

**The Neurocognitive Self:
Conceptual System Research in the 21st Century
And The Rethinking of What a Person Is**

**George Lakoff
University of California at Berkeley**

The genre of futurism begins with a look at the past. Go back 50, 100, or several hundred years and list a group of technological innovations that are commonplace now that were barely dreamed of then: personal computers, satellite communication, men on the moon, birth control pills, genetic engineering. Then project forward, trying to guess from ideas on the drawing board what the science of the future will bring.

Since Cognitive Science isn't all that old, we only need go back 25 years to a time before there was a Cognitive Science. Twenty-five years ago, in 1968, I was a linguist in my 20's teaching at Harvard. The elements of early cognitive science were falling into place. Back in those days, I believed that thought was symbol manipulation, that ideas were represented in the mind in logical forms -- symbolic structures akin to formulas in predicate logic or expressions in LISP. I believed that we represented situations in the world in terms of mental models, which were miniature mental versions of the set-theoretical models of formal logic. Back then I believed in a version of transformational grammar, that is, I believed that sentences were derived from underlying structures like those of predicate logic, and that the "derivations" took place not in real time but in some abstract mathematical time that, somehow, had cognitive reality. I believed that conceptual categories were defined by necessary and sufficient conditions.

Above all, I believed in functionalism, symbolic processing, and classical semantics. Functionalism is the idea that the mind can be studied in terms of the mental functions it performs without looking at the way the brain actually works. Those mental functions were assumed to be carried out by symbolic processing -- by algorithmic manipulation of arbitrary symbols without regard to what the symbols mean. Classical semantics assumes that the symbols being processed are to be given meaning by being associated with things in the world, thus making symbolic expressions into internal representations of external reality.

Those were the ideas that I shared with those from psychology, linguistics, anthropology, philosophy, and artificial intelligence that formed the first generation of cognitive science researchers. How was I to know that a mere 25 years later all those ideas would have been empirically disconfirmed, and entirely new views of thought and language proposed to account for the new data from cognitive science.

We have come a long way. The logical forms of a quarter-century ago characterized six real properties of conceptual structure: coreference, the binding of variables, predicate-argument structure, operator scope, propositional functions, and quantification. But the remarkable aspects of conceptual structure that have come to light since then have taken us far beyond old-fashioned predicate logic and LISP representations. Since the late 1970's and early 1980's, we have learned about all of the following aspects conceptual structure:

Image-schemas: Topological and orientational structures that are used in the characterization of spatial relations concepts. They characterize spatial logic, are used in complex spatial relations concepts, and are projected by conceptual metaphors onto concepts in abstract domains. Examples include the container schema (for concepts like IN and OUT), the SOURCE-PATH-GOAL schema (for concepts like TO and FROM), the contact schema, the verticality schema, and so on. A complex concept such as ON makes use of three elementary schemas: verticality, contact and support.

Frame semantics: Knowledge is organized in holistic structures called *frames* or *schemas* (or, when they involve time, *scripts* or *scenarios*). Frames bring together diverse pieces of information into gestalts. Words are defined relative to such frames.

A celebrated example is the simple commercial event script, in which the elements are a buyer, a seller, goods, and money. The scenarios has three parts: First, the buyer has the money and the seller has the goods. Then they exchange goods and money. And finally, the seller has the money and the buyer has the goods. Words like "buy," "sell," "goods," and "price" are defined with respect to this scenario. Frames are not isolated structures with simple boundaries like the knowledge representations of two decades ago. Rather they partake of complex overlappings and interactions that seem more like the kind of phenomena that connectionism was developed to handle.

Conceptual metaphor: Most abstract concepts are conceptualized in terms of more concrete concepts via conceptual metaphors. A conceptual metaphor is a mapping from a source domain onto a target domain. For example, time is conceptualized in terms of moving objects, as in sentences like:

The time for action has arrived. The time for celebrating is long since gone. A parade will follow the awarding of prizes.

Here the present is conceptualized in terms of an observer situated at the present time and facing toward the future. Future times move toward the observer from front to back. A conceptual system contains thousands of such mappings. Such mappings preserve image-schematic structure; for example, source domain paths

are mapped onto target domain paths; source domain containers onto target domain containers; and so on. By this mechanism the logic of space is mapped onto the the logics of the full range of abstract domains, with spatial inference patterns mapped onto abstract inference patterns. For example, Boolean logic is the metaphorical projection of the logic of containers (that is, bounded regions in space).

Radial categories: These are categories that have centers and extensions of various types. The major mechanisms of category extension from center to periphery include (1) similarity, (2) conceptual metaphor, and (3) shared frames.

A radial category of type 1 is *mother*. Central cases of mothers are defined by four converging folk models: a birth model, a nurturance model, a marriage model, and a genetic model. The prototypical mother gives birth to the child, nurtures it, is married to the father, and is the female from whom it gets half its genes. Other types of mothers -- birth mothers, stepmothers, foster mothers, genetic mothers, and so on -- bear similarities to central mothers, but are mothers relative to fewer models.

A radial category of type 2 is *harm*. The central case of harm is physical harm. But there are also metaphorical forms of harm: psychological harm, social harm, political harm, economic harm, and the harm that is characterized as the thwarting of someone's interests. Metaphorical forms of harm are nonetheless forms of harm. All are part of the same concept, but there are differences -- differences recognized by law.

A radial category of type 3 is the Dyrirbal *balan*, as discussed in Lakoff, 1987. The central case is women, but other members are in the category by virtue of shared frames. For example, the myth that the sun is the wife of the moon is a frame shared by the sun and women, placing the sun in the *balan* category. Fire and the sun share the frame of hot things, which puts fire into the same category as the sun and women. And so on.

Prototype systems: Not all members of a conceptual category have the same status; some are better examples of the category than others. Moreover, there are many types of prototypes: Typical cases, on which default inferences are based. Social stereotypes, used for making snap judgments. Graded prototypes, indicating the degree to which something is a good example. Ideal prototypes, used as standards of judgement. Salient exemplars, used for making probability judgments. And radial category centers, used as the basis for the extension of a concept to noncentral cases.

Basic-level concepts: These are concepts that arise from the optimal interaction of people with their external environment. They are based on gestalt perception, the capacity to form mental images, and motor programs. Compare, for example, the concepts *furniture* and *chair*. Several things make a chair basic-level and furniture superordinate (without basic-level properties). People can perceive chairs as gestalts, but not pieces of furniture. One can get a mental image of a chair, not of a general piece of furniture (as opposed to a specific kind like a chair, table or bed). People have motor programs for interacting with chairs, but not with furniture in general.

Contested concepts: These are concepts where everybody seems to have a different idea of what the concept is. Common examples include democracy, art, and feminism. Our best current theory of contested concepts is that there is an uncontested, but underspecified core concept, which gets extended on the basis of ideologies or other sets of beliefs about more general subject matter.

Mental spaces: These have replaced the mental models of the 70's. They differ in that they can simultaneously represent multiple models and complex links across them.

These phenomena go well beyond what predicate logic, LISP-like formalisms, and set-theoretical semantics can deal with. A cognitive semantics in which such phenomena can be accurately characterized is now under development, and looks very little like predicate logic. Just as representations based on predicate logic have been superceded, so algorithmic symbol manipulation has been superceded by connectionist computation, which attempts to mirror the kinds of operations carried out by neural structures of the brain. Such connectionist systems are, by their very nature, constraint satisfaction systems.

Within contemporary cognitive linguistics, old-style transformational grammars have been superceded by cognitive grammars and construction grammars. The basic units of grammar have become constructions, which are schema-like structures incorporating semantic, pragmatic, and syntactic information. Constructions, like frames, combine by superposition, and each construction defines a set of constraints. A grammar is a network of constructions that combine via a constraint satisfaction system that defines a language. This view of grammar not only does better at handling purely grammatical problems, but it meshes better with what is known about neural computation.

What unites all these results is a very new view of conceptual systems and of the mind itself. On the old view, the mind was disembodied -- a mathematical object, an abstract device for manipulating symbols without regard to their interpretation, floating free of the body and only incidentally "implemented" in the brain. The new view requires that the mind be seen as essentially embodied:

Basic-level concepts are characterized partly in terms of gestalt perception and motor programs.

Color categories are grounded in the neurophysiology of color vision.

Spatial relations seem to be embodied in the perceptual system (as we shall see shortly).

And conceptual metaphors are part of a set of imaginative mechanisms that project from concepts with a direct bodily grounding to abstract concepts. The new view of concepts as getting their meaning via bodily grounding and metaphorical projection supercedes the old view of concepts. On the old view, concepts were seen as abstract symbols that get their meaning by being associated directly with things and categories in an external world that was assumed to all the structure built into it to account for the structure in our conceptual systems. We now know that much of the structure of our conceptual systems has as much to do with our bodies and brains as it has to do with the world external to our bodies.

Two Futuristic Hypotheses

Nonsymbolic Neural Representations of Concepts. A particularly important recent result indicates that the representation of concepts using symbols is an artifact of the fact that we write, rather than an intrinsic part of the nature of concepts. Terry Regier, in his Berkeley dissertation (1992) constructed a connectionist system in which spatial concepts are learned and represented in terms of neural structures without symbols like those used in formal logic, linguistics, or classical AI. Using connectionist networks, Regier made use of neural structures of the kind found in topographic maps of the visual field, orientation sensitive cells, and center-surround receptive fields. The system learned to acquire synaptic weights which enabled the neural structures to perform the functions of image-schemas and to represent the meanings of spatial relations concepts. Regier's results go beyond the capacities of symbol-manipulation systems and indicate that we may have at hand a method for beginning to represent concepts using neurally-inspired models rather than models with abstract symbols. Moreover, these results give a very concrete idea of what it would mean for concepts to be grounded in the body.

My first futuristic hypothesis is what I will call the Neural Representation Hypothesis: that Regier's work will lead us within the 21st century to a general

technique for representing concepts using neurally-inspired nonsymbolic structured connectionist models. They will model how concepts are grounded in the body, as Mark Johnson, myself, and others have hypothesized. In particular, I foresee an extension of Regier's work to characterize spatial inferences, and a further extension to characterize how metaphorical mappings can be represented in neurally-inspired connectionist models so as to map spatial relations and their inference structures onto abstract domains, such as time, events, the emotions, etc.

Regier's models not only *represent* spatial concepts, but they *learn* spatial concepts and the words and morphemes that express them linguistically. Regier has used connectionist learning techniques along with biologically-motivated and structured neural nets to accomplish the learning -- with no explicit negative examples! I believe that Regier's work will spark a new direction in conceptual learning using biologically-motivated structured connectionist models.

The Convergence Zone Hypothesis. Results about the structure and function of the brain in recent years have tended to converge with all the above results about the nature of conceptual structure and grammar. Earlier generations of linguists and cognitive scientists believed in a simple-minded modularity hypothesis that I will call **Localized Modularity** -- namely, the belief that specialized complex brain functions are completely performed in isolated modules. According to **Localized Modularity**, whenever a deficit of some cognitive function results from a lesion in a particular region of the brain, it follows that the *entire* cognitive function is carried out in that region in people without the lesion. Thus, face recognition, according to the **Localized Modularity**, is entirely carried out in one particular region in the brain, which, if lesioned, results in a loss of the capacity for face recognition.

This view of brain modularity led to a hypothesis about mind modularity: The mind too was seen as made up of modules, and accordingly, there were supposed to be isolated modules in the brain for specialized cognitive functions such as language.

In recent years, the **Localized Modularity** hypothesis has become a thing of the past, largely due to the emergence of pet scan and MRI research. Capacities such as face recognition, which can be destroyed by lesions in certain locations, have been shown to involve activation of many portions of the brain, not just the regions where lesions destroy the capacity. This has led to the hypothesis that there exist "convergence zones" where information is brought together and "bound" neurally. For example, in face recognition, many individual features of a particular face must be bound together to form the overall gestalt of that face. The convergence zone hypothesis says that the binding of those features occurs "temporally" via the simultaneous activation of all the disparate brain regions

that compute those features. There are neural connections from the areas where those features are computed to a higher level where there is a neural ensemble whose job it is to govern the binding of those disparate features. A lesion in the convergence zone is, according to the theory, responsible for the destruction of the binding function. But where there is no lesion, the computation may be done in widely scattered regions of the brain. The convergence zone is where the activation of disparate functions is coordinated, but not where the functions are actually carried out. Instead, of being localized in a module, the computations characterizing brain functions are widely distributed throughout the brain, while being coordinated from a variety of convergence zones.

What makes the research on convergence zones by Hannah and Antonio Damasio and their co-workers especially powerful is that it explains why certain lesions in the visual cortex but not in the frontal area can result in the loss of basic-level categories for animals but not for artifacts. Some people experience the loss of the ability to distinguish basic-level animal categories, say, camels from elephants, but not the ability to distinguish basic-level artifact categories, say knives from forks. Basic-level animal categories make use of gestalt perception and mental imagery, while basic-level artifact categories also make use of motor programs. On the convergence zone hypothesis, the convergence zone for basic-level artifact categories would occur at a level in the brain where both visual and motor information could be brought together, while the convergence zones for the animal categories would occur at a lower level, where no motor information is involved. Hence a lesion at the lower level (more towards the rear of the brain) could result in the loss of one kind of basic-level category but not the other. Such explanations by the Damasios give support to the idea of convergence zones.

Extending the work on convergence zones by the Damasios, I propose what I call The Neurocognitive Convergence Hypothesis (NCH). It is an attempt to provide a unified framework for many of the relatively new discoveries mentioned above: Complex image-schemas, semantic frames, conceptual metaphors, radial categories, basic-level concepts, and grammatical constructions. In each of these cases, disparate information is brought together to form a schema with gestalt properties, and in each case the schemas combine by superposition and form a constraint satisfaction system. The different types of phenomena -- basic-level categories, grammatical constructions, conceptual frames, conceptual metaphors, etc. -- result from the location of convergence zones in different locations in the brain.

The Neurocognitive Convergence Hypothesis is a hypothesis that proposes that convergence zones in the brain explain why a wide variety of seemingly disparate cognitive phenomena have the same basic properties. It uses a hypothesis about brain structure to unify research on conceptual systems and

language. I do not know if it is true, but I suspect that it is, and I optimistically hope that it will guide research into the 21st century to discover whether or not it is true.

Grammar and Neuroscience

I believe that research in neuroscience will utterly transform the landscape of linguistics. The first major casualty will be generative grammar, which requires the assumption that language is autonomous, that is, independent of the rest of cognition. Chomsky has hypothesized that there is a "language organ", a separate module of the brain devoted exclusively to grammar. To guarantee that language is independent of all the rest of cognition, Chomsky's "language module" would have to have *no input from the rest of the brain*. If it did have such input, then other aspects of cognition could influence grammar, and the central hypothesis of generative grammar -- total autonomy from meaning and other aspects of cognition -- would be ruled out. But there is no such thing in the brain as a module with no input. For this reason, the properties of the brain module required by Chomsky's theory does not accord with those of real brain modules. For a discussion by a prominent neuroscientist, see Edelman (1992, pp. 211 - 252.)

Incidentally, the autonomy hypothesis is not a casual aspect of generative grammar that might be jettisoned. It is central to the endeavor. The reason is that the mathematical foundations of generative grammar -- the theory of formal grammars -- is defined as the manipulation of symbols without regard to what they mean or how a brain might process them. There is no place in the mathematical foundations of the theory for input from other aspects of cognition. Moreover, the mathematics requires the existence of "operations" that are, like deductions in formal logic, 'performed' in abstract time, not in real time. The brain is not capable of performing operations in "abstract time." In short, the kind of things that brains do is very different from the kind of things that generative grammars do.

The Damasio's Convergence Zone Hypothesis has two important consequences for grammar. First, it explains a set of findings in cognitive linguistics, namely, (1) Grammar is organized in terms of *constructions* which pair semantic and pragmatic information with surface (that is phonological) form, e.g. pronunciation order of constituents, grammatical morphemes, etc. (2) Constructions are *gestalts* in which disparate kinds of information are brought together into a coherent whole. (3) Constructions combine by superimposition. (4) Less central constructions inherit constraints from more central constructions ; that is, peripheral constructions use as much of they can of the content of the central constructions. These are exactly the properties that one would predict from the CZH,

but not from the theory of formal grammars.

Second, the CZH explains away what had been taken as evidence for the autonomy of syntax. Aphasics with lesions in Broca's area suffer from agrammatism, a disorder in which people know the meanings of words but cannot put them together grammatically to form grammatical sentences. Under the Localized Modularity hypothesis, it follows that if a lesion in an area affects only grammar and leaves semantics intact, then that area must perform the function of grammar alone, and not any semantic function. But if the Convergence Zone Hypothesis is right, then Localized Modularity is wrong, and the conclusion no longer follows. The existence of agrammatism does not entail the independence of syntax from semantics. It could mean instead that a convergence zone is located in Broca's area, a zone that links information about the semantics of grammatical constructions with information about the phonological form of constructions -- the order of pronunciation of the constituents, the phonological forms of grammatical morphemes, etc. That is, it is consistent with the idea that syntax is the pairing of categories of semantic elements with phonological restrictions (where pronunciation order counts as "phonological." In short, agrammatism is as consistent with cognitive grammar as with generative grammar, which eliminates a classical argument for generative grammar and the autonomy of syntax.

The Convergence Zone Hypothesis is a hypothesis about the structure of the brain. The Neurocognitive Convergence Hypothesis is a hypothesis about how the CZH provides a unified account of such apparently disparate phenomena as basic-level concepts, metaphors, and grammatical constructions.

Twenty-five years ago, I could not have imagined any of these developments that I now teach routinely in Cognitive Science 101. At that time there were in our intellectual landscape no prototypes, no basic-level categories, no radial categories, no image-schemas, no conceptual metaphors, no mental spaces, no theory of grammatical constructions, no connectionism, no convergence zones. I have no doubt that the current intellectual landscape too will change just as radically in the next 25 to 100 years. But given the license to engage in futuristic speculation, let us consider where the new developments that have just been outlined might lead.

Conceptual System Research

It is my opinion that the most dramatic effect of cognitive science on culture in general will come from conceptual system research. That research has the capacity to change our understanding of everything from law and politics to philosophy to interpersonal relations and the nature of public discourse. The most radical changes in our perceptions of ourselves and our social lives will, I believe, come from three sources: conceptual metaphor, radial categories, and contested

concepts.

Conceptual metaphor is the most radical of these developments. We have discovered, over the past decade and a half, that a conceptual system contains an enormous subsystem of thousands of conceptual metaphors -- mappings that allow us to understand the abstract in terms of the concrete. Without this system, we could not engage in abstract thought at all -- in thought about causation, purpose, love, morality, or thought itself. Without the metaphor system there could be no philosophizing, no theorizing, and little general understanding our everyday personal and social lives. But the operation of this vast system of conceptual metaphor is largely unconscious. We reason metaphorically throughout most of our waking, and even our dreaming lives, but for the most part are unaware of it. At present, the metaphor system of English has barely begun to be worked out in full detail, and the metaphor systems of other languages have been studied only cursorily. Working out the details would be a huge job -- not as big as the human genome project, but most likely, more beneficial. For what is at stake is our understanding of ourselves and our daily lives, and the possibilities for improvement through that understanding. The other radical innovations are radial categories and contested concepts.

Radial categories occur where a concept has a central case that gets extended to other cases, often by conceptual metaphor. For example, the concept of HARM, has physical harm as a central case, with metaphorical extensions to mental, social and economic harm, as well as to causal harm -- the thwarting of one's purposes. Radial categories occur in every aspect of our lives, and the difference between central cases and extensions has repercussions in many realms of social, interpersonal, and intellectual life. Truth, as I have argued elsewhere, is another radial category. It is important to describe the radial structure of important concepts in full detail, and to bring to general awareness the fact that radial structure exists.

A great many concepts are contested concepts, especially social concepts like democracy or feminism or justice. An enormous numbers of misunderstandings and pointless disputes arise from a failure to understand the nature of contested concepts -- especially the fact that what is contested is usually the product of divergent belief systems. It is commonly recognized that there are disagreements as to what, say, feminism or democracy are. It is vital to understand that such disagreements are not over matters of fact, but over what ideology or more general belief system should prevail.

Imagine for the moment that the general public recognized the importance of understanding our mental life (as well as the mental lives of those in other cultures) and that research on conceptual systems was reasonably well funded. Imagine that the metaphor system underlying English and many other languages was extensively worked out, and that the system of radial categories and contested

concepts was well advanced. What use could be made of this knowledge about our own modes of understanding?

Law. Steven Winter, one the country's most important young legal theorists, has been demonstrating for some years that conceptual metaphor and radial concepts are central to law. The prevailing legal fiction is that the law applies general principles to particular cases, and that the judgments about particular cases follow by logic from the legal principles plus knowledge about the case at hand. In short, law engages in the mythology that legal decisions are strictly literal, that legal categories are defined by necessary and sufficient conditions (rather than radial structures), and that only logic and not metaphor enters into legal decisions.

Winter has argued, to the contrary, that major judicial decisions involve the application of conceptual metaphor not logic. Take for example, the legal metaphor that **CORPORATIONS ARE PERSONS**. This metaphor allows corporations a range of legal rights such as standing in court and the ability to sue. There is, on the other hand, no legal metaphor that **ECOSYSTEMS ARE PERSONS**, and ecosystems are thereby not given rights or standing in court if they are harmed.

Winter has shown that a great many judicial decisions that have come to be taken as precedent are cases where a judge has used a metaphor to extend a body of law from central to noncentral cases. Indeed, just about all Supreme Court cases involve the extensions of some concept radially from central to noncentral cases. What would result if the use of metaphor and radial categories that presently exists in the law were officially recognized as such?

The current legal mythology, Winter observes, permits a great deal of judicial arbitrariness. When judges extend legal categories by metaphor and claim that they are just showing how the case logically fits existing necessary and sufficient conditions, the effect is that judges get to make law outside of democratic procedures. Winter proposes, in an extensive rethinking of standing law, that the concept of standing might be redefined by legislation as a radial category. The existing concept of standing in court would be preserved at the center of the category while metaphorical extensions of the concept of standing could be added by legislation. Such a redefinition of standing, he argues, would eliminate many categories of injustices.

The official recognition of cognitive science and conceptual system research by the legal community could have far-reaching effects in bringing law under more democratic procedures. It would certainly drastically change the nature of legal argumentation, and perhaps even how laws are framed.

Bringing cognitive science into law would also change the most basic understanding of legal theory. Take the question of rights. The traditional liberal view

is that rights are “inalienable” -- essential properties of human beings, part of the essence of what it is to be human. Certain adherents of the Critical Legal Studies movement, however, have noted that, according to Continental philosophy, there are no essences, and hence if rights are essences, there are no rights, and it is thus pointless to fight to defend rights. Their point is to portray law as the arbitrary use of power by judges.

Winter, using results from the study of conceptual metaphor, demonstrates that rights are understood via a collection of conceptual metaphors. But he observes that this does not make them fictional. He notes that metaphors can be made real by social institutions. For example, we have a metaphor in this culture that **TIME IS A MONEY-LIKE RESOURCE** that can be wasted, saved, budgeted, invested, spent, squandered, and so on. Time isn't herently a money-like resource, and many indigenous cultures around the world do not have this conceptual metaphor. But we have social institutions such paying people by the hour, deadlines, and so on that structure society so that this conceptual metaphor fits a wide range of experience in our culture. The institutions make the metaphor real -- it *is* possible that someone could really *waste* an hour of my time, and I have certainly *squandered* many an hour. Similarly, Winter argues that, if we like the idea of rights and want to have them despite the fact that they are metaphorical creations, we have to build and maintain institutions to define and guarantee those rights. That takes work and social activism. Thus, we cannot blithely assume the existence of rights as if they were essences, nor need we give up on them as being fictions. By recognizing their metaphorical character and the fact that conceptual metaphors can be realized through social institutions, we can see why continued activism is required if rights are to be made and kept real.

From these examples we can see that our very understanding of the nature of law at its most fundamental level changes when results about our conceptual system and its embodied nature are taken seriously.

Philosophy Philosophy, since the time of the Greeks, has seen itself as the final arbiter in characterizing concepts and determining the validity of arguments. But the development of conceptual system research within Cognitive Science fundamentally changes the role of philosophy. Cognitive Science has made it an empirical question as to what conceptual systems are like -- and they have turned out to be very unlike what philosophers had surmised. Conceptual analysis is now an empirical matter and Cognitive Science has developed the tools to do analyses of basic-level concepts, prototypes, frames, conceptual metaphors, etc. The empirical results in these areas fit neither the Anglo-American nor the Continental tradition in philosophy. What is needed is a new philosophical tradition that takes into account the results of Cognitive Science. That tradition has begun to be built.

Mark Johnson, the major philosopher to make use of empirical results from Cognitive Science in doing philosophy, is the principal initiator of that new tradition. In his classic *The Body in the Mind*, Johnson argues on empirical grounds that concepts are grounded in the body and extended by imaginative mechanisms such as metaphor, prototypes, frames, etc. This view counters, first, the most common Anglo-American view that meaning lies in the relationship between symbols and a mind-free world -- independent of human psychology. It also counters the most common Continental view (not counting Merleau-Ponty) that meaning is arbitrary and ungrounded. What emerges instead is Johnson's "experientialist" position, a position very much in keeping with results in contemporary neuroscience.

Johnson has also applied Cognitive Science to the traditional questions of ethics. In *Moral Imagination: Ethics in the Age of Cognitive Science*, Johnson argues against the traditional view that moral laws are like axioms in a logic that will either fit or not fit a given situation and will objectively tell one by the application of pure reason what to do in that situation. Instead, he argues, our very concept of morality is based on a collection of conceptual metaphors that are conventional in our culture and that conflict with one another. Citing research by Sarah Taub, Johnson argues that the major metaphor for morality in our culture, as well as in many others, is the metaphor of Moral Accounting.

Moral Accounting works roughly like this: Say I do something to harm you. That puts me morally in your debt -- I owe you something. And if I do something to help you, then you are in my debt, and you owe me something. Debts can be paid according to a metaphorical moral arithmetic: Giving something bad = getting something good; and giving something good = getting something bad. Thus, if I do something to harm you and am morally in your debt, the books can be balanced in two ways: You can do something bad back to me (retribution) or I can do something good for you (compensation). It is also possible to accrue moral credit, to be morally bankrupt, to pay someone back with interest, and so on. Taub demonstrates that there are a number of mutually exclusive moral positions possible within the moral accounting metaphor: compensation, retribution, revenge, turning the other cheek, potlatch, and so on.

Johnson observes that real moral decisions often involve two layers of metaphor -- the choice of a metaphor for morality and, commonly, the choice of a metaphor for understanding the situation at hand. The result, Johnson argues, is that conceptual metaphor is always used in making moral decisions, and that moral dilemmas are real and may have to do with which metaphor should be used to characterize a situation.

Expanding on Johnson's views on ethics, Johnson and I set out to apply cognitive semantics not just to ethics, but to philosophy in general. We reasoned as follows: Each philosophical theory is a miniature conceptual system. Each

philosophical theory would be what we call a “conceptual paradigm” -- a collection of theoretical statements making use of conceptual metaphors, prototypes of various kinds, frames, and so on. Cognitive science has the tools to characterize such conceptual paradigms precisely. Suppose we take philosophical theories as data to be described in terms of the conceptual metaphors, frames, and so on. What would a philosophical theory look like when characterized from the perspective of cognitive science?

What we proposed was a subfield of cognitive science with philosophy as its subject matter: The Cognitive Science of Philosophy. At present, we have done preliminary analyses of Presocratic metaphysics and Enlightenment epistemology and ethics. Here’s what we have found so far: Each philosophical theory uses a collection of metaphors -- typically a consistent subset of those found in the culture at large. Those metaphors characterize a significant portion of the ontologies of the philosophical theories. For example, in Enlightenment faculty psychology, the Society of Mind metaphor is used: The mind is conceptualized as consisting of personified aspects of mind: Reason, Will, Perception, Judgement, and so. Each of these has a job to do. Perception acquires information about the external world, reason calculates what actions will best serve the person’s interests, Will carries out those actions, and so on. Philosophical theories operate within the confines prescribed by the major metaphors of the philosophy, and theorize about those metaphorical entities to form a conceptual paradigm. When a conceptual paradigm uses versions of metaphors taken from our everyday conceptual systems, that conceptual paradigm seems “intuitive”, since it uses metaphors that we already think in terms of.

A particularly clear case of a Conceptual paradigm that seemed intuitive because it made use of everyday metaphors is the philosophy of mind that accompanied the early days of cognitive science when it was thought that the mind was a computer and that thought was symbolic computation of the kind used in classical artificial intelligence. Here are some of the everyday metaphors that went into that conceptual paradigm:

The Thought-as-Object-Manipulation Metaphor: Here thoughts are seen as objects that can be manipulated, e.g., put together, taken apart, rearranged, added to, stored, retrieved, put into words (which are linguistic containers for thoughts), sent to others, and received from others.

The Thought-as-Mathematical-Calculation Metaphor: Here ideas are seen as mathematical objects (numbers), reasoning is addition (“He put two and two together”), considering an idea is including it in the count (“Should we count that?”), inferences are sums (“What’s the bottom line?”), and explaining is accounting (“Give me an account of why that happens. Can you account for this

phenomenon? It just doesn't add up.'').

The Thought-as-Language Metaphor: Here, Thinking is Language Use (Speaking or Writing); Simple Ideas are Words; Complex Ideas are Sentences; and Fully Communicating A Thought is Spelling It Out. The metaphor can be seen in expressions such as:

It's Greek to me. She can't translate her ideas into well-defined plans. I can't hear myself think. He's an articulate thinker. His thoughts are eloquent. What is the vocabulary of basic philosophical ideas? The argument is abbreviated. The theory is spelled out very well. Do I have to spell it out for you. Follow the letter of the law. He's reading between the lines. He's computer literate. I wouldn't read too much into what he says. He's a big question mark to me. Be home by midnight -- period! She's like an open book. Let me make a mental note of that. Note that ... Take note that ... He's the author of the theory that ...

The Mind-As-Machine Metaphor: Here, The Mind is a Machine; Ideas are Products; Thinking is the automated step-by-step assembly of simpler thoughts into more-complex thoughts; Good ideas are products that work; Normal thought is the normal operation of the machine; Inability to think is a failure of the machine to function. Examples illustrating this metaphor are:

He had a mental breakdown. I'm a little rusty today. The wheels are turning now. He's turning out theories at a great rate. That argument doesn't work.

These, when spliced together, yield: **The Mind-As-Computer Metaphor, where a computer is understood as being a MACHINE that CALCULATES by MANIPULATING ABSTRACT OBJECTS in a LANGUAGE.**

The conceptual paradigm of early cognitive science thus seemed intuitive because it was based on metaphors that we already had.

If cognitive science comes to be used to analyze philosophical theories in a systematic way, philosophy will change dramatically. The actual content of much of philosophy depends on certain outdated philosophical views of what concepts are and what reason is. When old views of concepts and reason are replaced by more empirically sound views coming out of cognitive science, then the actual content of much of philosophy will change as well. As I observed, the objectivist tradition in Anglo-American philosophy and the subjectivist tradition in Continental philosophy are both inconsistent with results from cognitive science. Taking cognitive science seriously should result in the abandoning those philosophical views and developing views that are consistent with empirical

results from the cognitive sciences. The very idea that philosophy should have to be consistent with empirical results about the nature of the mind should radically transform philosophy.

Politics and Public Policy Michael Barzelay, a professor at the Kennedy School of Government at Harvard, has made a remarkably innovative use of the study of conceptual systems in cognitive science. After acquainting himself with results about the nature of conceptual metaphor and entailments of metaphorical thought, Barzelay set to analyzing the subject matter that he taught -- governmental bureaucracies.

Barzelay realized that bureaucracies had been conceptualized metaphorically as factories and that bureaucratic officials had been trained to be factory managers. Citizens dealing with the bureaucracy were seen as objects to be processed and people who worked in the bureaucracy as cogs in a machine. The different functions of the bureaucracy were to be assigned to different branches, so that each function could be carried out most efficiently. To be free from bias, the bureaucracy was to function impersonally, and to guarantee impersonal, efficient, machine-like functioning, rules and procedures were handed down for bureaucratic employees to follow precisely.

Barzelay noticed that government bureaucracies had been designed and perpetuated according to a conceptual paradigm governed by a major conceptual metaphor: Bureaucracies are factories. Most of citizens' complaints about government bureaucracies, he realized, were about features that had been designed into the system and consciously perpetuated. The Bureaucratic Paradigm, as he has called it, was not designed to serve the needs of citizens.

Awakened by an insight from cognitive science, Barzelay asked how bureaucracies might be redesigned to serve citizens' needs. What was needed, he reasoned, was a new major metaphor and a new conceptual paradigm based on it. The old metaphor was taken from manufacturing industries. But since then major industries had developed to serve people's needs -- service industries. Barzelay asked what it would be like if bureaucracies were redesigned according to a new metaphor: **A BUREAU OF GOVERNMENT IS A SERVICE INDUSTRY**. He called this the Post-bureaucratic Paradigm. Citizens to be served are to be seen as customers or clients whose needs are to be met. The major metaphors that define the Bureaucratic and Post-bureaucratic Paradigms have very different entailments (Barzelay, 1992, pp.8-9):

- A bureaucratic agency is focussed on its own needs and perspectives. A customer-driven agency is focused on customer needs and perspectives.

- A bureaucratic agency is focused on the roles and responsibilities of its parts. A customer-driven agency is focused on enabling the whole organization to function as a team.
- A bureaucratic agency defines itself both by the amount of resources it controls and by the tasks it performs. A customer-driven agency defines itself by the results it achieves for its customers.
- A bureaucratic agency controls costs. A customer-driven agency creates value net of cost.
- A bureaucratic agency sticks to routine. A customer-driven agency modifies its operations in response to changing demands for its services.
- A bureaucratic agency fights for turf. A customer-driven agency competes for business.
- A bureaucratic agency insists on following standard procedures. A customer-driven agency builds choice into its operating systems when doing so serves a purpose.
- A bureaucratic agency announces policies and plans. A customer-driven agency engages in two-way communication with its customers in order to assess and revise its operating strategy.
- A bureaucratic agency separates the work of thinking from that of doing. A customer-driven agency empowers front-line employees to make judgments about how to improve customer service and value.

Barzelay's book, *Breaking Through Bureaucracy* is a case study of how such a change was actually carried out in the State of Minnesota. Besides Barzelay, others have proposed such a metaphor shift, most notably David Osborne in *Reinventing Government* and Gareth Morgan in *Images of Organization*.

An End of Innocence If the Cognitive Revolution takes hold, it is conceivable that colleges and universities will teach what conceptual systems are like and how to do conceptual analysis. Analyses of conceptual systems may even become commonplace. At that point, a major form of innocence will come to an end. We will be forced to confront the conceptual systems we normally use in minute detail. What was unconscious and automatic in our thought can be made conscious. The metaphors that define conceptual paradigms, whether in academic disciplines or in public life or in interpersonal relations, can be brought to awareness and their entailments revealed. Our implicit conceptual paradigms can be made explicit. Conceptual explicitness can be of great use.

Take interpersonal relations. Each spouse in a marriage has different ways of conceptualizing the marriage metaphorically. Common conceptualizations of

marriage in America include: A journey through life together, a partnership, a home, a constructed object, a life, a haven from the outside world, a struggle, and so on. Each of these has different entailments, and those entailment typically conflict. In short, spouses typically conceptualize their marriages in inconsistent ways. The details of those inconsistencies and their metaphorical sources can be discovered.

Or take public discourse about politics: George Bush convinced the American public that the Gulf War was a Heroic Rescue: A monstrous villain (Iraq) had attacked a weak and innocent neighbor (Kuwait) and a hero (the US) had to come to the rescue. It was a metaphor that hid a great many realities: As Bush said during the 1992 presidential campaign, the war was mainly fought to protect US oil supplies. When Serbia attacked Bosnia, where there was no oil, Bush did not respond with a Heroic Rescue. The Rescue metaphor for the Gulf War hid many other realities: Kuwait was not all that innocent. it had provoked Iraq in a variety of ways. It was also a brutal police state, and our "saving" it did not change that. The US invasion killed 200,000 innocent Iraqis and left more a million more injured or in misery. The war allowed Saddam Hussein to consolidate his power by further suppressing the Kurds and Shiites. Indeed, it was American policy to keep the Kurds and Shiites from getting out from under the rule of Saddam Hussein. The war did not destroy the military power of Saddam's regime. Indeed, the American strategy was designed to keep the monstrous villain reasonably strong. The disparities between the metaphor and what the metaphor hid were enormous.

The American press largely accepted the metaphor. The press did not point out the metaphor and what it hid. Indeed, the press systematically fails to call into question the ways in which the government frames policy. In short, the tools of conceptual analysis have not yet made their way into the press and into public discourse. Cognitive science is not required for training in journalism. Someday it may be. Someday it could become normal for reporters, or experts in cognitive science, to regularly analyse the conceptual framing of major public issues, and what that framing hides. It is conceivable that the tools for conceptual analysis developed in cognitive science could be used as a matter of course in public discourse. It would greatly improve public discourse.

What is a Person?

Every culture has at its core fundamental assumptions about what a person is. In Western culture, since the Greeks, there has been a conceptual paradigm for characterizing a person. That paradigm can be characterized roughly as follows:

This conception of the person is, of course, behind Western religion. In the Judeo-Christian tradition, the locus of consciousness is identified with the soul. Since it is separate from the body, and not subject to physical constraints, it is seen as being able to live on after the death of the body.

That view of the person also lies behind the traditional European distinction between the natural sciences and the humanities: What is subject to physical law can be studied scientifically -- the physical world, including biology. But anything having to do with the human mind is seen as not being capable of being studied scientifically, since it is radically free and not subject to any laws at all. For this reason, Cognitive Science has not been taken seriously within traditional humanistic fields of study.

The traditional Western view of the person is at odds with the fundamental results from neuroscience and cognitive science that I cited above. The Neuro-cognitive Self has neither a separation of mind and body, nor a single locus of consciousness, nor transcendental reason, nor a monolithic consistent world view, nor radical freedom, nor interest maximization, nor an objectivist conception of reality, nor a literalist conception of truth. In short, the conception of the person that emerges from neuroscience and from the cognitive sciences is radically different from the conception that we have inherited from the Western cultural tradition.

Neuroscience and The Self

Let us begin with some very basic results from neuroscience.

-The brain is highly structured from birth; there are hundreds of portions of the brain that either coordinate or perform specialized functions.

-There is also a fair amount of plasticity in the brain, within the boundaries of its basic structure.

-During the first years of life, a significant proportion of the neurons and neuronal connections that a child is born with die off. Which cells and connections remain depends on which ones are used. Thus, experience -- interaction with the world -- shapes our brains in important ways, within the boundaries of innate brain structure.

-Learning involves a change in synaptic weights, adding connections and losing connections not used. When you learn something, your brain physically changes.

-Neuronal systems necessarily form categories. They are not classical categories

The Traditional Western Person

-People are distinguished from animals by having a mind and a capacity for reason.

-The mind is separate from the body. The body is subject to physical constraints; the mind is not.

-There is a single, unitary locus of consciousness with the capacity for reason and will. Reason is conscious.

-Reason is transcendental; it transcends the limitations of the body, applies universally, and for all time. Since concepts are what one uses to reason with, concepts likewise transcend, and are not restricted by, the body.

-In order for transcendental reason to apply to the world, independently of any merely human limitations, the concepts used in reasoning must be capable of fitting what is objectively in the world. Thus, concepts must be literal and reason, being transcendental, cannot make use of mechanisms of the human imagination such as metaphor.

-Perception (especially vision) is veridical; what one perceives (especially what one sees) is, except for relatively rare illusions, really there in the world.

-Conceptual systems are self-consistent and monolithic. Hence, each person has an internally consistent, monolithic view of the world.

-People have free will. That freedom is radical -- completely unconstrained.

-The essence of a person is reason, free will, and a natural tendency to maximize pleasure (or gains) and minimize pain (or losses). Thus, a person is a radically free, rational, maximizer of self-interest.

The separation of mind and body is at the heart of this conception of the person. Descartes claimed that the body and mind were made of different kinds of substance -- the body from physical substance (spatially extended and subject to physical laws), and the mind from mental substance (not spatially extended and not subject to physical laws). The body was subject to physical forces, but the mind was not, and was hence radically free. Similarly, reason transcends everything physical. In short, the essence of the traditional Western person is not physical but mental.

defined by necessary and sufficient conditions. As neural beings, it is impossible for us not to categorize.

-There are concepts that are neurally grounded, at the very least, the color concepts. It has been determined that the neurophysiology of color vision determines the fundamental focal colors and the graded structure of color categories.

-There is no single physical locus of consciousness in the brain. Thought and perception are distributed among many areas of the brain.

-If the Convergence Zone Hypothesis is correct, then there is a biologically-imposed structural constraint on concepts. We would expect complex concepts to integrate diverse lower-level concepts forming gestalts, and we would expect them to combine by simultaneous activation. In short, the brain places constraints on the possibility of conceptual structure. The CZH would also lead us to expect that concepts would be grounded in lower-level perceptual-motor structure.

-At least in the case of color, perception is not veridical, that is, it does not mirror the external world; it is a product of whatever the world is like plus body-and-brain mechanisms that create and structure the categories in which we perceive.

-At least in the case of color, concepts are not arbitrary. They have a physical grounding, and are not merely products of historical accident.

-If the CZH is correct, then concepts in general would be neurally grounded, non-veridical, and nonarbitrary.

These are some of the constraints that neuroscience places on what a person is. Let us now add constraints from Cognitive Science of the sort discussed at the beginning of this paper:

-Most thought is unconscious, automatic, and effortless.

-Basic-level categories and image-schemas have a bodily grounding.

-Abstract concepts are metaphorical projections of more concrete concepts, and conceptual metaphors have a grounding in bodily and cultural experience.

-Knowledge is organized in holistic structures called frames or schemas or idealized cognitive models.

-Conceptual systems are not internally consistent. For example, a single concept

is often characterized by a multiplicity of different and mutually inconsistent conceptual metaphors.

-Categories have complex structures, often radial structures, and make use of many kinds of prototypes.

-Many aspects of conceptual structure are universal; many other aspects are not.

-If the Neural Representation Hypothesis is correct, then concepts are not symbolic in nature but are represented neurally in a distributed fashion such that highly-articulated neural structure will characterize highly articulated conceptual structure.

Put together, all this entails a very different concept of a person:

The Neurocognitive Person

-The mind and body are not separate. Concepts are embodied. They are grounded in perceptual and motor experience and represented neurally. They do not exist without a body and brain.

-Reason is not transcendental; it is bodily in nature.

-Most reason is neither conscious nor deliberate.

-Perception is not a passive mirror of nature; much of what we perceive is constructed by brain mechanisms.

-Much of our conceptualization of the world is metaphorical, much of our knowledge is formulated using metaphorical concepts, and most reasoning involves metaphorical concepts. Therefore, our conception of reality is not strictly literal. We may, however, believe our metaphors, especially when they allow us to function well.

-Since our conceptual systems are neither self-consistent nor monolithic, people do not have consistent monolithic world views.

-We require our conceptual systems to represent knowledge, choose our purposes, and make decisions. But conceptual systems are limited by the possibilities of bodily grounding, of metaphorical extensions, and of the formation of cognitive models -- all of which are limited by the nature of our brains and well as by whatever accidents of history have shaped our conceptual systems.

-Therefore, we are not radically free. Instead, we have only a situated freedom, a freedom to make decisions within certain boundaries given by what we can conceptualize.

-Since most of our thought is unconscious and uncontrolled, we mostly do not consciously choose which aspects of our conceptual systems to use in making most decisions. Most of the time we think and make choices using normal conceptual apparatus such as prototypes and frames. And most of the time there is no possibility of doing otherwise. Kahneman and Tversky have shown, in a wide variety of experiments, that prototype- and frame-based thought in many situations goes against one's self-interest (as defined by theories of self-interest based on probability theory and logic). Therefore, the very nature of our conceptual systems rules out the possibility that we could be maximizers of self-interest. That does not mean that many of us might not try, consciously, to maximize self-interest. It does mean that we could not possibly succeed a significant proportion of the time. Moreover, since our conceptual systems are not consistent and monolithic, it is often the case that there will be no consistent notion of what our self-interest is so that we can maximize it.

Here we have the Neurocognitive Person, with neither a separation of mind and body, nor a single locus of consciousness, nor veridical perception, nor transcendental reason, nor a consistent monolithic world view, nor radical freedom, nor interest maximization, nor an objectivist conception of reality, nor a literalist conception of truth.

At present this is simply a technical, scientific view of what a person is, based on what I believe is the best of what contemporary neuroscience and cognitive science have to offer. As those sciences progress, more details will come in, and perhaps there will be some changes. But I believe that, on the whole, the Neurocognitive Person will stand the test of time, scientifically.

There is however a serious question about whether such a scientific conception of a person can possibly work its way into popular consciousness. There are reasons to be pessimistic:

-This view of what a person is directly contradicts the received cultural view on virtually every point. It is a sophisticated scientific view. Most people don't get educated in such sophisticated scientific views. Moreover, those who do tend not to learn any such views until well after the age of eight to ten, by which time the dominant cultural view has been learned mostly unconsciously and used during the years in which much of conceptual structure has been shaped into the brain. By that time, the traditional view is so entrenched and automatic that it cannot be

completely unlearned.

-Computer technology reinforces the most central part of the traditional view. The computer metaphor for mind comes along with computer technology, and that metaphor has the traditional view of the disembodied mind built into it.

-There are powerful forces in society that depend on the traditional view of the person. Without a disembodied mind, it is hard to conceive of a soul and, hence, of life after death. Western religions have an important stake in the disembodied mind.

Luckily, people do not have consistent, monolithic conceptual systems. It is possible for the view of the Neurocognitive Person to be learned consciously after the traditional view has been learned unconsciously, and there is room for both to coexist with different statuses in our brain/minds.

What Kind of Person?

A culture's view of what a person is has many consequences. It organizes and makes sense of a vast number of beliefs and practices. The traditional Western view of the Person has been subject to considerable criticism in recent years, and has been seen as the locus of many ills. Here are some of the common critiques:

It denigrates the body, since the mind is what is seen as important.

It motivates religions without ecological concerns and with an eye on the future not the present -- religions that care mostly about the soul, which is non-physical, and the afterlife.

Via transcendental reason, it justifies imperialism since nonWestern thought is seen as irrational and primitive, and hence outside of the essence of what is human.

Because of transcendental reason, it is ahistorical and hence fails to understand the importance of worldview difference and worldview change.

It oversimplifies people, seeing them as mere bundles of attributes.

Since it sees itself as having the right concepts for accurately comprehending the world, it sees itself as capable of understanding any issue using its concepts and

of grasping absolute truth.

Since it assumes that all reason is conscious, it ignores most of the unconscious reasoning that is done.

It justifies self-seeking on the grounds that it is an essential human attribute.

In assuming radical freedom, it fails to focus on those aspects of culture and of conceptual systems that limit freedom.

In these critiques, we can see why our concept of what a person is matters so much: It tells us how we should live our lives and how society should be organized. A new concept of a person requires a new way of looking at life and at society.

The Neurocognitive Person contrasts radically with the traditional Western Person:

It is an integrated Person, in that it does not have a mind-body split.

It is a Person that values the body, since it is inseparable from the body.

It is a Person that requires a philosophy of living life to the fullest while the body exists since the Person is identified with the body.

It is not a Person that denies spirituality, at least in certain senses of the term. It is not only consistent with spiritual experience that comes out of meditative traditions, but moreover its essential mind-body connection fits very well with certain of what are sometimes called the Wisdom traditions.

It is humanistic Person, in that it recognizes universals of human experiences and the aspects of conceptual systems based on them.

It is an unself-righteous Person, in that it recognizes that it cannot have a hold on objective truth, that what one takes to be truth depends on understanding via metaphor, and that multiplicities of metaphors naturally give rise to multiple, inconsistent world views.

It is a respectful Person -- respectful of the diversity of human conceptual systems.

It is a historical Person, in that it recognizes that one's personal history is

embodied and importantly shapes one's conceptual system. It also recognizes the contribution of accidents of history in shaping conceptual systems and, hence, in shaping the world view that one inherits as a member of one's culture and generation.

It is a complex Person -- it cannot be pigeonholed by a list of attributes. Each person, instead, has multiplicities -- a complex, multifaceted conceptual system that tolerates multiple world views and aspects of the Self.

It is an open-ended Person, in that it recognizes that there is sufficient plasticity of brain/mind to permit significant change at all stages of life.

It is a creative Person, in that it recognizes the important role of imaginative mechanisms (such as metaphor and cognitive models) in creating understandings of the world and of one's place in it.

It is an ecological Person, in that it is not separable from its environment and cannot be defined apart from its environment.

It is a Person with a meaningful life, with meaning grounded in the body and in the fullness of experience.

I very much like the Neurocognitive Person. I think it is not only the product of what I take to be the best of contemporary neuroscience and cognitive science, but I think it is also a considerable improvement on the traditional view. If our society were adopt such a concept of the Person, a great many things would change.

Whether the new Neurocognitive concept of the Person becomes part of mainstream culture, however, will depend on just how much of conceptual system research in Cognitive Science comes to be known, understood, and assimilated by the general public during the 21st century.