiences — in politics these might be conservative and progressive worldviews. You might understand a range of experiences using one world-view or the other, but not both at once.

A Note on Gestalt Nodes

What I will call a "gestalt node" is a subcollection of neurons forming a subcircuit within a given circuit. The gestalt node acts like a traffic cop, directing activation within the circuit and directing activation into and out of the circuit. This allows the gestalt node to form a computational "unit" out of the larger circuit that it is part of; e.g., a "unit" like a frame, an event sequence, a metaphor, etc.

Though all the neurons in the brain are ultimately linked to all other neurons by some pathways or other, or by many pathways, that does not mean that everything activates everything else. The existence of such traffic cop circuits allows the circuitry of the brain to do a vast number of special jobs.

A Schema Circuit:

- A collection of nodes, say, A, B, C, and D and a “gestalt node” G.
- When G is firing, all of A, B, C, and D fire.
- When a sufficient set of A, B, C, or D is firing, G fires, which results in all other nodes firing. One especially salient node can be sufficient in some cases, or there can be a threshold where the total activation summed over all the nodes is above G’s threshold and results in G firing.
- When G is inhibited, at least one of the other nodes is inhibited.

Schema Circuits characterize the structure of frames. Frames are special cases of schemas. The entire frame corresponds to G and the roles correspond to A, B, C, D, etc.. Schemas circuits also characterize the circuitry that allows image-schemas and X-schemas to function as schemas.

In a gestalt, the whole is more than just the sum of its parts. Accordingly, in a schema circuit, the whole — G — cannot be inhibited and all of its parts activated. The activation of even some of the salient parts activates the whole. And the activation of the whole activates all the parts.

A Linking Circuit:

- Two nodes, A1 and A2, a circuit L connecting A1 with A2, and G, the gestalt node of circuit L.
- When A1 and G are firing, A2 is firing. But when A2 is firing, A1 need not be firing. Thus, activation flow is asymmetric from A1 to A2.
- When A1 is firing and G is not, the linking circuit L is not active. (That is, G “gates” the connection L.)
- When A1 and A2 are both firing, gestalt node G is firing and the linking circuit L is active.

Note: A1 can fire without A2 firing. This can happen if G is inhibited from firing, thus making the link L from A1 to A2 inactive. Think of G as a traffic cop regulating the flow from A1 to A2.

Linking Circuits are used in metonymy: Within a frame F, one semantic role A may “stand for” another B. A metonymy is characterized by:

(1) A schema consisting of gestalt node F and at least nodes A, B, and X;
(2) A connection L linking A to B asymmetrically, with gestalt node G gated by X. That is, G fires only if X is firing.

For example, in *The ham sandwich wants his check*, the frame F is the restaurant frame, *the ham sandwich* plays the role Dish (A), *his* refers back to the entity that plays the role Customer (B), and L characterizes the metonymic link from the Dish (A) to the Customer (B), while X is the condition that the waiter/waitress identifies the Customer B primarily in terms of the Dish B.

Binding circuits:

When neural binding occurs, the nodes bound together act as if they characterize the same entity: every circuit activated by one is activated by the other, and every circuit activating one activates the other.

A common theory of binding holds that the nodes that are neurally bound fire in synch. There are other theories, and none is accepted as having been proven.

ID Links:

You are the same person you were as an infant, though you certainly have changed. That change means that circuits characterizing you as an infant must not be the same as those characterizing you now. One suggestion for how a neural system accomplishes this uses the concept of an *Essence* as a semantic role of the frame for an entity. On this hypothesis, we can characterize an ID Link.

An ID Link between A and B consists of:

(1) a linking circuit from the circuit for one entity A to the circuit for another entity B;
(2) a neural binding identifying the Essence role of A and the Essence role of B.

ID links characterize connectors across mental spaces, identifying A in one space as the same entity as B in another space, even though they may have different properties in different spaces.

Two-way linking circuits:

A two-way circuit linking nodes A1 and A2 is comprised of two opposite one-way linking circuits, with a gestalt node creating a unit from the two linking circuits. Here are the properties of two-way linking circuits:

- Nodes A1 and A2. Linking circuits L1 and L2. Gestalt nodes G1 and G2. Gestalt node G.
- First linking circuit: From A1 to A2 via linking circuit L1, with activity directed by gestalt node G1.
- Second linking circuit: From A2 to A1 via linking circuit L2, with activity directed by gestalt node G2.
- The overall gestalt circuit: Linking circuits L1 and L2 with gestalt node G.
- When G is activated, both links are activated. When G is inhibited, both links are inhibited.

Two way-linking circuits provide the kinds of connectivity used in grammatical constructions and lexical items, where there is a two-way connection between a lexical meaning and a lexical form, or a grammatical meaning and a grammatical form structure.

Mappings

Mappings are unidirectional multiple linking circuits, each governed by a single gestalt node, and with those linking gestalt nodes governed altogether by a single gestalt node. Here are some examples:

Metaphor

Metaphor mappings apply to schemas of all sorts, whether image-schemas, frames, action sequences, or narratives. They map "source domain" schemas and their roles to "target domain" schemas and their corresponding roles.

Metaphorical Mapping:

Two schemas S1 with roles A1, B1, …, and S2 with roles A2, B2, … .
Linking circuits LS, LA, LB, that respectively link schema S1 to Schema S2, role A1 to role A2, role B1 to role B2, and so on.
Gestalt nodes GS, GA, GB, … governing linking circuits LS, LA, LB, … .
Gestalt node G governing GS, GA, GB,... .

When G is inhibited, GS, GA, GB, etc.. are all inhibited and activation is shit off in all the links.

When G is activated, DS, GA, GB, etc. are all activated, and activation flows from Schema node S1 to Schema node S2 and through all the links between roles.

When Schema 1 and Schema 2 are activated, G is activated.

Mental Space Mapping:

A "mental space" M consists of circuitry used to run a simulation activated by a Semantic Structure.
A Semantic Structure SS consists of the circuitry characterizing such semantic "units" as frames, image schemas, X-schemas, metaphorical maps, metonymic maps, blending maps, etc. together with neural bindings integrating them into a whole.
A cross-space relation R consists of two mental spaces M1 and M2 and a semantic structure SS with roles bound to elements of each space.
A mental space mapping MSMap consists of:
A linking circuit L from M1 to M2.
A collection of ID Links from elements EM1[i] of M1 to elements of EM2[i] of M2.
A gestalt node G governing L and all the ID Links.

A "space builder" is a linguistic element or construction B that activates R and G.

For example, take the sentence *If Clinton had been President of France, there would have been no scandal over his affair.* The mental spaces are: M1 = The US during Clinton's presidency with EM1(1) = Clinton and SS1 = his affair in the US, and M2 = France at that time, EM2(1) = A Clinton-correlate and the role played by EM2(1) is the President of France who has an affair in France with no scandal; L1 is the circuit that identifies A1 (the real Clinton) with A2 (the Clinton correlate ≠ Clinton). The SS indicates that M1 is taken as fitting reality and M2 is not.

Extension Circuit:

- A semantic structure SS1, containing nodes A1, B1, C1, D1,… and a gestalt node G1 governing SS1.
- Gestalt node G2 governing nodes C2, D2, … .
- A Linking Circuit L with gestalt node G linking G2 asymmetrically to G1.
- Mutually inhibitory links between C2, D2, …, and C1, D1, … respectively.
- SS2 consisting of SS1 with C2, D2, … replacing C1, D1,… .

When G is firing, SS2 is activated and SS1 is inhibited.
When G1 is firing, SS1 is activated and SS2 is inhibited.

Extension Circuits characterize radial categories (See Lakoff, 1987, Case Study 3). For example, suppose SS1 characterizes the concept of mother and SS2 characterize the concept of stepmother, where the birth frame of mother is inhibited while the marriage frame remains active.

X-schema circuit:

- A gestalt node
- State nodes
- Action nodes
- Connections, both activating and inhibiting
- Conditional Choice nodes
- Timing nodes

X-schemas, or "executing schemas," do things, via bindings that activate other circuits. Every action node is preceded and followed by a state node, with activation spreading from states to actions to states. Timing nodes coordinate the lengths of states and actions (which may be instantaneous or elongated). Iterated actions are formed by loops from the state following an action to the state preceding the action. Conditional actions are formed by gatings — cases where activations from both nodes A and A' are needed to activate node B. Conditional Choice nodes have outputs going to two or more other nodes, with gatings that determine the choices, perhaps probabilistically.

The gestalt node activates the initial state and the final state inhibits the gestalt node. Actions typically have initial and final states, initiating and concluding actions, central actions, and may have purposes. A purposive action is one with a desired state. The purpose is met if the desired state is active after the central action, and if so, the action is concluded. Each action can be neurally bound to the gestalt node of another complex X-schema, to produce quite complex actions.

X-schemas characterize the structures of states and actions, referred to as “aspect” in linguistics. Aspects can be durative or instantaneous, stative or active, completive or open-ended, iterative or non-iterative.

When connected to the body via the primary motor cortex, pre-motor/SMA X-schemas can carry out actions. X-schemas can also define scenarios within frames or narratives and carry out chains of reasoning, by sequentially activating mental simulations.

Blending Maps:

Blending maps are complex maps: They take two or more sources and map them to a single target. The source has two or more gestalts, G1, …, Gn. The target has a single gestalt, H. The linking circuits consist of either neural bindings or identification links (with bindings of essences).

Examples will be given below.

Conceptual Blends

Conceptual blends are carried out by blending maps, where the crucial circuitry is neural binding circuitry. We will discuss this further below. In short, blending can be seen as a form of complex binding.

The point of these characterizations of circuit types is that, in NTL, one has to be explicit about the computational properties of neural circuitry. Any cognitive analysis must be able to be carried out by the brain and by the relatively simple circuit types of this sort. As we shall see, different mental operations require different types of neural circuitry that performed very specific neural computations.

Neural Systems Are Best-Fit Systems

It is a common cognitive phenomenon that a fact that fits an overall conceptual organization is remembered better than a fact either in isolation or one that contradicts an overall conceptual organization. Ideas make sense when they fit a whole system of ideas.

Similarly, a linguistic compound makes sense when it fits into a coherent context. Take the classic example of “pumpkin bus” — coined on a school outing. There were two buses and the road home passed a pumpkin patch. One of the buses was designated to stop there for students who wanted to buy a pumpkin. It was called the “pumpkin bus,” and the compound was instantly understandable because it fit the context.”

Compare two sentences: “Bill drank a soda” and “Bill drank an elephant.” To get the meaning of the sentences, you need to do a mental simulation, in which Bill is drinking and a frame is activated in which a soda is bound to the patient role in the frame of drinking, which requires that it be a liquid and consumable, which it is. In “Bill drank an elephant,” again the drink frame requires a consumable liquid. Since an elephant is neither — binding the concept of an elephant to the patient’s role in the drink scenarios runs up against neural inhibition. However, context may change things. *Elephant* is a brand of Danish beer, and so the sentence may refer to Danish drinking experience. Or second, one could imagine a context in which an elephant was sacrificed by being cut up and put in a blender and liquefied so that one could drink it.

What determines “fit”? Maximizing the number of overall neural bindings — including context and overall knowledge — without contradiction, that is, without encountering any mutual inhibition. By maximizing bindings you engage as much as possible of what is already in the brain. And making maximal use of strong synaptic weights, that is, circuitry that has the highest prior probability of being activated.

Image schemas and Cogs

Terry Regier (1997) has constructed a neural computational model for how a range of spatial relations concepts can be computed by a neural network that shares certain propoerties with human brains. Narayanan (1997) has constructed a neural computational model of the structure of events, that is, X-schemas. Dodge and Lakoff (2006) have speculated on many of the details involved. Gallese and Lakoff (2005) have shown that certain action circuitry has the structure of frames. They have further speculated that the meanings of grammatical elements and constructions are characterized by “Cogs,” that is, secondary neural structures (e.g., pre-motor/SMA cortex) that bind to structures in primary cortex, e.g., motor and visual. This would explain why grammatical meanings are “abstract” in the sense that they have a very general structure but lack specific details.

We are now ready to discuss how all of this changes old metaphor theory into the neural theory of metaphor.

The Old Theory

Metaphors We Live By was written back in 1979, before the era of brain science and neural computation. Nonetheless, certain results from that era have stood the test of time:

- Metaphors are conceptual mappings; they are part of the conceptual system and not mere linguistic expressions.
- There is a huge system of fixed, conventional metaphorical mappings.
- The system exists physically in our brains.
- Certain metaphors are grounded via correlations in embodied experience (e.g., More Is Up is grounded via the correlation between quantity and verticality — you pour more water in the glass and the level goes up).

- Metaphorical mappings are typically across conceptual domains (as in Affection Is Warmth).
- Mappings (as in A Competition Is a Race) may also be from a specific case (a race) to a more general case (a competition).
- Mappings operate on source domain frame and image-schema structure.
- Via metaphorical mappings, source domain structures (image-schema and frame structures) are used for reasoning about the target domain. Indeed, much of our reasoning makes use of conceptual metaphors.
- Metaphorical mappings are partial.
- Metaphorical language makes use of conceptual metaphors.
- Many different linguistic expressions can express some aspect of the same metaphor.
- A conceptual metaphor may be used in understanding a word, even if that word is not realized in the source domain of the metaphor.
- Most conceptual metaphors are part of the cognitive unconscious, and are learned and used automatically without awareness.
- Novel metaphorical language makes use of the existing system of conventional metaphors.
- We commonly take our conceptual metaphors as defining reality, and live according to them.
- Target domain entities and target domain predications can result from metaphors.
- Two of the relevant sources of data are generalizations over inference patterns (in the source and target domains) and generalizations over lexical items (that can be used of both source and target domains).

These results will be familiar to any student of conceptual metaphor.

To those who have read *The Contemporary Theory of Metaphor* another result that has stood the test of time will be familiar:

- Complex metaphors are made up of simple metaphors bound together and/or bound to commonplace frames.

For example, Love is a Journey is composed of such conceptual metaphors as:

Purposes are Destinations
Difficulties are Impediments to Motion
A Relationship is a Container
Intimacy is Closeness

plus commonplace literal frame-based knowledge that:

A Vehicle is an Instrument for Travel,
A Vehicle is a container in which the travelers are close together,
People are expected to have life goals,
Lovers ideally have compatible life goals.

These are put together in such a way that:

The life goals are destinations;
The lovers are travelers trying to reach those destinations;
Their relationship is a vehicle such that the lovers are *in* the relationship
They are *close*; and
The relationship (when working) helps them achieve life goals; and
The relationship difficulties are impediments to motion (e.g., a long, dusty road; being on the rocks or off the track).

Such compositional structures were noticed during the 1980's. But such descriptions hid certain neural realities. Each frame role is a node. The metaphor mappings are physical circuits—linking circuits that, when activated, form links within integrated neural circuits. "Nodes" are part of such circuits. Each frame is a circuit, and each role in a frame is node in that circuit. Thus, there is a Travel frame circuit and Destination is a node in that circuit. When we write down, as we just have,

The life goals are destinations
The lovers are travelers trying to reach those destinations

We have written down the word destination twice, but in the brain both uses of the word "destination" are activating the same node in the same frame circuitry. Thus, activating a metaphor activates very complex integrated brain circuitry, and a given node may occur in many circuits made active by the metaphor. The more integrated the circuitry activated by the metaphor mapping, the better the "fit" of the metaphor to other brain structures. Thus, the life goals of the ideal lovers are special cases of, and thus fit, the life goals that are understood as destinations. A vehicle used for travel is typically a container, which is the same as the container in the metaphor *A Relationship is A Container.*

A metaphor mapping is a complex circuit which, when activated, activates many other circuits via linking and binding circuitry. This makes possible metaphorical inferences: Source domain inferences that are mapped combine with target domain knowledge via binding to produce new inferences: If lovers are "stuck" in relationship, if the relationship isn't "going anywhere," then they are not making progress toward common life goals. If the lovers are "going in different directions," then they may not be able get to the same destinations, which means metaphorically that their common life goals may be inconsistent.

The NTL perspective provides a very different way of thinking about such complex metaphors. The brain is a best-fit system. Inferences are new activations that arise via prior activations. Bindings and linkings form integrated circuits that result in activations that arise via prior activations. This bindings and linkings give rise to inferences.

Consider our existing conceptual system where:

(1) *A Relationship is a Container,*
(2) A Vehicle is a Container in which the Travelers are close together,
(3) Intimacy is Closeness,
(4) Lovers are intimate,
(5) *A Vehicle is an Instrument for Travel*,
(6) and Lovers are Travelers.

Container is (1) names the same neural structure as Container in (2).
Closeness in (2) names the same neural structure as Closeness in (3)
Intimacy in (3) names the same neural structure as Intimacy in (4)
Vehicle in (5) names the same neural structure as Vehicle in (2)
Lovers in (4) names the same neural structure as Lovers in (6)
Travelers in (6) names the same neural structure as Travelers in (2)

The result is a very tightly integrated neural system—a system that fits together very well. The Love Is a Journey metaphor mapping fits our knowledge, including primary metaphors like (1) and (3) and commonplace knowledge like (2), (4) and (5).

Primary Metaphors

The neural theory of metaphor got its real impetus from three Berkeley dissertations done in 1997 — by Srini Narayanan, Joe Grady, and Christopher Johnson. Narayanan's dissertation was key. He modeled metaphors as neural mappings and formulated certain metaphors for international economics. He then showed that the results of source domain inferences from the domain of physical motion and action are mapped onto the international economics target domain, interact with the logic of the target domain and produce metaphorical inferences.

Johnson studied metaphor acquisition in young children and found 3 stages:
Stage (1) source domain only;
Stage (2) experiences where the source and target domains are both active ("conflated"); at this stage, children learn to use source domain words with target domain meanings and grammar, then later
Stage (3) children use the words metaphorically, with nothing from the source domain present.

Putting together the Johnson and Narayanan results yields the following hypothesis: In situations where the source and target domains are both active simultaneously, the two areas of the brain for the source and target domains will both be active. Via the Hebbian principle that *Neurons that fire together wire together*, neural mapping circuits linking the two domains will be learned. Those circuits constitute the metaphor.

Grady called such metaphors "primary metaphors" and observed that they are learned by the hundreds the same way all over the world because people have the same bodies and basically the same relevant environments, and so will have very much the same experiences in childhood in which two domains are simultaneously active, and so will learn neural metaphorical mappings linking those domains naturally, just by functioning in the world. Just living an everyday life gives you the experience and suitable brain activations to give rise to a huge system of the same primary metaphorical mappings that are learned around the world without any awareness.

By best fit, different cultural frames will combine with those primary metaphors and give rise to different metaphor systems. The Love Is a Journey metaphor is a good example. The primary metaphors that ground the Love Is a Journey metaphor are:

- Purposes are Destinations: Every day there is a correlation between achieving a purpose and reaching a destination, as when you have to go to the refrigerator to get a piece of fruit or a cold beer.
- Difficulties are Impediments to Motion: A difficulty is something that inhibits your achievement of some purpose, which is metaphorically reaching a destination. Hence, difficulties are conceptualized metaphorically as impediments to motion to a destination.
- A Relationship Is a Container (a Bounded Region of Space): People who are closely related tend to live, work, or otherwise spend time in the same enclosed space — you family in your home, your co-workers at the office, and so on.
- Intimacy is Closeness: The people you are most intimate with are typically the people you have spent time physically close to: your family, spouse, lover, and so on.

In each case, a correlation in experience is realized in the brain as the co-activation of distinct neural areas, which leads to the formation of circuits linking those areas.

A Structural Prediction

The neural theory says that complex metaphors that are extensions of existing primary metaphors bound together should be easier to learn and understand than conceptual metaphors that are totally new — since they just involve new binding and other connecting circuitry over existing conceptual metaphors. They should also seem more natural.

Take, for example, the sentence *My job is a jail*.

(1) A jail is restricts someone freedom of motion to desired external destinations, thus producing frustration and other negative emotions.
(2) The metaphors that *Achieving a Purpose is Reaching a Destination* and *Actions are Motions* exist in our conceptual system.
(3) Binding the *restriction on freedom of motion* to *Actions are Motions*, we infer a restriction on freedom of action.
(4) Binding *desired external destinations* to *Achieving a Purpose is Reaching a Destination*, we infer *achieving external purposes.*
(5) *My Job is a jail* this metaphorically infers that my job restricts my freedom of action in achieving external purposes, thus producing frustration and other emotions.

Thus, given the existing system, maximization of binding produces the meaning of the sentence. We predict (below) that this should be easy to understand and to process.

Compare this sentence with a sentence like *My job is an aardvark*. An aardvark is an African animal with a long proboscis that eats ants by sticking its proboscis in anthills. There are no primary metaphors in our normal conceptual systems that provide a natural metaphorical interpretation for this sentence. However, that sentence can be metonymic, say, when said by a zookeeper whose job is taking care of an aardvark. The metonymy is: In the Animal Keeper Frame, The Animal stands for The Job of Taking care of that Animal

The neural theory in general predicts that the most immediate component metaphors for a complex metaphor will be activated and used in the mapping. In short, in most cases, new conceptual metaphors that are easy to learn and make sense of are using conceptual mappings that pre-exist, frame-based knowledge that pre-exists, and adding connections in the form of circuitry that binds, links, maps, extends, and forms gestalts.

A Processing Prediction

The neural theory of metaphor makes an important prediction in the case of conventional conceptual metaphorical mappings that are realized by fixed brain circuitry. When you hear a metaphorical expression, the literal meanings of the words should activate the source domain circuitry and the context should activate the target domain circuitry, and together they should activate the mapping circuit. The result is an integrated circuit, with activation of both source and target domains and processing over both at once. Thus, understanding language that is makes use of a conventional conceptual metaphor should take no longer than normal frame-based nonmetaphorical processing. This result has been shown over and over, as in the example, *My job is a jail*.

The neural theory thus contradicts old two-step theories (prior to conceptual metaphor theory) that claim that the source domain is processed first and then the mapping operates to process the target domain. Time of processing studies contradict this view.

Asymmetry

Each neuron fires asymmetrically, with the flow of ions from the cell body down the axon, spreading out from there. Different neurons have different firing capacities depending on the receptors at the synapses that regulate ion flow. Those neurons that fire more tend to develop greater firing capacities. And those involved in physical bodily functioning tend to fire more. For this reason, the metaphorical maps learned are asymmetric and tend to have physical source domains (though some have social source domains).

The literature abounds with obvious examples.

- More Is Up: Our bodies are constantly monitoring physical height more than computing abstract quantity.
- Affection is Warmth: Temperature is always there to be monitored; affection isn't.
- Intimacy Is Closeness: We constantly monitor how close we are to objects, more than we judge intimacy.

The preponderance of our system of primary metaphors is acquired in childhood, and childhood experience has an important influence on the system of primary metaphors that we learn. Consider the following important examples:

- Governing Institutions are Families: Our first experience with being governed is in our family. Thus the social domain of the family will be used more when the metaphor is learned.
- Speech Act Force is Physical Force: Parents teach their young children by manipulating their bodies as they give directives. Thus, verbal directives are learned as having a "force."
- Arguments are Struggles: All small children struggle with their parents when their parents guide them physically in teaching them how to behave. Early verbal arguments are commonly about meeting behavioral expectations. As we grow up and learn about wars and battles, the source domain of struggle is specialized and expanded to battles and wars.

During learning, much of the abstract domain is structured by fixed projections from the embodied domain. When processing source domain words in the context of a target domain subject matter, the fixed connections result in co-activation of the two domains. Thus, source domain activations arising from inferences are projected onto the target domain via the pre-established mapping.

The use of conceptual metaphors

The pre-neural theory of conceptual metaphor was vague on a number of details. Metaphors were cross-domain mappings — from a frame in one domain to another domain, also structured by frames. Such mappings were seen as applied to target domain situations as understood in the context of commonplace knowledge. Inferences were mapped from the source to target, with as much as possible frame and image-schema structure "preserved" from the source domain. Thus, in use, you had:

- The metaphorical mapping (from source domain frame to target domain frame)
- The specific situation being discussed, fitting the target domain.
- Target domain commonplace information.
- Source domain commonplace information.

Metaphorical inferences require (1) source domain inferences, (2) mappings of the results of such inferences to the target domain frames; (3) combining of those mapped inferences with target domain information to give new "metaphorical" inferences.

The neural theory of metaphor provides an explanatory mechanism for metaphorical inferences that can be modeled precisely (Narayanan, 1997) using neural computational modeling. At the heart of the modeling of metaphorical inferences is the notion of mental simulation, which represents specific situations. Let us look first at inferences in NTL, and then at metaphorical inferences.

Inferences

A meaningful node in a neural circuit is a node that can activate a mental simulation.

An inference occurs when:

- the activation of a collection of meaningful nodes (the antecedent situation) in a neural circuit leads to the activation of one or more other meaningful nodes (the consequence)
- and when the activation of the antecedent nodes is necessary for the consequence
- and when the inhibition of one or more consequence nodes results in the inhibition of one or more antecedent nodes.

Inferences are simply consequences of the meaningfulness of nodes in simulation semantics, the spreading of activation, and best-fit constraints (the consequences fit the antecedents best). Recall that the maximization of binding is one of the characteristics of the best-fit property of any neural system. In short, maximizing binding can lead to inferences.

Metaphorical inferences

A metaphorical inference occurs when:

- a metaphorical mapping is activated in a neural circuit,
- there is an inference in the source domain of the mapping,
- and a consequence of the source domain inference is mapped to the target domain, activating a meaningful node.

For example, suppose the sentence is *We're driving in the fast lane on the freeway of love*. In the travel domain, driving in the fast lane on the freeway activates the inferences that

(1) the vehicle the travelers are in is going a lot faster than usual,
(2) the driving is exciting and
(3) it can be dangerous (the travelers can suffer physical harm).

"Freeway of love" activates the target domain of love and source domain of travel, resulting in the activation of the Love Is a Journey metaphorical mapping. The metaphorical inferences are that:

(M1) the relationship the lovers are in is developing a lot faster than usual,
(M2) the development of the relationship is exciting, and
(M3) it can be dangerous (the lovers can suffer psychological harm).

These inferences are activated when the circuitry is activated in the processing of the sentence. The totality of source domain inferences does *not* have to proceed before any of the target domain inferences. Source domain inferences will be mapped as soon as they are activated.

Mapping "Gaps"

A mapping gap occurs when there is a metaphorical mapping, but part of the source domain frame has no correlate in the target domain. For example, take the sentence "*I gave Sam that idea*." In this metaphor, the communication of an idea is the transfer of an object from the speaker to the hearer.

(A) Source domain knowledge: the giver loses the object when he gives it to the recipient.
(B) Target domain knowledge: the speaker does *not* forget the idea when he tells it to the listener.

Because we know (B) about the target domain, no mapping from (A) to (B) can be learned. Thus, what appears to be a "gap" is not a gap; it is just that an impossible mapping doesn't take place in the learning of the metaphor. Recall that the learning of the metaphor involves repeated co-activation of the corresponding source and target nodes, and the absence of such co-activation implies that no such maps are learned.

Image-schema "Preservation"

As Regier (1995, 1996) and Dodge and Lakoff (2006) have argued, primitive image-schemas (e.g., container, source-path-goal, degree of closeness, direction and amount of force) are computed by brain structures that are either innate or form early. Action schemas and frames are structured using such primitive image schemas. For example, *putting in* makes use of the container schema, the source-path-goal schema, a force schema, a direction schema, and an aspectual schema.

Metaphorical *putting in* — as in *The Founding Fathers put freedom of speech into the Constitution* — uses physical putting-in as a source domain. The inference patterns of those schemas as bound together in the source domain are then used in metaphorical inferences. For example, if you put something into a physical container, it isn't there before you put it in and it is there afterwards and it remains there until something happens to remove it. That is also true of the freedoms the Founding Fathers put into the Constitution.

In pre-neural theories of conceptual metaphor, we spoke of "preservation" of source domain image schemas. In the neural theory, there is no "preservation" since the source-domain image-schemas are directly used via the neural mapping in target domain inferencing.

Mental Spaces

A "mental space" from an NTL perspective is a mental simulation characterizing an understanding of a situation, real or imagined. The "space" is not a specific place in the brain, but rather the totality of the neural circuitry used wherever in the brain it is located.

The entire space is governed by a gestalt node, which makes the mental space a “unit” which, when activated, activates all the elements of the mental space.

Blending

What is called “blending” is the result of blending maps—complex maps that use neural binding or identification links. Consider the monk blend. There are two mental spaces each structured by frames. In each there is a mountain and a path. On day 1, the monk walks up the path to the top of the mountain, sleeps overnight there and on day 2 walks down the same path to the bottom

Day 1 is one simulation space structured by a Gestalt, day 2 is another. The pair of Day 1 and Day2 form the complex source. There is an ID link between the monk on Day 1 and the monk on Day 2 indicating that they are the same person, though they occur in different simulations. The target of the blend has the Day 1 Gestalt superimposed on the Day 2 Gestalt, with mountain to mountain, path to path bindings, but no monk to monk binding; there are two different monks in the target.

The blending map consists of two ID linking circuits, one from the Day1 monk in the source to the Day 1 monk in the target, and another from the Day 2 monk in the source to the Day2 monk in the target. But the identity linking map between the Day1 monk in the source and the Day2 monk in the source has a neural binding of their Essences, making them understood as the *same* monk. The reflexive pronoun *himself* in “where he meets himself” is the grammatical correlate of that neural binding across Essences.

Question: Is there a single place on the path where the monk is located at the same time on both days? There is a single place where the two different monks meet in the target gestalt. But since those two monks are the same person in the complex source Day1 followed by Day2 source, the answer is “Yes.”

Note that there is no metaphor here.

Metaphors versus blends

Let us now consider a *metaphorical blend*, where the two sources of the blend are the source and target domains of a metaphor.

To see the difference between a metaphor and a metaphorical blend, consider the metaphor More Is Up. In a sentence like, “the temperature went up,” we are understanding quantity in terms of verticality. But they are different things. Amount of heat in itself is not vertical. The metaphor imposes verticality.

Now consider a thermometer. In a thermometer oriented vertically, the mercury goes up physically. There are also scalar marking on the thermometer indicating numerical quantity. We understand a thermometer through the following blend:

Blend Source 1: The verticality of the mercury.

Blend Source 2: The quantitative scale.

Blend Target: The thermometer, where there is a binding between the top the mercury column and the corresponding marking on the quantity scale.

The independently existing metaphor MORE (Quantity) IS UP (Verticality), mapping from Blend Source1 to Blend source 2. Blend Source 1 is the metaphor source; Blend Source 2 is the metaphor target.

The blending map:

-The height of the vertical entity in Blend source 1 is linked by an identifying linking circuit to the top of the column of mercury in the Blend Target.

-A quantity in Blend Source 2 is mapped by an indentifying linking circuit to a corresponding quantity on the scales indicated on the thermometer.

The language for describing the temperature on the thermometer comes from the source domain of the metaphor, Blend Source 1. Example, *the temperature rose.*

This characterizes our understanding of the thermometer, which is a physical object. Note that the general case of points on the conceptual quantity scale functions like an *Essence* of every special case of a physically indicated quantity scale. Similarly, the height on a general Verticality scale function like an *Essence* of particular physical heights. This, the identifying linking circuits are neurally binding Conceptual entities (markings) to Essences (points on a conceptual scale) of the conceptual understanding of a physical object.

Note that the metaphor exists independently of the blend.

Let’s consider another contrast. Suppose you are explaining arithmetic to a child. You draw a line. And you say, “Think of a number as being a point on this line. Say this is zero. And to get to one you take a step from 0 to 1, located here on the line. To add 3 to 1 you take three steps from 1, like this, and you get to 4. To subtract 1 from 4, you take a step backward, and you get to 3.” And so on. Here you just using the metaphor that numbers are points on a line. It is just a metaphor. No blending.

But if you go to the Cartesian plane where you have a number line, then you not only have the metaphor of numbers as points on a line, but you have a blending map as well: The number and the point on the line are understood as *identical* — the same entity! This metaphorical blend is actually in the mathematics of the Cartesian plane.

Again, a mere metaphor (understanding the target in terms of the source) is crucially different from that metaphor plus a blending map from of source entities to a target.

Optimality in Blending

A great deal follows from the understanding of blending in terms of blending maps, given that neural systems work by spreading activation and best-fit principles. Best-fit principles include the maximization of neural bindings and prior neural activation weights. That means the maximal use of conventional frames, metaphors, commonplace knowledge already in the brain, plus context. What results are "emergent" inferences.

From this account of blending we get a set of predictions — exactly the well-known properties of optimal blends as described by Fauconnier and Turner:

- **Integration:** The scenario in the blended space should be a well-integrated scene.

Each neural binding and identifying linking circuit (a link plus a binding) across conceptual structures serves to "integrate" those conceptual structures.

- **Web:** Tight connections between the blend target and the blend sources should be maintained, so that an event in one of the blend sources, for instance, is construed as implying a corresponding event in the blend target.

Such correspondences are given by neural bindings, metaphor maps, maps across simulations (mental spaces), and blending maps.

- **Unpacking:** It should be easy to reconstruct the inputs and the network of connections, given the blending target.

Neural bindings have the property that they can be "relaxed" — that is, neurally bound structures can be conceptualized without the binding, as when you can separate off the blueness of a blue Volkwagen and think of it as red.

- **Topology:** Elements in the blend should participate in the same sorts of relations as their counterparts in the inputs.

This follows immediately since adding a neural binding and identifying linkages to a structure A does not remove any structure in A. It just adds more structure.

- **Good Reason:** If an element appears in the blend, it should have meaning. And if it arises by inference, it will be tied into the logic of the blend.

Since blending maps apply to simulation spaces to yield another simulation space, and since simulations get meaning through their embodiment, this follows immediately.

- **Metonymic Tightening:** Relationships between elements from the same input should become as close as possible within the blend. For instance, Western images of personified Death often depict the figure as a skeleton, thus closely associating the event of death with an object that, in our more literal understandings, is indirectly but saliently associated with it.

When there is a neural binding between A and B, all the circuitry activating and activated by either A or B act together as if A and B were a single entity. This is "tightening." Since blending maps contain indentifying linking circuits, which contain bindings, they give rise to "tightening."

In the case of the Grim Reaper discussed by Fauconnier and Turner, a commonplace metonymy is activated: The Cause Stands for the Result. Here the Cause is Death and the Result is the Skeleton. The Blend has the following structure:

Blend Source 1: Death
BlendSource 2: Skeleton
Blend Target: The Grim Reaper—Death understood as personified in the form of a skeleton.
The Metonymy: Cause (Death) stands for Result (Skeleton).
The Blending map, using identifying links:
-The metonymic source, Death, is identified with the Grim Reaper.
-The metonymic target, Skeleton, is identified with the Grim Reaper.

The use of "Death" to name the Skeleton is a result of the metonymy.

Thus, all of the Fauconnier-Turner optimality properties producing "good" blends are explained by the neural theory of meaning: by frames, metaphors, metonymies, simulation semantics, spreading activation, and best-fit, which governs optimality in biological neural networks.

Emergence

Emergence is the occurrence in a blend of an entity or proposition that does not exist in any of the blend "inputs." Emergence is explained by inference in neural systems. Blending maps and other maps and bindings across conceptual structures can give rise to inferences not present in any "input."

Consider the example, "In France, Clinton's affair wouldn't have mattered." In the blend target, Clinton, the American Chief Executive is identified with the position of the French Chief Executive in France. Since the French don't care about politicians' sexual liaisons, we get the inference that "In France, Clinton's affair wouldn't have mattered." This "emergent" inference does not occur in either of the inputs: France, where Clinton was not chief executive of France, and the US, where Clinton's affair did matter. It arises by maps and inferences arising from simulations when those mapping circuits are activated.

Better Analyses with Metaphoric Blends

Certain classic analyses in the blending literature which are seen as non-metaphoric blends really should be seen as metaphoric blends. For example, there is a common metaphor in which Breaking a Record Is Winning a Race Against the Previous Record-holder. Thus, a few years ago when Mark McGwire and Sammy Sosa were both attempting to break Babe Ruth's home run record, the press represented the situation metaphorically as a race with Ruth — and each other. In the daily papers, McGwire and Sosa were represented by how many games they were "behind" or "ahead" of Ruth's 60 home run performance. They were spoken of as “catching up” or “falling behind.” The classic blending analysis misses this metaphor.

The same metaphor occurred in the situation many years back when the yacht Great America tried to break the San Francisco to Boston record through the Northwest Passage set a hundred years before by the yacht Northern Light. Accordingly, the metaphor had the Great America in a “race” with the Northern Light — even though they sailed 100 years apart. The newspapers daily reported how many days “ahead” of the Northern Light the Great America was. Again, the classic blending analysis misses the metaphor.

The moral: A neural theory analysis forces us to notice analyses we might otherwise miss. The reason is this. As soon as record-breaking is at issue, the general metaphor will be made active. *If it is used, then there is a better fit than if it is not used.* In short, the best-fit principle by which neural networks operate prefers such analyses. Remember: the more existing high-activation-weight circuitry that is activated, the better the fit between the current situation and the prior state of the brain.

This may sound counterintuitive. If you are a scientist trying to give an analysis of some data, you use Ockham’s Razor—the minimum that you need to hypothesize. So if you are a traditional descriptive linguist seeing your job, not as involving the brain, but as writing down the shortest possible written analysis of some sentences you encounter, you try to hypothesize the minimum. You don’t hypothesize a primary metaphor as part of the analysis unless some element of the sentence forces you to. But if you are a cognitive linguist or any other cognitive scientist whose job is to start with the brain, then you start with how the brain works and what is in it when you start. That includes all the primary metaphors and a whole conceptual system. And you start with the best-fit property, and that affects the analysis you give of what the brain would be doing.

Let’s consider another class of cases with the same moral. There are two widely used metaphors rarely analyzed as such.

- A Person who performs actions with certain characteristics is a Member of a Profession known for those characteristics.

Here, the mapping is from the frame of a member of a profession, with the characteristics that members of a profession are known by. Special cases, e.g., a surgeon frame, expands the general frame with the characteristics filled in by surgeon characteristics, while a butcher frame expands the general frame with the characteristics filled in by butcher characteristics.

In each case, the source domain of the metaphor is a stereotype, represented as a frame whose semantic roles include kinds of characteristics. For example, a *surgeon* is known for being precise with the beneficial results, while a *butcher* is known for being sloppy and acting more with force than with care, with messy results. Thus, we can say

- My lawyer presented my case with surgical skill.
- My lawyer butchered my case.

In the first, the lawyer was careful and skillful, with beneficial results. In the second, the lawyer was careless, sloppy, and heavy-handed, with messy results. Other examples can be quite diverse. Here is one for baseball fans:

- Ichiro slices singles through the infield like a surgeon.
- Frank Thomas hacks at the ball like a butcher.

This very general metaphor accounts for the classical examples:

- My butcher is a surgeon.
- My surgeon was a butcher.

The first case says the butcher cuts meat with the care of a surgeon, while second says that my surgeon handled my surgery in a careless, sloppy, and heavy-handed way.

A second example like this is the commonplace metaphor:

- A Person with characteristic properties is an Animal known for those properties.

Classic cases include *Man is a wolf*, *Our new salesman is a tiger, Harry’s a pig*, and *You’re trying to weasel out of this.* All are examples where there is a stereotype of an animal and we are understanding the person in terms of the characteristics of the animal stereotype.

There have been attempts to understand such cases non-metaphorically, just in terms of bindings based on similarity. Such an approach would claim that there is no conventional metaphor at all, and that all such cases are literal blends based on similar (partly identical) properties. We can see what is wrong with this approach by looking at cases outside the proposed conventional metaphors we just discussed. Consider sentences like:

- My surgeon is a Russian.
- My butcher is a Russian.
- My lawyer is a Russian.

There are common stereotypes of Russians, say that they are very sentimental and emotional, sometimes to the point of losing control. If the blending approach were correct, we would expect these sentences to act like *The butcher is a surgeon* and *The surgeon is a butcher*. Just as the butcher isn’t literally a surgeon by profession, nor is the surgeon literally a butcher by profession, so you would expect these sentences to be saying that the surgeon, butcher, and lawyer were not literally Russian by nationality; but they do say that. In addition, you would expect them to be saying that the surgery, butchering, and law practice is carried out in an overly sentimental, emotional, almost out-of-control way. But the sentences do not say that. The Russian-sentences are literal and work just as you would expect literal sentences to work. The surgeon-butcher sentences are metaphorical, using conventional conceptual metaphors, and they work accordingly.

The Role of Metaphor in Abstract Concepts

In *Whose Freedom?* I argue that metaphor is central to the core concept of freedom, and that this abstract concept is actually grounded in bodily experience.

Physical freedom is freedom to move — to go places, to reach for and get objects, and to perform actions. Physical freedom is defined in a frame in which there are potential impediments to freedom to move: blockages, being weighed down, being held back, being imprisoned, lack of energy or other resources, absence of a path providing access, being physically restrained from movement, and so on. Freedom of physical motion occurs when motion is unimpeded, that is, when none of these potential impediments is present.

Various metaphors turn freedom of physical motion into freedom to achieve one's goals. The Event Structure Metaphor, for instance, characterizes achieving a purpose as reaching a desired destination, or getting a desired object. Freedom to achieve one's purposes then becomes, via the Event Structure Metaphor, the absence of any impediments to reaching one's goals. Other ideas, like political freedom and freedom of the will build on that concept.

The concept of political freedom is characterized via a network of concepts that necessarily includes the Event Structure Metaphor and the inferences that arise via that metaphor. The ultimate grounding of the concept of political freedom is visceral. Arising from the experience of *not* being free to move and the frustration that engenders.

What is the role of metaphor in our concept of political freedom? Our understanding of conceptual systems in terms of neural systems shows that conceptual metaphor is used in our understanding of political freedom, though it is used indirectly.

Metaphor in Systems of Thought

In *Philosophy in the Flesh*, Mark Johnson and I argue that philosophical systems of thought rest on a relatively small number of metaphors treated as ultimate truths and used constantly in reasoning. The neural theory of metaphor allows us to understand more about such systems and people who think in terms of them most of every day.

Because the fundamental metaphors are used constantly, the synaptic strengths in the metaphors become very very strong and resistant to change. Secondly, spreading activation and best-fit properties (including maximization of binding) make such systems highly integrated, tightly connected, with a lot of inferences. As a result, such a system will dominate your thought, your understanding of the world, and your actions.

One will tend to see the world through the system; one will tend to construct neural simulations to fit the system; one will tend to plan the future using the system; one will define common sense though the system; and the system will tend to make experiences and facts that are consistent with it noticeable and important, and experiences and facts that are inconsistent with it, invisible.

This is especially true in politics, where progressive and conservative thought are each defined by a central metaphor and a system of thought that fits it (see my book *Moral Politics*).

By far the most detailed study of the role of metaphor in a system of thought is Rafael Núñez' and my book, *Where Mathematics Comes From*, which shows in great detail how many branches of higher mathematics are built up via layers of metaphor from embodied concepts.

Metaphorical Language

The neural theory of language allows us to understand better why language is so powerful. Let's start with words. Every word is defined via a linking circuit to an element of a frame — a semantic role. Because every frame is structured by a gestalt circuit, the activation of that frame element results in the activation of the entire frame. Now, the frame will most likely contain one or more image-schemas, a scenario containing other frames, a presupposition containing other frames, may fit into and activate a system of other frames, and each of these frames may be structured by conceptual metaphors. All of those structures could be activated simply by the activation of that one frame element that defines the meaning of the given word. In addition, the lexical frame may be in the source domain of a metaphor. In that case the word could also activate that metaphor. In the right context, all of these activated structures can result in inferences.

Let's suppose a word activates a network of frames, image-schemas, and metaphors. The metaphors may be only indirectly linked to the frame directly activated by the word. Is that word an instance of "metaphorical language"? That is not how the term is usually used.

We usually speak of metaphorical language when

- the frame element the word designates is in the source domain frame of the given metaphor
- the subject matter under discussion is in the target domain of that metaphor.

Thus, *up* in the sentence *Prices went up*, activates the verticality frame, *prices* activates the quantity frame, and together they activate the *More is Up* metaphor.

In addition, the word *up* — by virtue of the metaphorical mapping — acquires a link to the quantity frame, where it activates greater quantity.

Does *up* in *Prices went up* always activate the *More is Up* metaphor? In our neural systems, the *More is Up* metaphor is always present, with the idea of greater height always physically linked by an existing circuit to the concept of greater quantity. The metaphor is a physical circuit, ready to be activated. The language activates it.

Grammar can also play a role in activating a metaphor, as in the expression *freeway of love*, in which the construction sanctions an interpretation in which the head noun *freeway* comes from the source domain (travel) and object of the preposition *love* comes from the target domain. Grammatical constructions come with metaphorical constraints, as Karen Sullivan has observed. Compare *bright student* versus **intelligent light*: the modifier (*bright*) is from the source domain, while the head (*student*) is from the target domain; but the reverse doesn't work — except in a special class of cases, like *emotional intelligence*, where the modifier is a non-predicative adjective that defines a domain (emotion).

All this is natural in a neural theory because of the connectivity involved. The form elements (words and grammatical categories) are neurally linked to the elements in conceptual system, where metaphorical mappings are linked to frame elements, which are linked to words or grammatical categories.

Consider a poetic metaphor like Dylan Thomas's line, *Do not go gently into that good night.* The line does not overtly mention death as the subject matter, but the line contains three words that each evoke a source domain frame in a metaphor for death: *go* as in *Death is Departure*; *gently* as in *Life is a Struggle*; and *night* as in *A Lifetime is a Day and Death is Night*. This is natural from a neural perspective. Each word activates a frame element in a frame *go*, *gently*, *night*. The three frames are thereby activated and each provides some activation to the corresponding metaphors for Death. This is reinforced by the fact that the sentence does not have a direct literal meaning, in which each of these words is used literally. But the source domain meanings do important work in constructing a metaphorical image of a man moving into the night ready to fight. The next line, *Rage, rage against the dying of the light* uses *dying* metaphorically in the sense of light ceasing to exist. But the activation via the metaphor of source domain of death reinforces the interpretation of the first line. This use of "activation" makes sense in the neural model.

The Use of Metaphoric Language

The neural theory of metaphor also makes sense of the use of metaphoric language in context. We know that metaphor does not reside in words but in ideas. This is especially clear from cases of metaphorical ambiguity, where the same words evoke different readings using different metaphors. "It's all downhill from here" may in a given situation mean "it's getting easier" (Ease of Action Is Ease of Motion) or "it's getting worse" (Down is Bad). Either conceptual metaphor can apply to the spatial meaning of "down" in "downhill." In a neural account, both metaphors activate to the spatial meaning of "down," but the metaphors are mutually inhibitory. Only one can be activated, depending on context.

Consider a metaphorically ambiguous sentence like "Let's move the meeting ahead two days." If uttered on a Wednesday, it could refer to either Monday or Friday, depending on which metaphor for time is used — moving-ego or moving-time. Since they are mutually contradictory, the metaphors are mutually inhibitory. The neural theory can explain Lera Boroditsky's classic experiment at San Francisco airport. She showed that, for people waiting for a plant to come in, the motion of the plane toward them primed the moving time metaphor and they gave the answer "Monday," two days ahead of the moving time. Those who were on the plane and coming off were primed by being on the moving object, and they gave the answer "Friday," two days ahead of the moving ego.

The neural theory explains the priming in these cases. The two time metaphors are mutually inhibitory. What tips the scales is the priming – the neural activation of either a moving time or moving ego in the spatial domain.

What makes metaphorical language meaningful?

Language is meaningful when the ideas it expresses are meaningful. Conceptual metaphors are meaningful when they are grounded. They are grounded, first, by source domain embodiment, and second by the embodiment of the source and target domains of of the primary metaphors being used.

Summary: What Does the Neural Theory Provide?

The neural theory provides a much better understanding of how thought and language work and of how metaphorical thought fits into the picture. It also provides *explanations* for a host of phenomena. And it changes how one does metaphor analysis — and redefines what metaphor analysis is.

The neural theory explains:

- Why there should be conceptual metaphor at all; what conceptual metaphors are physically; why we have the metaphors we have, how the system is grounded, and why certain conceptual metaphors are widespread around the world.
- How metaphorical inferences work; why they should exist; how they operate in context, and how they interact with simulations.
- All of the properties of the old metaphor theory, the theory as described by myself and Mark Johnson in *Metaphors We Live By* and by myself in the essay "The Contemporary Theory of Metaphor."
- How metaphors can function *indirectly* in the characterization of abstract concepts.
- How a small number of metaphors can organize a whole system of thought and become the principles on which one lives one's life.
- How metaphorical language works as a simple extension of non-metaphorical language.
- Why metaphors differ from blends, and why blends do not do the job of metaphors.

The neural theory also clarifies what the study of metaphor is about, namely,

- Showing how metaphorical understanding is grounded in basic human experience via primary conceptual metaphors;
- Showing how primary metaphors contribute to complex conceptual metaphors;

- Showing how both primary and complex metaphors contribute to the meanings of words, complex expressions, and grammatical constructions;
- Showing how conceptual metaphor plays a role in abstract concepts and overall conceptual systems (as in politics, philosophy, and mathematics);
- And, finally, showing how conceptual metaphors contribute to the understanding of language and other uses of symbols.

How Does a Metaphor Analyst Make Use of All This?

Metaphor analysts rarely know neural computation, and they shouldn't be expected to. The Neural Theory of Language Project has figured out a way to let linguists be linguists and not computer or brain scientists. We have invented a notation that correlates with circuitry with the appropriate computational properties, but can be used by analysts without worrying about the computational details. Thus, consider a notation like:

```
Metaphor: LoveIsAJourney
Source Domain: Journey
Target Domain: Love
Mapping:
        Travelers               —>      Lovers
        Vehicle                 —>      Relationship
        Destinations            —>      LifeGoals
        ImpedimentsToMotion     —>      Difficulties
Evokes:
        Purposes Are Destinations Metaphor,
                with Destinations = Self.Source.Destinations
                     Purposes     = Self.Target.LifeGoals

        Difficulties Are Impediments to Motion Metaphor,
                With Impediments to Motion = Self.Source.ImpedimentsToMotion
                     Difficulties          = Self.Target.Difficulties

        Intimacy Is Closeness Metaphor,
                With Closeness = Self.Source.ClosenessOf TravelersInVehicle
                     Intimacy  = Self.Target.IntimacyOfLovers

        A Relationship Is A Container Metaphor,
                With Container    = Self.Source.Vehicle
                     Relationship = Self.Target.Relationship
```

The statement that this is a metaphor corresponds to the appropriate mapping circuit. The name of the metaphor corresponds to the appropriate gestalt node. The arrows ("—>") correspond to linking circuits. The statement of the mapping specifies what maps to what. The equal signs ("=") specify the neural bindings. The "evokes" statement sets up linking circuits activating the "component" metaphors, with neural bindings between LoveIsAJourney (called "Self" in the formalism) and the various component metaphors. There can be, and often is, a chain of "evokes" statements that ultimately lead to primary metaphors that ground the metaphor system in experience.

This formalism is easy for metaphor analysts to learn and use. It can be converted by algorithm to computational neural modeling programs that, say, take a sentence as input and produce an analysis as output. There are corresponding formalisms for grammatical and lexical constructions, metonymies, frames, image-schemas, and so on. The technical term for the notational system is *Embodied Construction Grammar*.

Conclusion

This is where we are in the Neural Theory of Metaphor as of January 2009. We have a reasonable early approximation to the kinds of computations that neuronal groups must perform to characterize frames, metaphors, metonymies, mental spaces, and blends. A parsing program to use these kinds of computations is being constructed. Thousands of frames and hundreds of metaphors have been analyzed informally to date and can readily be converted to the notation system. And we know enough about natural metaphor learning to understand how the metaphor system gets built up just by functioning in our everyday lives.

References

Dodge, E. and G. Lakoff (2006). "Image schemas: From linguistic analysis to neural grounding." In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. B. Hampe (ed.). Berlin, Mouton de Gruyter**:** 57-91

Fauconnier, G. and M. Turner. *The Way We Think*. New York: Basic Books. 2002.

Feldman, J. *From Molecules to Metaphors*. Cambridge, MA: Bradford MIT Press. 2006.

Feldman, J. and S. Narayanan (2004). "Embodied meaning in a neural theory of language." *Brain and Language* **89**(2): 385-392.

Grady, J. *Foundations of Meaning*. UC Berkely Dissertation. 1997.

Grady, J., Oakley, T., and Coulson, S. "Blending and Metaphor." In *Metaphor in cognitive linguistics* , G. Steen and R. Gibbs (eds.). Philadelphia: John Benjamins. 1999.

Gallese, V. and G. Lakoff. "The brain's concepts" In Cognitive Neuropsychology, 2005, vol. 22, n. 3-4, pp. 455-479.

Johnson, C. Constructional grounding. 1999. UC Berkeley Ph.D. dissertation.

Lakoff, G. *"The contemporary theory of metaphor." In A. Ortony,* Metaphor and Thought*. New York: Cambridge University Press. 2nd Edition. 1993.*

Lakoff, G., and M. Johnson. *Metaphors We Live By.* Chicago: University of Chicago Press. 1980; 2nd edition, 2002.

Lakoff, G. and M. Johnson*.* *Philosophy in the Flesh*. New York: Basic Books. 1999.

Lakoff, G. and R. Núñez. *Where Mathematics Comes From.* New York: Basic Books. 2000.

Narayanan, S. Karma: Knowledge-based Action Representations for Metaphor and Aspect. UC Berkeley Dissertation, 1997.

Regier, T. *The human semantic potential: Spatial language and constrained connectionism.* Cambridge, MA: MIT Press. 1997.

Sullivan, K. *Grammar in Metaphor: A Construction Grammar Account of Metaphoric Language.* UC Berkeley Dissertation. In progress, due 2007.

Talmy, L. (2000). *Toward a Cognitive Semantics*. Cambridge and London, MIT Press.