WHO ARE YOU?
THE SELF AS A SYSTEM OF MULTILEVEL INTERACTING MECHANISMS
Paul Thagard
University of Waterloo
pthagard@uwaterloo.ca
Draft 4, November, 2009; comments welcome.

ABSTRACT

This paper proposes a theory of the self as a multilevel system consisting of social, psychological, neural, and molecular mechanisms. This account provides integrated explanations of many phenomena concerning how people represent, control, and change themselves. The multilevel system theory of the self provides a scientific alternative to transcendental and deflationary views favored by many philosophers. The paper identifies more than sixty aspects of the self that divide naturally into nine groups, and provides multilevel accounts of one representative from each: self-concepts, self-consciousness, self-deception, self-presentation, self-criticism, self-esteem, self-affirmation, self-regulation, and self-development. In place of reductionist and holistic approaches to cognitive science, this paper advocates a method of multilevel interacting mechanisms.

WHAT IS THE SELF?

The concept of the self has been theoretically important in philosophy, psychology, and related social sciences, including sociology, anthropology, and political science. The nature of the self is relevant to explaining many interesting phenomena, including: self-abnegation, self-absorption, self-actualization, self-admiration, self-affirmation, self-appraisal, self-assessment, self-awareness, self-blame, self-concepts, self-confidence, self-consciousness, self-consistency, self-construction, self-control,

Despite the centrality of such phenomena in social and clinical psychology, in philosophy of mind, and in related social sciences, there is currently no general, rigorous theory of the self that can provide a principled, organized explanation of them. I propose that the self is best understood as a multilevel system, encompassing mechanisms that interact across four interconnected levels: social, psychological, neural, and molecular. Each of these levels can be understood as a subsystem consisting of environmental influences, component parts, interconnections between parts, and regular changes in the properties and relations of the parts. After specifying in detail the nature of each of these subsystems, I analyze the relations among them, which are far more complex than usually described in reductionist and antireductionist methodologies. My approach rejects both the holistic view that higher levels are autonomous from lower levels and the
individualistic view that higher levels can be entirely explained by mechanisms at lower levels. I introduce the term multilevelism to stand for the view that attention to multiple levels avoids the implausible assumptions and consequences of both individualistic reductionism and holistic antireductionism.

The proposed multilevel theory of the self generates integrated explanations of the full range of dozens of self-phenomena, naturally grouped into three classes concerning self-representing, self-effecting, and self-change. For each of these three, I will describe the most relevant levels of analysis and show how mechanisms at these levels can explain some of the most interesting phenomena identified by researchers in psychology and related fields. I pay particular attention to nine phenomena that are both important and representative of other aspects of the self: self-concepts, self-consciousness, self-deception, self-presentation, self-criticism, self-esteem, self-affirmation, self-regulation, and self-development.

My account of the self is radically different from most philosophical approaches, which tend to be either transcendental or deflationary. Transcendental views, held by philosophers such as Plato, Aquinas, Descartes, and Kant, take selves as supernatural entities – souls – that are not open to mechanistic explanation using the methods of natural science (Organ, 1987). At the other, deflationary extreme, some philosophers have been skeptical of the idea of the self as a determinate kind of thing, proposing instead that the self is just a bundle of perceptions (Hume, 1888), a convenient fiction amounting to a “center of narrative gravity” (Dennett, 1991), or simply a myth (Metzinger, 2009).
In contrast, social and clinical psychologists make substantial use of the concept of the self in their discussions of a wide range of phenomena (e.g. Baumeister, 1999; Alicke, Dunning, and Krueger, 2005). But they have largely shied away from the task of saying what selves are. The multilevel theory is intended to fill this gap while avoiding the metaphysical extravagance of transcendental views and the explanatory impotence of deflationary ones.

MULTILEVEL SYSTEMS

In order to identify a self as a multilevel system, we need to characterize the systems, levels, and mechanisms that constitute selves. My characterization synthesizes and adapts ideas developed by philosophers of science, particularly Bunge (2003) and Bechtel (2008). See Appendix A for terminological clarification and further references.

We can define a system as a quadruple, $<\text{Environment}, \text{Parts}, \text{Interconnections}, \text{Changes}>$, EPIC for short. Here the parts are the objects (entities) that compose the system. To take a simple example, a bicycle is composed of such parts as the frame, wheels, handlebars, chain, and pedals. The environment is the collection of items that act on the parts, which for a bicycle would include people who push on the pedals, roads that interact with the wheels, and air molecules that provide wind resistance to the handlebars. The interconnections are the relations among the parts, especially the bonds that tie them together. In a bicycle, key relations include the physical connections between the chain and the wheels and between the handlebars and the frame. Finally, the changes are the processes that make the system behave as it does, for example the turning of the bicycle’s chain and wheels. There are various ways of describing changes reviewed in Appendix B.
The self cannot be easily decomposed into a single EPIC system. Even a bicycle can be understood at multiple physical levels, for example with the wheel decomposing into various parts such as the hub, the rim, the tube, and spokes, each of which consist of molecules, which consist of atoms, which consist of sub-atomic particles, which may consist of quarks or multidimensional strings. For most purposes, it suffices to consider bicycles at the single level of observable parts such as wheels and pedals in interaction with each other, although an engineer attempting to optimize performance may have reason to work at lower levels, as when nanotechnology is used to design extremely light racing bikes.

To characterize multilevel systems, we can generalize the EPIC idea and think of a multilevel system as consisting of a series of quadruples, with the structure:

\[ <E_1, P_1, I_1, C_1> \]
\[ <E_2, P_2, I_2, C_2> \]
\[ \ldots \]
\[ <E_n, P_n, I_n, C_n> \].

At each level, there is a subsystem consisting of the relevant environment, parts, interconnections, and changes. A later section lays out the relations between environments, parts, interconnections, and changes at different levels.

What are the most important levels for understanding selves? The answer to this question depends on what mechanisms are needed to explain the many interesting self-phenomena. I conjecture that there are four main subsystems required for such explanations, operating at social, psychological, neural, and molecular levels, which are the levels that can be used to explain emotions, consciousness, and other important aspects of thinking (Thagard 2006, 2010). To spell out the claim that the self is a
multilevel system, we need to describe each of the four levels, specifying their parts, interconnections, environment, and changes.

LEVELS OF THE SELF

The Social Self

At the most familiar social level, the set of parts consist of individual persons. Even at this level, there is a hierarchy of additional sublevels of social organization, such as families, neighborhoods, regions, nations, and states, just as at the neural level there are additional levels of organization such as populations of neurons and brain areas. The social parts are influenced by an environment that includes all the objects that people causally interact with, including natural objects such as rocks and lightning bolts, artifacts such as houses and cars, and social organizations such as teams and governments. The interconnections at the social level consists of the myriad relations among people, ranging from mundane perceptual ones such as a person being able to recognize another, to deeper bonds such as being in love, to ones involving several persons such as belonging to the same sports team. Finally, the changes at the social level consists of the many processes of human interaction, ranging from talking to playing games to sexual intercourse. Humans are social animals (Aronson, 2003).

The Psychological Self

At the psychological level, the self consists of individual behaviors and the many mental representations that people apply to themselves and others. The most common representations are personality concepts, such as kind, mean, cheerful, morose, adventurous, cautious, agreeable, hostile, sociable, unfriendly, and hundreds of others. People use such concepts to form rule-like beliefs about individuals, such as that a friend
is optimistic, as well as about social groups, such as that Canadians are courteous. Behaviors are properties of individuals, but mental representations can be considered as parts of them if one adopts an information-processing rather than a commonsense view of the mind.

There are at least three different ways of talking about mental representations, found in everyday conversation, philosophical discourse, and current psychological theories. In everyday conversation, people speak of mental states such as beliefs, emotions, concepts, and ideas, in ways tied to dualist notions that mind is a non-material, supernatural substance. In contrast, my concern is with developing a scientific, evidence-based theory of the self, so I will pay no further attention to everyday concepts of mental entities that derive from unreflective introspections and theistic metaphysics.

Nor will I pay much attention to current philosophical theories of mental representation that view beliefs as propositional attitudes, which are supposed to be relations between persons and abstract entities (propositions) that are the meanings (content) of sentences. The doctrine of mental states as propositional attitudes has been critiqued elsewhere (Churchland, 2007; Thagard 2008, 2010). From the perspectives of folk psychology and standard philosophy of mind, it is odd to describe mental representations such as concepts and beliefs as parts of people. More commonly, concepts and beliefs are spoken about as if they are possessions of people, and the philosophical idea of propositional attitudes understands mental representations as relations between people and abstract entities. Some philosophers claim that to speak otherwise of mental representations is to commit a category mistake.
This objection, however, is scientifically naïve, because the point of theoretical development is to change concepts, not to stick with ordinary ones. Folk psychology has no more claim to truth than folk physics, chemistry, and biology, all of which have long since been superseded by scientific ideas. Since the 1960s, cognitive psychology has developed new, information-processing conceptions of concepts and other mental representations, by analogy to structures and processes used in computers. On this analogy, at least at a crude level, concept and beliefs are like the data structures (e.g. strings, lists, objects, arrays, etc.) that are part of a computer program, which is part of a running computer. Analogously, mental representations can be parts of people, in a way that is even more obviously true from the perspective of the neural level to be discussed below. Cognitive psychology abounds with ideas about what kinds of computational structures might be found in the mind. For example, there are diverse theories about concepts (e.g. Murphy, 2002), and processing theories about non-supernatural propositions (e.g. Anderson, 1983).

Thus, at the psychological level, the self consists of a subsystem where the parts are mental representations such as concepts, schemas, beliefs, attitudes, propositions, rules, situations, analogies, images, and so on – all the kinds of representations found in textbooks in cognitive science (e.g. Thagard, 2005). The environment for these parts consists of all the objects in the world that can be inputs to and outputs from mental processes, including objects in the world and other people. The interconnections of a system of mental representations consists of the relations between them, particularly the bonds that hold them together. For examples, concepts are organized by kind and part-whole relations: the concept bicycle is related to concepts machine and wheel, because a
bicycle is a kind of machine and its parts include wheels. Beliefs have concepts as parts, as when people put the concepts bicycle and heavy together to form the belief that bicycles are healthy.

Folk psychology can tell us nothing about the processes that cause the interactions of mental representations, and philosophical psychology has only limited theories of inference such as ones based on deductive logic. But cognitive psychology over the past 40 years has developed rich ideas about mental processing that apply to a wide range of mental representations, from concepts, to rules, to images. For example, theories of spreading activation among concepts explain many interesting phenomena about memory and language such as priming effects. Rule-based thinking has been modeled by processing systems such as ACT that provide detailed accounts of inferential mechanisms (Anderson, 2007). These theories and their attendant computational models provide mappings from the properties that apply to mental representations at one time and the properties that apply at a later time. Thus cognitive psychology provides accounts of the processes by which concepts, rules, and other mental representations change over time. Increasingly, cognitive theories are being tied to neural processes.

The Neural Self

Characterizing the neural subsystem is relatively straightforward. The most important parts of the brain are neurons, which are cells that also exist in related parts of the nervous system such as the spine. The interconnections of the neural system are largely determined by the excitatory and inhibitory synaptic connections between neurons, although glial cells in the brain and hormonal processes are also relevant (Thagard, 2006, ch. 7). The environment of the neural system is better described at a
smaller scale than the level of whole objects appropriate for the psychological and social levels. For example, photons of light stimulate retinal cells and initiate visual processing in the brain, and sound waves affect the structure of the ear and initiate auditory processing. Thus the environment of the neural system consists of those physiological inputs that influence neural firing. Finally, the changes in the neural subsystem include alterations in firing patterns resulting from excitatory and inhibitory inputs from other neurons, as well as alterations in the synaptic connections (see e.g. Dayan and Abbott, 2001; Eliasmith and Anderson, 2003; O’Reilly and Munakata, 2000).

Folk and philosophical psychology totally ignore the neural level, but in current cognitive science the neural and representational levels are increasingly becoming integrated (e.g. Anderson 2007, Smith and Kosslyn, 2007). I have used the term “representational” to refer to familiar structures such as concepts and beliefs, but the activities of neural populations can be representational too, by encoding features of the external and internal world. As an inert object, a single neuron does not represent anything, but there are special cases where the firing activity of individual neurons can stand for things in the world, for example specific actors such as Jennifer Aniston (Quiroga et al., 2005). More commonly, neural representations are accomplished by the joint firing activity of populations of neurons. We will see that particular self-representations can be performed by populations of neurons that fire in ways that causally correlate with aspects of the self and world.

The Molecular Self

Just as cognitive psychology has drawn increasingly on neuroscience in the past two decades, neuroscience has drawn increasingly on molecular biology. Neurons are
cells consisting of organelles such as nuclei and mitochondria, and the firing activity of neurons is determined by their chemical inputs and internal chemical reactions. Aspects of the self such as personality are influenced by biochemical factors including genes, neurotransmitters, and epigenetic factors that modify the expression of genes.

Genetic effects on behavior are displayed by studies that find higher correlations between some features in identical twins than in non-identical ones, for example in tendencies toward mental illnesses such as schizophrenia. Humans have variation in genes that determine the receptors for more than fifty different neurotransmitters that affect neuronal firing. For instance, there are variations in the gene DRD4 that controls the formation of the D2 receptor for the neurotransmitter dopamine. These variations are associated with behavioral effects such as the personality trait of novelty seeking (Benjamin et al., 1996) It would be naïve, however, to suppose that there are “genes for” particular behaviors, because there is increasing evidence for the importance of epigenetic effects on the operation of genes (e.g. Richards, 2006). Whether a gene expresses a particular protein depends not only on the gene, but also on the attachment of various chemicals such as methyl groups, which are affected by the overall environment of the cells that contain the genes.

In sum, a self is a system consisting of subsystems at four levels – social, psychological, neural, and molecular – each of which includes environment, parts, interconnections, and changes. Before getting to explanation of phenomena, I need to provide a clearer picture of the relations among the various aspects of different levels. In writing of the social, psychological, neural, and molecular selves, I am not taking a
person to consist of four separate selves. Rather, the self is the integration of all four levels, as can be shown by considering the relations among them.

**RELATIONS AMONG LEVELS**

From the EPIC perspective on systems, we need to look in detail at the relations between environment, parts, interconnections, and changes at different levels. The relations between parts are the most straightforward. As a first approximation, we can say that the parts at one level are composed of the parts at the next level down. This relation is most obvious at the intersection of the neural and the molecular levels, as biology makes it clear that the parts of neurons include molecular parts such as proteins and genes. But composition is more complicated in other cases. Does it really make sense to say that mental representations are parts of persons, and that neurons are parts of mental representations?

I already argued that the information-processing idea that representations are parts of people should not be rejected because of the commonsense idea that beliefs are properties of people. Concepts can be parts of people in the same way that data structures are parts of computers loaded with software programs. It also takes some conceptual revision to see neurons as parts of mental representations, which in the early days of cognitive science were largely viewed as functional computational entities not tied to any particular kind of physical instantiation. The rapid development of cognitive neuroscience, however, has made it more natural to think of concepts and mental representations as patterns of neural activity. But are neurons as things – nerve cells – parts of dynamic entities like neural activity, let alone parts of more abstract entities such as patterns?
It is easier to answer this question if we distinguish between occurrent and dispositional aspects of mental representations, following the traditional philosophical distinction between occurrent and dispositional belief. People have beliefs that they are not currently thinking about: five minutes ago, you were probably not thinking that Canada is in North America, but you probably believed it, in the sense that you had a disposition to say yes when asked if Canada is in North America. Once you are asked, the belief becomes occurrent when you are actually thinking that Canada is in North America. Analogously, a spoonful of sugar has the disposition to be soluble in water that makes it dissolve. Sugar has this disposition because of intermolecular forces arising from its chemical structure and that of water. Similarly, a pattern of neural activity occurs because of synaptic connections between members of a neural population. Hence, from the perspective of cognitive neuroscience, a dispositional belief is a pattern of neural connections that, given external and internal stimuli, will lead to a pattern of neural firing. Because a pattern of neural connections is a combination of neurons and their synaptic links with other neurons, it is natural to say that neurons are parts of mental representations in the dispositional sense. It is only a small step to acknowledge that neurons are also parts of patterns of firing activity in neural populations, in the same way that the colored threads in a quilt are part of the pattern on the quilt.

It might seem that this discussion of composition implies or presupposes a simple reductionist view of the self, with molecules as parts of neurons, which are parts of mental representations, which are parts of persons, which are parts of groups. However, this unidirectional, asymmetric ordering does not imply that causality needs to be
similarly unidirectional: we will see later that social processes can causally affect molecular processes.

Now we can consider the relations between environments that operate in the multilevel system of the self. At the extreme, the large objects that influence the social system are very different from the minute ones that influence the molecular system. Within adjacent levels, however, there seems to be much overlap between environments. Large scale objects in the world such as buildings and rivers influence persons (operating at the social level) and mental representations (operating at the psychological level). Such objects also have effects at the neural level, through psychophysical processes of perception, as when light reflects off a building and photons stimulate the retina to initiate a cascade of neural processing. It seems, then, that the relation between levels of environment is sometimes identity, sometimes part-whole (as when the light reflects of the windows of a building), and sometimes a more complex causal process. The complexity of environmental influences derives from the fact that environments are also multilevel systems ranging from microbes to large-scale terrains and climates, with which humans as multilevel systems interact at multiple levels ranging from the cellular to the social.

The third aspect of the EPIC account of systems concerns interconnections, the set of relations that hold between objects, especially the bonds that hold them together. How can we characterize the abstract connection between bonds that operate at one level and bonds that operate at lower ones? Consider a simple physical case. When two pieces of wood are joined by a nail, their bond is the result of physical forces operating at a lower level, connecting the molecules of the nail with the molecules of the two pieces
of wood, where these molecular bonds are in turn the result of subatomic, quantum-mechanical processes. Similarly, for each bond at a higher level in a multilevel system, we should look for a causal process at the next level down that produces it. Higher bonds do not have lower bonds as parts, but rather emerge from causal processes involving lower bonds.

Similarly, in the multiple levels that comprise the self, the bonds at each level are the causal results of processes operating at lower levels. At the social level, groups are formed by bonds between persons that are partly the result of the operations of mental representations at the lower level. For example, when two people become friends, their friendship results from a complex of mental representations that each has about the other, including concepts such as *nice*, beliefs such as “She likes me”, and emotions such as feeling happy when the other person is around.

It is harder to connect the bonds between mental representations with underlying neural processes, because detailed knowledge of the relevant neural mechanisms is still lacking. But for some simple cases such as association between concepts, informed conjectures are possible. There is a bond between the concepts *cat* and *dog*, in that both cats and dogs are kinds of animals that are often pets. Activating the concept *cat* will therefore likely lead to activation of the concept *dog*, in a way that can be understood at the neural level. If the two concepts are both patterns of neural firing, then their association results from synaptic links between the neurons involved in one pattern and the neurons involved in the other pattern, which may include some overlapping neurons and links. Hence the bond between the two concepts that leads to their association plausibly results from the underlying neural structure and activity.
Similarly, the bonds between two neurons – their synaptic connections – are the results of molecular processes that link the axons of the presynaptic neuron with the dendrites of the postsynaptic neuron. Bond relations, like part-relations, seem to be unidirectional and therefore asymmetric: bonds at a higher level result from causal processes at a lower level, but bonds at a lower level are independent of bonds at the higher level. In contrast, the relations between changes at different levels are not asymmetric in this way, as changes at higher levels can cause changes at lower levels (see examples below and Appendix C).

Identifying relations between changes requires considering the parts at both levels, as well as the properties and relations that alter over time. Changes in systems can be described in many ways, using words, diagrams, and mathematical equations (appendix B). How do changes in groups relate to changes to persons, mental representations, neural populations, and molecular configurations? The simplest answer would be the reductionist one that property changes at the higher level always result from property changes at the lower level. Such determinations are often the case, when changes in group interactions result from changes in mental representations that result from neural and molecular changes. For example, consider the social change of John approaching Mary, because she smiled at him, because she was mentally representing him as attractive, because of firing of neural populations in her visual cortex and dopamine-rich nucleus accumbens. Often, therefore, the reductionist picture is correct in portraying molecular changes that cause neural changes that cause psychological changes that cause social changes.
Often, however, causality runs in the opposite direction. When Mary smiles at John, this social interaction is clearly the cause of a course of changes in him that are psychological, neural, and molecular. He perceives her smiling and probably infers that she likes him, which are changes in mental representation that are also neural changes. Then social changes cause psychological, neural, and molecular changes as is evident in may other situations, for example when men react to their sports team winning by increased testosterone levels. More contentiously, I want to claim that psychological changes can cause neural changes, for example when John’s inferring that Mary likes him (a change in mental representation) cause increased neural activity in various brain areas such as the nucleus accumbens. Hence contrary to the reductionist view that causality is always from lower levels to higher, I prefer the interactive view presented in figure 1.

**Figure 1.** Causal relations among different levels. For a more detailed analysis, see figure 3 in the conclusion.

My account of levels in this paper is largely compatible with discussions by philosophers such as Bechtel (2008), Craver (2007) and Wimsatt (2007). Yet Craver and Bechtel (2007) aggressively reject the idea of downward causes. See Appendix C for a defense of the reality of causal relations from higher to lower levels. With these general aspects of the relations between levels making clear the nature of the claim that the self is
a multilevel system, I can now proceed to organize the many self-phenomena and explain select ones as arising from multiple mechanisms.

GROUPING THE SELF-PHENOMENA

The main aim of this paper is to show that conceptualizing the self as a multilevel system makes possible unified explanation of a wide range of important phenomena. The task is daunting, as there are more than sixty frequently discussed topics that I call the self-phenomena. More accurately, each of these topics should be understood as a group of phenomena. For example, there are many experimental findings about self-esteem that should count as distinctive phenomena to be explained, so there are potentially hundreds of findings for which a scientific theory of the self should be able to account.

Fortunately, the task of accounting for all of the self-phenomena, through causal explanations of the large number of empirical findings about them, can be managed by grouping the phenomena according to three primary functions involving the self: representing, effecting, and changing. The representing self encompasses all phenomena that involve different ways in which people depict themselves, either to themselves or to others. The effecting self concerns ways in which people facilitate or limit their traits and behaviors. The changing self concerns phenomena that involve lasting alterations in how people represent and control themselves. All of the more than sixty self-phenomena fall primarily under one of these functional groups, although a few are related to more than one group. Figure 2 summarizes the proposed organization of self-phenomena that is now discussed in more detail.
The Representing Self

Representations are things that stand for other things, as when English speakers use the word “elephant” to represent elephants or a picture of an elephant to represent a particular elephant. Many of the self-phenomena identified in the introduction concern ways in which people represent themselves. They can roughly be divided into three subgroups concerned with (1) depicting oneself to oneself, (2) depicting oneself to others, and (3) evaluating oneself according to one’s own standards.

The most general terms for depicting oneself to oneself are self-knowledge and self-understanding, which seem roughly equivalent. Self-concepts and self-schemata
are both mental ingredients of self-knowledge, serving as cognitive structures to represent different aspects of the self such as personality traits. (See below for a much more detailed account of self-concepts.) Self-interest consists of the collection of one’s personal goals, conscious or unconscious. Self-identity and self-image are also ways in which one represents oneself to oneself, although they may also contribute to how one represents oneself to others.

Several aspects of depicting oneself to oneself assume conscious experience: self-awareness, self-consciousness, self-recognition and usually self-absorption. Such experience is not purely cognitive, as it can also involve prominent affective components such as moods and emotions. I will give a multilevel system account of self-consciousness below. Another set of phenomena that involve depicting oneself to oneself includes self-deception and self-delusion, in which the representation of self is false, discussed in more detail later.

The second sub-group within the group of self-representing phenomena involves depicting oneself to others. In all of self-expression, self-presentation, and self-promotion, the point of depicting the self is to communicate it to others.

The third sub-group of self-phenomena concerns the evaluation of the self, either as an ongoing process or as the product that results from the evaluation. Phenomena concerned with the process of evaluation include self-appraisal, self-assessment, self-criticism, self-monitoring, and self-reflection. There are many phenomena that result from this process, including both general assessments (e.g. self-confidence, self-consistency, self-esteem, self-regard, self-respect, and self-worth) and particular emotional reactions (e.g. self-blame, self-hatred, self-loathing, self-love, and self-pity).
The processes and products of evaluation may have implications for other functions of self that involve effecting and changing.

**The Effecting Self**

The self does more than just represent itself, it also does things to itself, including facilitating its own functioning in desirable ways and limiting its functioning to prevent undesirable consequences. Phenomena involved in the self having a facilitating effect include self-actualization, self-affirmation, self-determination, self-help, self-indulgence, self-motivation, self-possession, self-preservation, self-protection, and self-reliance.

On the negative side, self-evaluation can produce the self-knowledge that unconstrained actions may have undesirable consequences, as in excessive eating, drinking, drug use, and dangerous liaisons. Accordingly, there are a set of important phenomena concerning limits that people put on their own behavior, including: self-abnegation, self-control, self-denial, self-discipline, self-effacement, self-management, self-regulation, and self-restraint, and even self-sacrifice. All of these self-effecting phenomena involve people encouraging or discouraging their own behaviors, but they do not bring about fundamental changes in the self, which is the third and probably rarest function of the self.

**The Changing Self**

Over a lifetime, people change as the result of aging and experience. A small number of self-phenomena concern processes of change: self-construction, self-definition, self-development, self-enhancement, self-maximization, and self-transcendence. The changes can involve alterations in self-representing, when people come to apply different concepts to themselves, and also self-effecting, if people manage
to change the degree to which they are capable of either facilitating desired behaviors or limiting undesired ones. Whereas short-term psychotherapy is aimed at dealing with small-scale problems in self-representing and self-efficacy, long-term psychotherapy may aim at larger alterations in the underlying nature of the self.

The proposed grouping of self-phenomena summarized in figure 2 is not meant to be exhaustive, as there are aspects of self that are described by words without the “self” prefix, such as autonomy, resilience, and personhood, as well as more esoteric terms that do use the prefix. But the diagram serves to provide an idea of the large range of phenomena to be accounted for by a theory of the self. My goal is to show the applicability of the multilevel theory of the self to this range of phenomena, by selecting one phenomenon from each of the nine classes in the bottom of figure 2. It would be tedious to apply the multilevel theory to all sixty or so phenomena, but I take a broad sampling that includes: self-concepts, self-consciousness, self-deception, self-presentation, self-criticism, self-esteem, self-affirmation, self-regulation, and self-development. Each of these has aspects that need to be understood by considering the self as a system that operates at social, psychological, neural, and molecular levels.

I do not mean to suggest that there are three separate selves capable of representing, effecting, and changing, any more than I implied that there are separate social, psychological, neural, and molecular selves. I especially want to avoid the ridiculous suggestion that a person might consist of twelve different selves combining three different functions at four different levels. My goal is to display the unity of the self, not just its amazing diversity. Unification arises first from seeing the
interconnections of the four levels described earlier, and second from recognizing how
the interconnected mechanisms produce all three of the self’s functions.

EXPLAINING THE REPRESENTING SELF

The scientific value of understanding the self as a multilevel system depends on
its fruitfulness in generating explanations of important experimental findings concerning
the various self-phenomena. I will attempt to show the relevance of multiple
mechanisms to understanding six phenomena that are involved in representational aspects
of the self: self-concepts, self-consciousness, self-deception, self-presentation, self-
criticism, and self-esteem. The first three of these primarily involve representing oneself
to oneself.

Self-concepts

People have various concepts that they apply to characterize themselves with
respect to features such as gender, race, ethnicity, nationality, religion, occupation,
hobbies, personality, and physical characteristics. For example, a man might think of
himself as a middle-aged, intellectual, Canadian, father. In social psychology, social
identity theory attempts to understand the psychological basis of group discrimination by
noticing how people categorize themselves in ways that identify themselves with some
groups that are contrasted with other groups that are viewed less favorably (Tajfel,
1974). Much experimental research on the working self-concept has found that different
aspects of the complete self-concept will be accessible at different times depending on
varying social situations (Markus and Wurf, 1987). What is the nature of the concepts
that people apply to themselves, and what are the mechanisms underlying these
applications?
The psychological level of mental representations is clearly highly relevant to understanding concepts including ones about the self. Psychological mechanisms such as priming carried out by spreading activation between concepts explain how different concepts get applied in different situations. For example, people at parties may be more prone to think of themselves as extraverted. Such explanations require also taking into account social mechanisms such as communication and other forms of interaction. Then the causes of applying the concept *extraverted* to oneself include social mechanisms as well as the psychological mechanism of spreading activation among concepts.

Unfortunately, there is no single currently available psychological theory of concepts that can be applied to self-concepts. Debate is ongoing about whether concepts should be understood as prototypes, collections of exemplars, or theoretical explanations (Murphy, 2002). Moving to the level of neural mechanisms provides a way of seeing how concepts can function in *all* the ways that psychologists have investigated – as prototypes, exemplars, and theories, if concepts are understood as patterns of neural activity (Thagard, 2010). Another advantage of moving down to the neural level is that it becomes easier to apply multimodal concepts such as ones concerned with physical appearance. People who think of themselves as thin or fat, young or old, and quiet or loud, are applying to themselves representations that are not just verbal but also may involve other modalities such as vision and sound. The neural level of analysis makes it easier to see how human concepts can involve representations tied to sensory systems (Barsalou et al., 2003).

Once social concepts are understood partly in neural terms, the relevance of molecular mechanisms becomes evident too, because of the important role of affect and
emotion in self-concepts. For most people, thinking of themselves as young and thin carries positive affect, whereas thinking of themselves as old and fat carries negative valence. When such valences are interpreted neurologically, molecular mechanisms involving neurotransmitters and hormones can be applied. For example, the pleasurable feelings associated with young, thin, and other concepts that people enjoy applying to themselves plausibly result from activity in neural regions rich in the neurotransmitter dopamine, such as the nucleus accumbens. On the negative side, negative feelings such as anxiety are associated with activity in the amygdala, whose neurons have receptors for the stress hormone cortisol as well as various neurotransmitters. Hence if we want to understand why people much prefer to apply some concepts to themselves and different concepts to others, it is helpful to consider the molecular mechanisms that underlie emotion as well as social, psychological, and neural mechanisms.

Self-concepts illustrate the complex interactions among multiple levels that were illustrated in figure 1, belying oversimplified views that see causality as only emanating from lower to higher levels or from higher to lower lives. For example, a social interaction such as a job interview can have the psychological effects of applications of particular concepts (e.g. nervous or competent) to oneself. Activation of these concepts consists of instantiation of patterns of firing in neural populations, attended by increases and decreases in levels of various chemicals such as cortisol and dopamine. Changes in chemical levels can in turn lead to social changes, as when high cortisol makes a person socially awkward, producing counterproductive social interactions that then lead to self-application of negative concepts. Under such circumstances, the four levels can comprise an amplifying feedback loop, from the social to the neural and back again.
Self-consciousness

A second kind of representing oneself to oneself is less specific than applying particular concepts. Being self-aware or self-consciousness is a more general representation of oneself, broader than having a self-concept. It involves having a mental experience that may be difficult to express in words, in part because the experience is tied to sensory capabilities such as vision and hearing.

Self-consciousness has social, psychological, neural, and molecular aspects. The social mechanisms are most evident when people feel awkward or embarrassed in difficult social situations. In routine situations, people may be able to operate automatically to accomplish social and other tasks; but, when placed in a more trying situation such as a job interview, they become much more aware of their own mental states and location in a social interaction. This awareness has an identifiable psychological side, as people become alert to their own personal behavior and the way in which they are mentally representing what is happening in their environments, as well as their own beliefs and emotional feelings.

Social environments seem to have played a role in the evolution of self-recognition, which (at least in humans) overlaps with self-consciousness. The ability to identify oneself visually is rare, appearing only in humans (after 18 months), a few apes, dolphins, elephants, and magpies (Prior, Schwarz, and Güntürkün, 2008). Magpies show a high degree of social intelligence needed to compete with other birds for finding memorized hoards of food. Hence self-recognition in magpies and the few other species capable of it is part of a mechanism of social interaction in which animals are able to distinguish themselves from others.
Although no one has yet developed a rich theory of the neural mechanisms of self-consciousness, neural explanations of perceptual consciousness are increasingly being proposed (e.g. Tononi and Koch, 2008). Thagard and Aubie (2008) develop a model of how conscious emotional experience is produced in the brain as the result of many interacting areas coordinated in working memory. These brain areas integrate perceptions of bodily states of an organism with cognitive appraisals of its current situation. Emotions are neural processes that represent the overall cognitive and somatic state of the organism. Conscious experience arises when neural representations achieve high activation as part of working memory. How could self-consciousness operate in this kind of system?

Consider what happens when a person, call her Alice, becomes aware that she is enjoying the warm spring sunshine. At the psychological level, this requires combining several mental representations: of the sunshine, of enjoyment and of Alice herself, as I. In a sentence, this just becomes *I am enjoying the sunshine*. Having a corresponding neural representation requires having patterns of neural activity that represent sunshine, enjoyment, and I that are somehow bound together into a unified representation of the whole situation that involves self-awareness.

In current research on the modeling of neural processes, there are currently two main approaches to understanding how the brain combines such representations: by synchronization of neural firing (e.g. Hummel and Holyoak, 2003) and by convolution of information encoded in patterns (Eliasmith and Thagard, 2001; Eliasmith and Stewart, forthcoming; Thagard and Stewart, forthcoming). Either convolution or synchrony would be sufficient to produce the effect of binding together in a purely neural fashion.
the key representations for self-consciousness, so that the neural equivalent of *I am enjoying the sunshine* is just the convoluted (or synchronized) neural result of integrated activity of the neural populations whose connections and activity encode the information that Joan is experiencing the enjoyable sunshine. Hence it appears that a neural explanation of self-consciousness is becoming feasible.

Brain size and structure seem to be important to self-recognition, as all the mammals who can do it have comparatively large brains. Dolphins and elephants have even more neurons than humans, although not as many cortical neurons. Magpies, like other birds, do not have a cortex at all but they belong to a family—corvids—with a relatively high ratio of brain size to body size, comparable to primates. Corvids have a highly developed forebrain that may be functionally analogous to mammalian cortex, for example in enabling tool use.

Neural explanations of self-consciousness concerning emotional experience will also need to make reference to molecular mechanisms. In discussing concepts, I argued that molecules such as dopamine and cortisol are relevant to understanding emotional thinking, and the same account applies to emotional aspects of self-consciousness. Because feeling good or bad needs to be understood partly at the molecular level, and because a major part of self-consciousness is thinking about how well we are feeling, molecular mechanisms are relevant to understanding self-consciousness. Hence the phenomena associated with conscious experiences of the self require a multilevel, social-psychological-neural-molecular approach.

**Self-deception**
Judgments of self-consciousness and self-recognition are generally accurate: when people report their experiences of self-awareness, we rarely have occasion to doubt them. People’s self-concepts, on the other hand, are not always veridical. People are typically accurate in reporting obvious characteristics such as sex and nationality, but self-attributions of personality traits are open to many kinds of biasing, such as the Lake Wobegon effect, named after the fictional town where all the women are strong, all the men are good-looking, and all the children are above average. Positive illusions abound in people’s inferences about the world and themselves (Taylor, 1989). In the most serious cases, people engage in self-deception, making inferences about their situations that go against the evidence that they themselves hold. In a classic example, the clergyman Dimmesdale in Hawthorne’s *The Scarlet Letter* believes that he is a good minister, despite his adulterous relationship with his parishioner (for a detailed analysis, see Thagard, 2006, ch. 13). Self-deceptions generate representations of self and other aspects of a situation that are distorted because of what people want to believe about themselves and their situations.

Self-deception and more moderate positive illusions can be understood in terms of the psychological mechanism of motivated inference, in which one’s personal goals lead to selection and distortion of evidence (Kunda, 1990). Motivated inference is not just wishful thinking in which people believe whatever they want, but is a more complex process in which evidence and alternative views are disfavored because they do not fit with personal goals. A psychological mechanism for motivated inference has been modeled computationally using connectionist methods in which competing hypotheses
are evaluated in accord with goals as well as with respect to evidence (Thagard, 2006, ch. 8).

But psychological mechanisms are not the only ones relevant to understanding self-deception, which often occurs in social contexts. Understanding why Hawthorne’s clergyman deceives himself requires reference to his overall social situation: despite behavior that he himself considers to be immoral, Dimmesdale still cares about the opinion that other people such as his parishioners have about him. Hence the causes of his self-deception are partly social and should take into account mechanisms involving social interactions such as communication and approval.

Moreover, a richer understanding of self-deception can be gained by moving down to the neural level as well. Some philosophers have been puzzled about how self-deception could be possible, assuming a rational, unified self. But neuroscience paints a very different picture of the self as a product of many brain areas involving billions of interacting neurons with limited access to each other. There are interactions among brain areas such as ones involved in rational deliberation (e.g. the dorsolateral prefrontal cortex) and others involved in emotional assessment (e.g. the nucleus accumbens and amygdala). Hence it unsurprising that people can engage in self-deceptive practices such as failing to recognize situations where their judgments are affected by conflicts of interests (Thagard, 2007).

Molecular mechanisms are also relevant to understanding the neural processes of self-deception. Believing that you have ideal characteristics makes you feel good about yourself, activating the mid-brain reward system dependent on dopamine-driven neurons,
and avoids the kind of stressful social pain that involves other neurotransmitters and cortisol.

The relevance of molecular mechanisms is even more evident in cases of non-veridical self-representation that constitute self-delusion in paranoid schizophrenics. Because schizophrenic symptoms can be reduced by dopamine antagonists such as Thorazine, we know that molecular mechanisms are relevant to understanding misguided self-representations found in people who obsessively and falsely think that they are being persecuted.

**Self-presentation**

The modes of self-representing discussed so far largely concern how one thinks about oneself, although some aspects of self-image and self-identity also sometimes concern how one wants others to think about oneself. Self-presentation is the central phenomenon for representing of oneself to others. It has been extensively been discussed by sociologists such as Erving Goffman (1959) and by social psychologists concerned with processes of impression management (Leary and Kowalski, 1990). I want to show that self-presentation involves multilevel interacting mechanisms.

The social mechanisms involved in self-presentation are obvious: the environment is the overall physical and social context in which people interact, the parts are individuals, the interconnections are interpersonal relationships and communication, and the changes are the transitions in people’s characteristics and relationships. Self-presentation occurs when social situations lead people to act in ways intended to control the impressions that other people have of them.
However, a social explanation of self-presentation behavior is incomplete without an account of the psychological mechanisms that cause people to want to impress others. People have a basic need for relatedness, for belonging to groups of people that they care about (Baumeister and Leary, 1995; Deci and Ryan, 1990). People know that they are more likely to be accepted by others who have a positive impression of them, so it is natural that people care intensely about how they present themselves to others. At the psychological level, the process of self presentation involves a complex of representations that people apply to themselves and to groups that they want to associate with, as well as all the motivations and emotions that contribute to behavior in interpersonal contexts.

Self-presentation is also dependent on neural mechanisms. People naturally fear not being accepted by others, and a variety of studies have found that the social pain of rejection involves some of the same brain areas as physical pain, such as the anterior cingulated cortex (Macdonald and Leary, 2005). On the other hand, being accepted by others produces activity in brain areas associated with positive emotions, such as the nucleus accumbens. Of course, these neural processes are also molecular ones, with dopamine and opioids associated with positive social experiences, and stress hormones like cortisol associated with negative ones. For example, when people have to give a public speech, often a painful instance of self-presentation, their cortisol levels increase, which may even produce behaviors such as excessive nervousness that undermine the effectiveness of their attempts to produce a good impression. Thus self-presentation involves the complex interaction of social, psychological, neural, and molecular mechanisms.
Self-criticism

The third major kind of self-representing is more dynamic than using existing concepts to represent oneself to oneself or to others. People often engage in the process of self-evaluation, which can involve processes such as self-appraisal and self-monitoring, and have as products resulting mental attitudes ranging from self-love to self-loathing. I will discuss self-criticism as a sample process of self-evaluation, and self-esteem as a sample product.

Self-criticism might occasionally be viewed as a purely internal matter, for example if someone decides to undergo a Socratic self-examination for purely philosophical reasons. But self-criticism usually has social causes, most extremely in the Maoist version in which people are coerced into publically confessing their political shortcomings. Less dramatically, people can be pressed into self-criticism by failures in social relations such as conflicts with family members or co-workers. Hence self-criticism occurs as part of a social mechanism of interpersonal interaction and communication.

The psychological mechanisms of self-criticism include those for self-concepts and self-presentation. In addition, self-criticism requires a comparison between aspects of one’s current self and aspects of one’s ideal self. Comparison processes have been studied by cognitive psychologists in connection with processes of similarity and analogical reasoning (e.g. Gentner and Markman, 1994). Social psychologists have discussed people’s needs to find consistency and avoid discrepancies between their actual selves and their ideal selves (Higgins, 1987). Self-criticism has many emotional states
associated with it, including feeling happy when one satisfies the goal of being what one wants to be, and sadness or even despair when one falls short.

The strong emotional causes and concomitants of self-criticism point to the relevance of neural and molecular mechanisms that underlie affect. Self-criticism involves an assessment of one’s current states with respect to the goals that one has for oneself, which suggests that the most relevant theory of emotions would be ones based on cognitive appraisal. But Thagard and Aubie (2008) argue for a neural synthesis of cognitive appraisal accounts of emotions with accounts that emphasize physiological perception. On this view, an emotional state arises from the interaction of brain areas such as the prefrontal cortex performing cognitive appraisal with brain areas such as the amygdala and insula performing perception of bodily states. Then, self-evaluation need not be a purely cognitive process, but can have physiological inputs (e.g. someone’s gut feelings) and outputs. Moreover, these emotional changes are associated with molecular changes of the sorts already discussed, involving levels of chemicals such as dopamine and cortisol. Researchers in the burgeoning field of neuroeconomics are investigating the neural and molecular bases for valuation in decision making, and their results should naturally apply to self-valuations (Glimcher et al., 2009).

Hence self-criticism and other processes of self-evaluation have underlying molecular and neural mechanisms.

Self-esteem

Similarly, the products of self-evaluation such as high and low self-esteem are the result of the full range of multilevel interacting mechanisms. At the social level, many kinds of experiences influence self-esteem, from childhood interactions with parents and
siblings to ongoing interpersonal relations with peers and authority figures. At the psychological level, self-esteem involves the application of self-concepts with positive or negative emotional valence, for example thinking of oneself as a success or failure in important pursuits such as love, work, and play.

I have repeatedly described the neural and molecular underpinnings of self-representations involving emotions, and the account of self-concepts as patterns of neural activity associated with particular kinds of neurochemical activity applies directly to self-esteem. There is some evidence that the causes of low self-esteem may be genetic as well as social (Roy, Neale, and Kendler 1995), which provides another reason for moving down to the molecular level in order to consider how genes affecting neural processing might be involved in self-esteem. Considering self-esteem at the neural and molecular levels may provide explanations for why in some individuals self-esteem is less influenced by life experience than psychological learning theories would explain. For example, not all successful people have high self-esteem, and the exceptions may arise from underlying neural and molecular differences that the psychological level does not capture.

Thus the many phenomena of self-evaluation, ranging from processes such as self-criticism to products such as self-esteem, appear to be explainable from the perspective of the self as a multilevel system.

**EXPLAINING THE EFFECTING SELF**

In addition to the dozens of self-phenomena concerned with self-representation, there are many phenomena concerned with the self attempting to modify its own states and behavior. These self-effecting phenomena fall into two groups, self-facilitating cases
where one attempts to foster positive aspects of oneself, and self-limiting cases where one attempts to prevent the behavioral expression of negative aspects of oneself. I will discuss self-affirmation as an important kind of self-facilitation, and self-regulation as an important kind of self-limitation.

**Self-affirmation**

According to the self-affirmation theory of Steele (1988), people are motivated to maintain the integrity of the self and respond to threats to this integrity by affirming positive aspects of themselves. For example, smokers may react to embarrassments about their habits by asserting that they can quit whenever they want to. One kind of self-affirmation urged by purveyors of positive thinking consists of making self-statements such as “Every day, in every way, I’m getting better and better.”

Self-affirmation operates at the psychological level, as people try to improve their emotional states and their mental representations of themselves. The operation at this level is more complex than self-help books realize, according to a study by Wood et al. (2009). They found that assigning people to say “I am a lovable person” only improved the mood of people who already had high self-esteem. In contrast, for people with low self-esteem, repeating the phrase actually made them less optimistic, perhaps because for them the positive self-statement was so unbelievable that it strengthened rather than weakened their negative view of themselves.

Self-affirmation also operates at social, neural, and molecular levels. Most threats to self-integrity arise in social contexts when the behavior of others leads people to doubt their preferred view of themselves. Hence self-affirmation results from the process of self-evaluation, whose social causes and context I have already discussed.
Self-affirmation also operates at the neural level as it involves applications of concepts such as *loveable* which, as I argued early, can be understood as patterns of activation in populations of neurons. The study of Wood et al. (2009) shows that self-statements can alter positive and negative moods, which plausibly involves alteration of activities of neurotransmitters such as dopamine. Better understanding of the neural and genetic determinants of low self-esteem could provide the basis for explaining why self-affirmation can have negative effects on people with low self-esteem.

**Self-regulation**

The biological aspects of the self are even more obvious in the self-limiting phenomena involved with the need to control or manage undesirable aspects of the self such as excessive desires for food, alcohol, drugs, sex, or inactivity. Such desires are all rooted in neural and molecular mechanisms that must be counteracted in order to overcome self-destructive behaviors such as overeating. I will not attempt a comprehensive account of all the phenomena concerned with limiting the self, but will focus merely on regulations of emotions, which is an important topic in clinical, social, and cross-cultural psychology (Vandekerckhove et al. 2008).

A specific example of emotion regulation, anger management, shows the need for multilevel explanations. The commonly recommended strategies for anger management operate at all four levels: social, psychological, neural, and molecular. Social strategies including expressing concerns with a sympathetic person and moderately communicating with the sources of anger. Humor involving pleasant social interactions can be a potent way of defusing anger. Temporary or permanent removal from anger-provoking social environments can also be helpful.
Psychological strategies for managing anger include the revisions of beliefs, goals, and attitudes. Cognitive therapy aims to help people by changing dysfunctional thinking, behavior, and emotion. Dysfunctional aspects of anger can be addressed by examining whether the beliefs and goals that underlie angry reactions are inaccurate and modifiable. According to the theory of emotions as cognitive appraisals, anger is a judgment that someone or something is thwarting one’s goals, so that anger should be reduced by realization either that the goals are not so important or by revision of beliefs about whatever is thought to be responsible for goal blocking.

Emotions such as anger, however, are not merely cognitive judgments, but also simultaneously involve brain perception of physiological states (Thagard and Aubie, 2008). Hence it is not surprising that anger management techniques include various methods for reducing physiological arousal, such as exercise and relaxation through deep breathing, mediation, and muscle tensing and release. Reducing physiological arousal reduces perception of body states performed by the insula and other brain areas, thereby reducing the overall brain activity that constitutes anger.

In severe cases of anger, pharmaceutical treatments may be useful, including anti-depressants such as Prozac that affect the neurotransmitter serotonin, anti-anxiety drugs that affect the neurotransmitter GABA, and sometimes even anti-psychotics that affect various other neurotransmitters. The onset of anger can also be affected by recreational use of drugs such as alcohol whose effects on brain chemistry are well known. Hence anger management is an aspect of self-regulation that operates at the molecular level as well as the higher ones. In sum, self-effecting phenomena of both the facilitating and limiting kinds are best understood at multiple levels.
EXPLAINING THE CHANGING SELF

Self-effecting phenomena involve local changes and behavior, but there is a final group of phenomena that involve more permanent changes to the self (Brinthaupt and Lipka, 1994). The most psychologically interesting is self-development, as people change in the course of their lives. Major developmental periods include early years when infants and toddlers begin to acquire identities (Bloom, 2004), adolescence when teenagers establish increasing independence from parents, and old age when physical decline imposes new limitations on the self. Each of these periods involves extensive social, psychological, neural, and molecular changes, but I will focus on old age, drawing on Breytspraak (1984) and Johnson (2005).

Social relations and the aspects of the self dependent on them change dramatically as people get older. Major changes can include the completion of child-rearing, retirement from employment, diminishing social contacts resulting from physical disabilities, and loss of friends and family to death or infirmity. These changes can all affect the quantity and quality of social interactions that are causally associated with a person’s behaviors and representations.

At the psychological level, there are changes in processes, representations, and emotions. Cognitive functioning measured by processing speed and short-term memory capability declines steadily from people’s thirties, and more precipitously in their sixties and later. Self-conceptions may be stable in some respects, but often alter in others, as people define themselves increasingly in terms of health and physical functioning rather than work roles. People in early stages of old age tend to be happier than those in middle age, but infirmities can bring substantial difficulties.
Neural causes of changes in the self are most evident in extreme cases like Alzheimer’s disease, when brain degeneration progressively eliminates anything but a minimal sense of self. There are also age-related disorders such as fronto-temporal dementia that can drastically diminish self-effecting phenomena such as self-control. Aging also brings about molecular changes, for example in reduction of levels of hormones such as testosterone and estrogen that affect neural processing. Hence for a combination of social, psychological, neural, and molecular reasons, self-development takes on important directions in old age. Similar observations could be made about other crucial stages of personal development such as adolescence. The changing self, like the representing and effecting self, operates through multilevel interacting mechanisms.

**OBSESSIONS**

The multilevel view of the self is open to objections from many directions. Some philosophers will think that I have slighted phenomenological aspects of the self - what it feels like to be you. They would argue that selfhood is more a matter of ongoing lived experience than the result of multiple mechanisms. My response is that qualitative experiences such as emotional consciousness are in fact amenable to mechanistic explanation, particularly at the neural level (Thagard and Aubie, 2008; Thagard, 2010). Emphasis on raw phenomenology over mechanism typically amounts to hanging on to transcendental views of the self as soul.

A second objection is that my account has neglected insights into the powerful role of embodiment in constituting thinking and personhood. Many philosophers, psychologists, and linguists have made interesting observations about the often neglected role of the body in human thinking (see e.g. Gibbs, 2005), but the multilevel approach is
Brains operate with a continuous flow of information from inside and outside the body, and the inclusion of environment in the characterization of systems at each level shows the compatibility of my account of the self with views that understand cognition as intimately coupled with bodies and the physical and social worlds with which they interact. I reject, however, extreme positions that claim that dynamic embodiment shows that minds do not require mental representations, which are a crucial part of all the self-phenomena. Cockroaches have dynamic embodiment, but they lack selves.

A final objection is that the account of selves as multilevel systems is terminally obscure, bereft of explanatory power. I grant that this account is very broad, but maintain that much of the details are being worked out through characterizing in detail the mechanisms at each level. At the psychological level, there have been decades of work on mental representations and the computational processes that operate on them. At the neural level, investigation of the kinds of high-level cognition relevant to understanding the self is much more recent, but the past decade has brought major advances concerning how brains represent and process information. Psychologists have tended to ignore the molecular level, but increased focus on neural mechanisms is inevitably leading also to increased attention to molecular mechanisms.

What is most obviously lacking in current discussions is much of an understanding of the relations among levels. I have maintained that there are interlevel feedback loops that account for much of the richness and unpredictability of human behavior. But much more research needs to be done to better comprehend the relations among the social, psychological, neural, and molecular levels. Insights from the
growing field of systems biology should be useful here. Any organism is also a multilevel system, and increased appreciation of the relations among bodies, organs, tissues, cells, genes, and proteins should be help to illuminate the analogous relations among the multiple subsystems that constitute the self.

CONCLUSION

So who are you? My answer is that a self – a person – is a complex system operating at four levels, each of which consists of an EPIC subsystem composed of environment, parts, interconnections, and changes. Each level includes mechanisms consisting of networks of parts whose interactions produce regular changes, as summarized in figure 3. Because the interactions in these subsystems typically involve nonlinear dynamics resulting from feedback loops that magnify effects of small differences in initial conditions, the behaviors of such mechanisms are often hard to predict. In particular, the behavior of the parts at each level is typically difficult to predict from the behavior of parts at lower levels. Forecasting is made even more difficult by the existence of causal relations among levels, for example social influences on molecular changes and vice versa. Moreover, at all levels the subsystem interacts with environments that include other complex systems such as climate and ecology, each of which can have changes that are difficult to predict.

The justification for adopting the multilevel system view of the self is that it is superior to alternative accounts in explaining a wide range of phenomena concerning human behavior. Unlike transcendental views of the self as a supernatural soul, the multilevel view understands the self as a natural but highly complex kind of entity, like a state, university, living body, organ, or molecule. Unlike deflationary views of the self
as a fiction, multilevelism maintains that a scientific concept of the self has sufficiently broad explanatory power to justify belief in selves akin to belief in atoms, viruses, fields, genes, organizations, and other important theoretical entities posited by successful sciences.

**Figure 3.** Diagram of the self as a multilevel system. Lines with arrows indicate causality. Thick lines indicate composition.

Our account of the self exemplifies an approach to the social sciences that might be called the method of multilevel interacting mechanisms (MIM). This method is implicit in various creative investigations of human behavior going back at least to the work of Herbert Simon (1962), but it has rarely been aggressively pursued. Simpler approaches, concentrating on one level or at most two, are cognitively simpler and less professionally risky. The cost of simplicity, unfortunately, is inability to explain many of the most important aspects of human behavior, such as ongoing political conflicts, economic crises, and the nature of the self.

The above discussions of many self-phenomena should make it clear that the MIM method is neither reductionist nor holistic. It is not holistic, because I do not
consider higher levels such as the social as independent from or exclusively determining what happens at lower levels. It is not reductionist, because I reject the common picture that causality moves only from lower levels up to higher. Not only do causal mechanisms operate at each level, but higher-level mechanisms can have causal influences on lower-level mechanisms. The parts at higher levels have emergent properties in the non-mystical sense that the properties belong only to parts at that level, not to parts at lower levels or to simple aggregates of those parts (Bunge, 2003; Wimsatt, 2007).

Hence multilevelism is interactive rather than mystically holistic or simplistically reductionist. The justification for this approach should not depend only on its success in making sense of the self, but also in applications to many other important human phenomena, including emotion (Thagard, 2006), creativity (Thagard and Stewart, forthcoming), and conflict resolution (Thagard and Findlay, forthcoming).

The MIM approach to the self potentially has implications for many other problems in philosophy, psychology, and social science. It suggests an understanding of agents as far more complex than is generally assumed in philosophical discussions of autonomy and personhood, in psychological and sociological discussions of identity, and in economic and political discussions of rational choice and power. Moreover, the MIM view of the self can naturally be generalized to consideration of the interacting mechanisms that operate in all social organizations, from families to nations, which are also multilevel systems.

My concern in this paper has been narrower: to make sense of the self by considering it as a multilevel system consisting of interacting social, psychological,
neural, and molecular mechanisms. I have shown the relevance of all of these levels to nine important phenomena: self-concepts, self-consciousness, self-deception, self-presentation, self-criticism, self-esteem, self-affirmation, self-regulation, and self-development. These nine are representative of three general classes (self-representing, self-efficacy, and self-changing) that cover more than sixty self-phenomena important in psychological, philosophical, and sociological discussions of the self. The self is neither simple nor fictional, but can be understood, from a sufficiently rich, multidisciplinary perspective, as a complex system.

APPENDICES

Appendix A: Systems and mechanisms - terminology

The terms “systems” and “mechanism” are widely used in science, but with varying meanings. My EPIC analysis of systems derives from the account of Bunge (2003), who defines a system as a quadruple consisting of <Composition, Environment, Structure, Mechanism>, CESM for short. My terminology is very similar: Composition=Parts, Environment=Environment, Structure=Interconnections, and Mechanism=Changes. The major difference is that I use the term “mechanism” in a broader sense more compatible with current practice in biology, medicine, and the philosophy of science. Whereas Bunge uses the term to mean a totality of processes (changes), my use of the term encompasses the parts and their organization, in keeping with the discussion of Bechtel (2008; Bechtel and Richardson, 1993). On this usage, mechanisms are whole subsystems, not just one aspect of a system as discussed by Bunge. Another terminological variant for describing mechanisms is due to Machamer, Darden, and Craver (2000), who speak of entities and activities rather than parts and
interconnections. I think the differences among these approaches is more terminological than substantive: All are consistent with the insight that advanced science provides explanations by delineating complex systems with nonlinear processes resulting from components interacting to produce changes. Other discussions of mechanisms, levels, and systems in the philosophy of science include Craver (2007), Darden (2006), Salmon (1984), Thagard (1999, 2006), and Wimsatt (2007).

The idea of levels of explanation is common in cognitive science, for example in Simon (1962), Newell (1990), and Churchland and Sejnowski (1992). What I call multilevelism is similar to the explanatory pluralism of McCauley and Bechtel (2001), and to the systemism of Bunge (1996, 2003). It would be nicer to have a term as catchy as “holism” and “reductionism”, but the Greek word for levels, epipedos, would yield the term epipedism, which sounds like a sexual perversion or skin disease.

The term “mechanism” is sometimes used pejoratively as a contrast to complex systems, but it is common in science and engineering to talk about highly complex mechanisms such as those in the immune system, nervous systems, and advanced robots. There are simple and complex mechanisms, just as there are simple and complex systems. My use of the term “system” is broadly compatible with complex systems theory developed in physics and biology, and with “system of systems” engineering. A complex system has nonlinear dynamics resulting from (1) interactions among components internal to the system, (2) interactions with components of external systems operating at the same level of organization, and (3) interactions with components of multiple subsystems operating at lower levels of organization.

Appendix B: Describing change
The changes in a system can be described in many ways: verbal, mathematical, and pictorial. In addition to using ordinary prose to represent changes, scientists often use formulas, as in chemical reactions such as: $2 \text{H}_2\text{O} \rightarrow 2 \text{H}_2 + \text{O}_2$.

Various kinds of mathematical equations can be used describe change, starting with simple ones such as $f=ma$. Differential equations use derivatives to express rates of change. Difference equations describe changes that occur over discrete time steps. Sets of equations define a dynamic system: the processes of change are characterized using variables and equations that generate a state space that captures how variables change over time as functions of other variables.

Pictures can also be useful in describing changes, either by using simple diagrams such as those found in biology textbooks or by using videos available in films and on the Web. There is no single canonical way of describing change, and science portrays systems and mechanisms using a combination of words, other symbols, mathematical equations, and pictorial representations.

We need to consider the relation between the changes in properties at higher levels and the changes in properties at lower levels. If a system can be characterized mathematically, we can think of it in terms of a structure <objects, variables, functions>, where the objects are the parts at that level, the variables have numerical or true/false values, and the functions are mappings from values of the variables at one time to values of variables at another. We can then ask about the relations between the values of variables at one level and the values of variables at other levels. The reductionist view claims that the variables at one level are always and only functions of variables at the next lower level. The holist view claims that the variables at higher levels are usually
independent of the variables at lower levels. The MIM view allows that variables at higher and lower levels can be functions of each other.

**Appendix C: Top-down causation**

Craver and Bechtel (2007) argue that the idea of interlevel causes is mysterious and philosophically suspect. If they are right, then the arrows between levels shown in figures 1 and 3 are wrong or confused. They claim that it is more appropriate to say that the relation between levels is constitution, not causality. They say that changes at one level are not caused by changes at other levels, but are “mechanistically mediated” through the fact that higher levels are composed of lower levels.

Craver and Bechtel claim that common assumptions about the nature of causation preclude the possibility of causal relations between parts and wholes. If causation involves transmission of a mark (e.g. a signal, or energy), and if causes precede their effects, and if causation is asymmetric in that causes produce their effects and not vice versa; then there are no interlevel causes between mechanisms and their components, because mechanisms and their components are not distinct events, objects, or processes.

The examples given of downward causation from the social to the molecular should make it clear why these worries about causation are misplaced. Claims such as that a social insult can cause an increase in cortisol levels are unproblematic on all reasonable accounts of causality, even though they cross levels. On probabilistic accounts, the probability of high cortisol levels given an insult is greater than the probability of high levels without an insult. On manipulation accounts, intervening in a social situation by generating an insult clearly results in the higher cortisol levels. On mark-transmission accounts, the social interaction transmits energy in the form of sound
waves to the hearer, changing the flow of energy all the way down to the molecular level. The social interaction clearly is a distinct event from the raising of cortisol levels and precedes it, even though people decompose into underlying parts. Changes at time t at one level cause changes at time t+1 at another level. This relation is easier to understand if changes are represented by difference equations or movies rather than by differential equations or static diagrams.

Hence there is no need for Craver and Bechtel’s retreat to the view that changes at higher and lower levels are related only by composition and mediation rather than causation.

ACKNOWLEDGEMENTS

This research has been supported by the Natural Sciences and Engineering Research Council of Canada.
REFERENCES


