

**WHO ARE YOU?**  
**THE SELF AS A SYSTEM OF MULTILEVEL INTERACTING MECHANISMS**

*Paul Thagard & Joanne V. Wood*  
*University of Waterloo*  
*Draft 8, July, 2010; comments welcome.*

**ABSTRACT**

This paper proposes an account of the self as a multilevel system consisting of social, individual, neural, and molecular mechanisms. This framework 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 seventy aspects of the self that divide naturally into nine groups, and provides multilevel accounts of important representatives from each, including: self-concepts, self-consciousness, self-deception, self-presentation, self-criticism, self-esteem, self-enhancement, 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, self-criticism, self-deception, self-defeat, self-definition, self-delusion, self-denial,

July 19, 2010

self-denigration, self-deprecation, self-determination, self-development, self-discipline, self-disclosure, self-disgust, self-effacement, self-enhancement, self-esteem, self-expansion, self-experience, self-evaluation, self-expression, self-focus, self-fulfillment, self-handicapping, self-hatred, self-help, self-identity, self-image, self-importance, self-improvement, self-indulgence, self-interest, self-knowledge, self-loathing, self-love, self-management, self-maximization, self-monitoring, self-motivation, self-pity, self-possession, self-presentation, self-preservation, self-promotion, self-protection, self-recognition, self-reflection, self-regard, self-regulation, self-reliance, self-respect, self-restraint, self-sacrifice, self-schemata, self-transcendence, self-understanding, self-verification, self-worth, and so on. Many other important ideas, including autonomy, personhood, and responsibility, are tightly connected with these phenomena concerning the self, which we will call the *self-phenomena*.

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. We propose that the self is best understood as a multilevel system, encompassing mechanisms that interact across four interconnected levels: social, individual, 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, we analyze the relations among them, which are far more complex than usually described in reductionist and antireductionist methodologies. Our 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. We 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 account 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, we 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. We pay particular attention to thirteen phenomena that are both important and representative of other aspects of the self: self-concepts, self-consciousness, self-deception, self-presentation, self-disclosure, self-criticism, self-esteem, self-enhancement, self-protection, self-verification, self-regulation, self-enhancement, and self-development.

Our 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 account 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. Our 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, < Environment, Parts, Interconnections, and 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:

$$\begin{aligned} &\langle E_1, P_1, I_1, C_1 \rangle \\ &\langle E_2, P_2, I_2, C_2 \rangle \\ &\dots \\ &\langle E_n, P_n, I_n, C_n \rangle. \end{aligned}$$

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. We conjecture that there are four main subsystems required for such explanations, operating at social, individual, 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 consists 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 consist of the many processes of human interaction, ranging from talking to playing games to sexual intercourse. Humans are social animals (Aronson, 2003).

### **The Individual Self**

At the individual level, the self consists of personal 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, our concern is with developing a scientific, evidence-based theory of the self, so we will pay no further attention to everyday concepts of mental entities that derive from unreflective introspections and theistic metaphysics.

Nor will we 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 individual 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 generate 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 individual 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). We 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, although 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 D<sub>4</sub> 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, individual, neural, and molecular – each of which includes environment, parts, interconnections, and changes. Before getting to explanation of phenomena, we need to provide a clearer picture of the relations among the various aspects of different levels. In writing of the social, individual, neural, and molecular selves, we are **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?

We 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 individual 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 off 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.

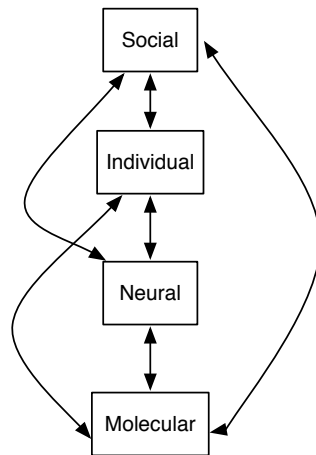
It is harder to connect the bonds between mental representations with underlying neural processes, because detailed knowledge of the relevant neural mechanisms is still lacking. But for some simple cases such as association between concepts, informed conjectures are possible. There is a bond between 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 individual 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 individual, 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 individual, neural, and molecular changes evident in many other situations, for example when men react to their sports team winning by increased testosterone levels. More contentiously, we want to claim that individual 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, we 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.

Our 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 inter-level causes. See Appendix C for a

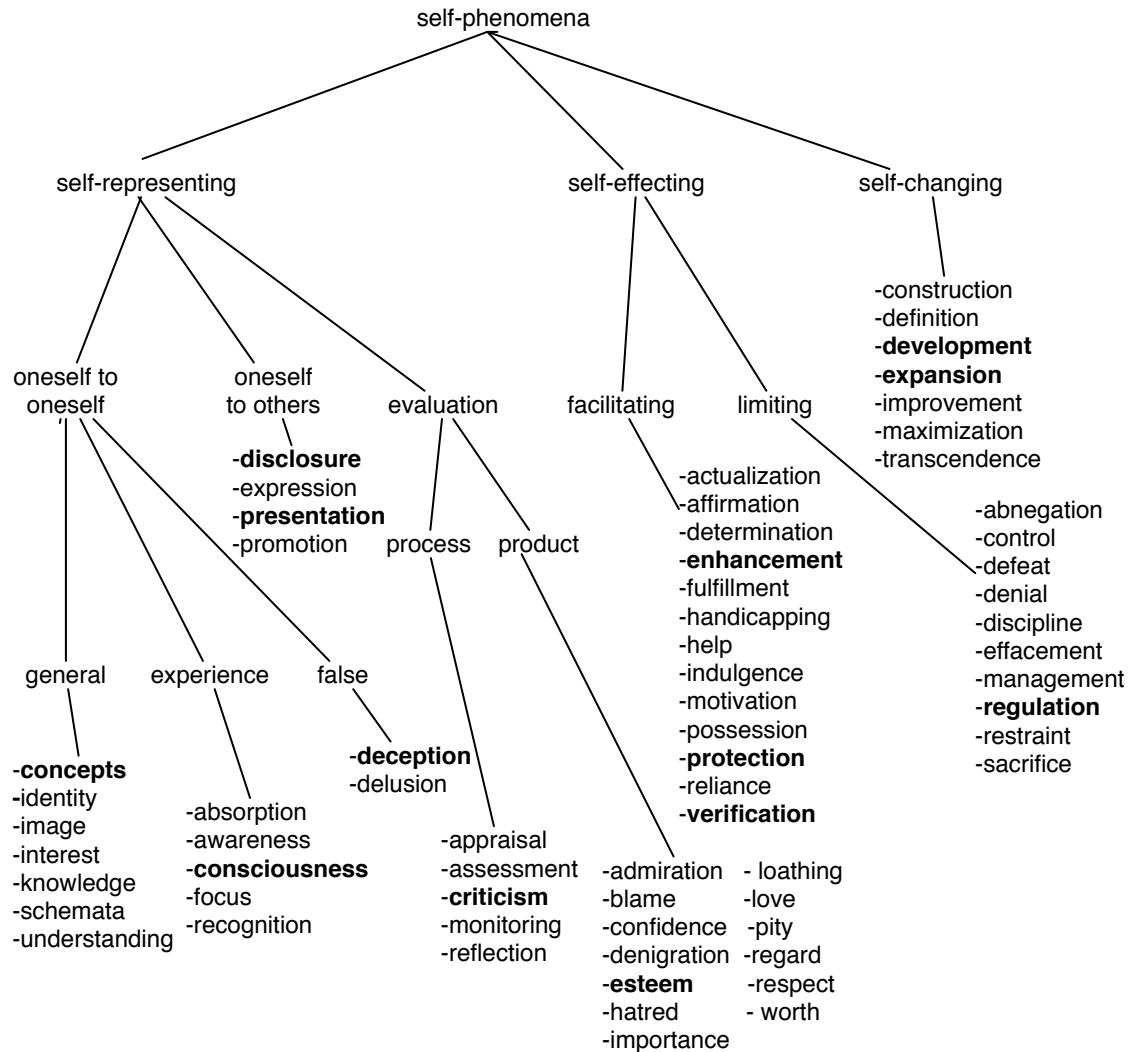
defense of the reality of causal relations from higher to lower levels. With these general aspects of the relations between levels making clear the nature of the claim that the self is a multilevel system, we can now proceed to organize the many self-phenomena and explain select ones as arising from 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 seventy frequently discussed topics that we 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 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.



**Figure 2.** Grouping of many self-phenomena into nine classes. The items in boldface are discussed more thoroughly later.

## 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 in 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. We 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, self-disclosure, 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-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-enhancement, self-help, self-indulgence, self-motivation, self-possession, self-preservation, self-protection, self-reliance, and self-verification.

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-expansion, 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. Our goal is to show the applicability of the multilevel account of the self to this range of phenomena, by selecting at least 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 seventy phenomena, but we take a broad sampling that includes: self-concepts, self-consciousness, self-deception, self-presentation, self-disclosure, self-criticism, self-esteem, self-enhancement, self-protection, self-verification, self-regulation, self-expansion, and self-development. Each of these has aspects that need to be understood by considering the self as a system that operates at social, individual, neural, and molecular levels.

We do not mean to suggest that there are three separate selves capable of representing, effecting, and changing, any more than we implied that there are separate social, individual, neural, and molecular selves. We especially want to avoid the ridiculous suggestion that a person might consist of twelve different selves combining three different functions at four different levels. Our 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.

### **REPRESENTING ONESELF TO ONESELF**

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. We 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-disclosure, self-criticism, and self-esteem. The first three of these primarily involve representing oneself to oneself.

#### **Self-concepts**

Psychologists distinguish between self-concept, which involves *content*—one's thoughts, beliefs, and knowledge about the self—and self-esteem, which involves *evaluation*—evaluation of oneself as good, bad, worthy, unworthy, and so forth. Here we focus on self-concepts.

Self-researchers no longer think of people as possessing a single, unified “self-concept,” but as possessing self-views in many domains (Baumeister, 1999). 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. Moreover, not all content of those various self-views can be held in mind at once. The part of self-concept that is present in awareness at a given time has been called the “working self-concept” (Markus & Kunda, 1986). What is the nature of the concepts that people apply to themselves, and what are the mechanisms underlying these applications?

The individual 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 individual mechanism of spreading activation among concepts. The vast literature on self-concept points to the interplay of the individual and social levels in a myriad of ways. First is research on “social comparison,” which shows that one’s working self-concept depends on the other people present (Wood, 1989). Skinny models can make one feel fat, and unkempt people can make one feel well-groomed. When asked to describe themselves, people tend to list characteristics that make them distinctive in their immediate social setting. A woman in a group of men is especially



likely to list her gender, and a white man in a group of African-American men is especially likely to list his race (e.g., McGuire, McGuire, Child, & Fujioka, 1978).

More permanent aspects of one's social surround can have more consequential effects on self-concept. For example, college graduates' career aspirations depend on their standing relative to their peers at their own college, regardless of the college's standing relative to other institutions (Davis, 1966). A student who earns high grades at institutions where grading is easier tends to have higher career aspirations than an equally qualified student at a more competitive college. This phenomenon has been called "the campus as a frog pond"; for the frog in a shallow pond aims his [or her] sights higher than an equally talented frog in a deep pond (Pettigrew, 1967, p. 257). 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).

Self-concepts are also influenced by the culture in which one lives. Markus and Kitayama (1994) proposed that whereas Westerners have more "independent self-construals," in which the self is autonomous and guided by internal thoughts and feelings, Asians have more "interdependent self-construals," in which the self is connected with others and guided, at least in part, by others' thoughts and feelings. In an interesting demonstration of how various self-concepts can be activated, Ross, Xun, and Wilson (2002) showed that for Chinese-born Canadian students, their Asian, interdependent self-concept was primed when the screen saver on a computer showed Chinese characters, whereas their independent self-concept was primed ...\*. Neural correlates of these different self-construals also have been demonstrated. When East Asian participants were

primed with an independent self-construal, right ventrolateral PFC activity was more active for their own face relative to a coworker's face, whereas when primed with an interdependent self-construal, this region was activated for both faces (see Lieberman, 2010, for references).

Another way that the individual and social levels intersect with respect to self-concept involves the “looking-glass self” or “reflected appraisals”—the idea of that people come to see themselves as others see them. This idea has been prominent in social science for some time (e.g., Cooley, 1902; Mead, 1934), but research in social psychology in the last few decades leads to a different conclusion: People do not see very clearly how others, especially strangers, see them, and instead believe that others see them as they themselves (see Tice & Wallace, 2003, for a review). Instead of others' views influencing one's self-view, then, one's self-view determines how one thinks others view oneself. However, within close relationships, the reflected self may play a greater role in shaping the self-concept (Tice & Wallace, 2003).

Feedback from others can also affect self-concepts, and not just in the way one might expect. For example, although people may think of themselves as more attractive when they have been told they are attractive, people sometimes resist others' feedback in various ways (Swann & Schroeder, 1995). For example, when people with high self-esteem learn they have failed in one domain, they recruit positive self-conceptions in other domains (e.g., Dodgson & Wood, 1998). People are more likely to incorporate others' feedback into their self-views if that feedback is close to their pre-existing self-view than if it is too discrepant (Shrauger & Rosenberg, 1970).

Self-concepts also change with one's relationships. Two longitudinal studies showed that people's self-descriptions increased in diversity after they fell in love; people appear to adopt some of their beloved's characteristics as their own (Aron, Paris, & Aron, 1995). Several studies also indicate that cognitive representations of one's romantic partner become part of one's own self-representation (as reviewed by Aron, 2003).

Interactions with other people also affect the self-concept through a process called "behavioral confirmation," whereby people act to confirm other people's expectations (Darley & Fazio, 1980). For example, when male participants were led to believe that a woman they were speaking to over an intercom was physically attractive, that women ended up behaving in a more appealing way than when the man thought she was unattractive (Snyder, Tanke, & Berscheid, 1977). Presumably, a man's expectation that a woman is attractive leads him to act especially warmly toward her, which in turn brings to the fore a "working self-concept" for her that is especially friendly and warm. Evidence suggests that when people believe that others will accept them, they behave warmly, which in turn leads those others to accept them; when they expect rejection, they behave coldly, which leads to less acceptance (Stinson, Cameron, Wood, Gaucher, & Holmes, 2009). More consequential results of behavioral confirmation are evident in a classic study in which teachers were led to have high expectations for certain students (randomly determined), who then improved in academic performance (Rosenthal & Jacobson, 1968).

So far we have considered social effects on the self-concept. In turn, one's self-concept influences one's judgements of others, in many ways. In his review of this large literature, Dunning (2003) grouped such effects into three main categories. First, in the

absence of information about others, people assume that others are similar to themselves. Second, in their impressions of another person, people emphasize the domains in which they themselves are strong or proficient. Third, when judging others on some dimension, such as physical fitness, people tend to use themselves as a benchmark. Given a man who takes a daily 20-minute walk, athletes will judge him to be unfit, whereas couch potatoes will judge him to be highly fit.

Finally, researchers have examined not only the content of self-concepts, but their clarity. People with clearer self-concepts respond to questions about themselves more quickly, extremely, and confidently, and their self-concepts are more stable over time (Campbell, 1990). Recent research has pointed to social influences on self-concept clarity. For example, clarity of self-concepts regarding particular traits depends in part on how observable those traits are to others (Stinson, Wood, & Doxey, 2008). And when people with low self-esteem receive more social acceptance than they are accustomed to, they become less clear in their self-concepts; the same is true when people with high self-esteem encounter social rejection (Stinson, et al., in press).

In sum, social factors are highly relevant to understanding the operation of self-concepts. At the individual level, we can ask what kind of mental representations are 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, individual, 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 levels. 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, individual, 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 psychologists' examinations of self-consciousness have taken the form of studying "self-focused attention" (e.g., Carver & Scheier, 1981). Self-focus has been studied as both a state—a temporary increase in attention to the self—and as a trait, whereby people differ in their general tendencies to focus on themselves. Self-focused attention is associated with two main categories of effects. The first involves heightening the subjective intensity of whatever aspect of self is salient at the moment. For example, when people are in a self-focused state, they experience their emotions especially intensely (Scheier & Carver, 1977). Negative emotions, in turn, heighten self-focus (Wood, Saltzberg, & Goldsamt, 1990). When provoked to anger, people are more likely to retaliate when they are self-focused than when they are not (Scheier, 1976).

The second main category of effects of self-focus concerns its role in facilitating self-regulatory processes. When self-focused, people are more motivated to adhere to standards. They are less likely to cheat on tests and are more honest in their self-reports (see Carver, Leary, and Tangney, 2003, for references). Trick-or-treating children who are self-focused are more likely than children not self-focused to take only one piece of candy from an unattended bowl (Beaman, Klentz, Diener, & Svanum, 1979). Self-focus leads people to adhere to whatever standards are salient to them in the current situation. For example, self-focus leads people to allocate group earnings according to equity norms when they are salient, but equality norms when they are salient; and leads people who are opposed to stereotyping to refrain from stereotyping, but leads people who

condone stereotyping to stereotype even more (see Carver, Leary, and Tangney, 2003, for references).

Many studies in the vast self-focus literature demonstrate the interplay of the social and individual levels of the self. The allocation-of-earnings findings and the stereotyping findings just described illustrate that self-focus, a psychological state of the individual, can influence the individual's behavior toward the social world. Additional examples are that self-focus can lead people to compare themselves with other people (Scheier & Carver, 1983), and affect their susceptibility to social pressures (see Carver, Leary, and Tangney, 2003, for references). The presence of an audience can induce self-focused attention (Carver & Scheier, 1978).

Although self-focus can be a pleasant state when one is meeting one's standards, it can be highly aversive when one is not. In any case, self-focused attention is "highly evaluative and motivational. People start wanting to change themselves so as to approach their standards, or to escape from awareness of self when change is not possible" (Baumeister, 1998, p. 685). Moreover, high states of self-awareness are associated with depression, anxiety, and other forms of psychopathology (Ingram, 1990).

Possible links between self-focused attention and the neural level have just begun to be identified. Carver, Leary, and Tangney (2003) have pointed to brain research suggesting that the prefrontal cortex plays a role in self-awareness, and EEG research that may help to identify brain activity involved in the processes induced by self-focus. Trait level self-awareness has been associated with medial PFC activity (see Lieberman, 2010).

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, pigs, 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.

Neural correlates of self-consciousness are increasingly being identified, including activity in the medial and ventrolateral areas of the prefrontal cortex (Lieberman, 2007, 2010; Legrand and Ruby, 2009). Although no one has yet developed a rich theory of the neural mechanisms of self-consciousness, neural explanations of perceptual consciousness are 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 individual 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 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, we 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-individual-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. 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.

Empirical demonstrations of true self-deception are rare, presumably because of the difficulty of establishing that an individual both believes something and does not

believe it at the same time (Paulhus, 2007). Psychologists Gur and Sackeim (1979) examined self-deception in the laboratory. They exposed participants to the sound of their own voice amidst a series of voices, and asked them to indicate whether or not each voice was theirs. During this time, participants' Galvanic Skin Response (GSR) was measured. Drawing on the assumption that people typically dislike the sound of their own voice, Gur and Sackeim reasoned that if participants denied that their own voice was theirs, yet their GSR registered recognition, they were deceiving themselves. Many such false denials occurred. False denials increased when participants' self-esteem had been threatened, suggesting that when participants were motivated to avoid self-confrontation, they engaged in more self-deception.

If one permits a looser definition of self-deception, many empirically-established phenomena become relevant. A great deal of evidence, to be reviewed later in the self-enhancement section, suggests that many people, if not most, have inflated views of themselves. As just one example of such evidence, most people view themselves as superior to others. In a sample of high school seniors, zero rated themselves as below average in the ability to get along with others (Gilovich, 1991), and in a sample of adults, 90% believed that they are above-average drivers (Svenson, 1981).

In a highly provocative paper, Taylor and Brown (1988) interpreted such findings as evidence of "positive illusions," because, they reasoned, everyone cannot be better than everyone else. Taylor and Brown (1988) identified three main types of positive illusions: unrealistically positive views of one's abilities and traits, exaggerated sense of personal control over events, and unrealistic optimism about the future. They reviewed a wealth of social psychological research suggesting that these "illusions" are pervasive.

What was most controversial about their paper was that they argued that these positive illusions characterize normal, even high functioning. In contrast to the widespread lay and professional wisdom that contact with reality is the *sine qua non* of mental health, Taylor and Brown (1988) reviewed considerable evidence that it is people with low self-esteem and who are depressed who are more “in touch with” reality. For example, in one study, nondepressed people overestimated their personal control over successful outcomes and underestimated their control over failures, whereas depressed people accurately assessed their degree of control (Alloy & Abramson, 1979).

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 individual mechanisms involving mental representations 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.

Just as self-deception may have social causes, it may have social consequences. Taylor and Brown (1988) argued that positive illusions about the self are good for social relationships, because they facilitate one's ability to care for other people. Other researchers have challenged this idea, however, suggesting that people with inflated self-views have difficulty forming or maintaining friendships because of their self-absorption, excessive self-promotion, narcissism, and lack of empathy (e.g., Colvin, Block, & Funder, 1995; Paulhus, 1998). Carefully-conducted studies have provided evidence for both sides of this debate. Some studies suggest that people with overly positive self-views exhibit off-putting social behaviors, such as bragging and irritability, and are not very well-liked (e.g., Bonanno, Field, Kovacevic, & Kaltman, 2002; Bond, Kwan, & Li, 2000; Colvin et al., 1995), whereas other evidence suggests that they are as well-liked as people with less inflated self-views (Taylor, Lerner, Sherman, Sage, & McDowell, 2003b).

Daniel Goleman (1985) has identified many other connections between self-deceptive processes and the social level, albeit without the scientific scrutiny that backs up the other work we have described in this section. Goleman observed, for example, that many families harbor secrets that they keep from the outside world, and often amongst themselves, in order to preserve the view that they are a happy family. Self-deceptive processes also may operate at the group level, encouraging "groupthink," and at the levels of nations and cultures. Schoolchildren, for example, are sometimes exposed to, at best, a sanitized, minimized version of atrocities that a nation has committed.

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 is 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.

Positive illusions may also have consequences at the molecular level. Taylor and her colleagues have proposed that positive perceptions of the self help people to manage stress, which can result in long-term physical health benefits (e.g., Taylor, Lerner,

Sherman, Sage, & McDowell, 2003a). They tested this idea by exposing participants to stressors in the laboratory. Indeed, participants who were “high self-enhancers” (who rated themselves higher than their peers to the greatest extent) had lower cardiovascular responses to stress and more rapid cardiovascular recovery (but see Shedler, Mayman, & Manis, 1993), as well as lower baseline cortisol levels, than low self-enhancers (Taylor et al., 2003a). These results suggest that people with inflated self-views may experience less wear and tear on their stress regulatory systems than people with more realistic self-views. The potential health benefits of such self-deception are remarkable. A series of studies of people who were HIV-positive or diagnosed with AIDS showed that those who held unrealistically positive views of their course of illness showed a slower decline in their health and lived longer (for references, see Taylor et al., 2003a).

## **REPRESENTING ONESELF TO OTHERS**

### **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 oneself to others. It has been discussed extensively by sociologists such as Erving Goffman (1959) and by social psychologists (Leary and Kowalski, 1990). We want to show that self-presentation involves multilevel interacting mechanisms.

Thirty years of research by social psychologists highlight the interplay of the individual and social levels (Schlenker, 2003). One’s goals, at the individual level, affect the social level. 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 typically want to create a favorable impression. However, people's goals sometimes lead them to present themselves in socially undesirable ways (for references, see Schlenker, 2003). They may try to appear intimidating to generate fear, or self-deprecate to lower others' expectations.

The social level also affects the individual level. One's audience influences one's self-presentation goals. For example, people tend to be more self-aggrandizing with strangers and more modest with friends (Tice, Butler, Muraven, & Stillwell, 1995). Particularly striking evidence of the social level affecting the individual level comes from studies indicating that one's self-presentation to others can influence one's private self-concept (see Schlenker, 2003; Tice & Wallace, 2003). For example, in one study, participants who had presented themselves as extraverted were more likely than those who had presented themselves as introverted to later rate themselves as extraverted, and even to behave in a more outgoing fashion, by sitting closer and talking more to others (Fazio, Effrein, & Falendar, 1981). Such self-concept change does not seem to occur unless one's actions are observed by others (Tice & Wallace, 2003), which again emphasizes the social level. In reviewing the self-presentation literature, Baumeister (1998, p. 705) went so far as to say,

...people use self-presentation to construct an identity for themselves. Most people have a certain ideal image of the person they would like to be. It is not enough merely to act like that person or to convince oneself that one resembles that person. Identity requires social validation.

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.

At the individual 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 cingulate 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, individual, neural, and molecular mechanisms.

### **Self-Disclosure**

An important kind of self-presentation is self-disclosure: a key ingredient in the formation of new friendships and the success of long-term relationships is the extent to which relationship partners reveal their thoughts and emotions to each other (e.g., Reis & Shaver, 1988). What determines whether one self-discloses to another person? The crucial determinant seems to be one's perceptions of the other person's regard—how much one believes that the other person values and likes oneself and will respond in a caring and validating manner (Gaucher et al., 2010; Reis & Shaver, 1988). Studies have indicated, for example, that people are more self-disclosing (Curtis & Miller, 1986), and warmer and friendlier (Stinson et al., in press), when they believe they will be accepted by a new acquaintance than when they do not. Even in long-term, established relationships, people reveal themselves to the extent that they feel secure in the other person's regard (Gaucher et al., 2010).

Revealing oneself has social consequences. Self-disclosure typically leads others to like the discloser, and relationships often develop by a process of gradually escalating self-disclosure by both parties (Aron, 2003). In established relationships, partners who frequently self-disclose to each other experience more love and fulfillment, and are more likely to stay together, than partners who do not (e.g., Berscheid & Regan, 2005). These results again highlight the interplay of the social and individual levels.

## **EVALUATING ONESELF**

### **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 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. We will discuss self-criticism as a sample process of self-evaluation, and self-esteem as a sample product.

The psychological mechanisms of self-criticism include those for self-concepts and self-presentation. People assess their abilities and attributes in order to serve many goals, with two goals perhaps being paramount: to know themselves and to feel good about themselves. People seek accurate self-knowledge so as to navigate their lives successfully, and they seek positive self-evaluations so as to be happy and to cope better with stressors. In either case, people are thought to engage in self-evaluation continually, especially when the self is salient (Stapel & Tesser, 2001).

Self-evaluation requires a comparison between aspects of one's current self and some standard. For example, people may compare themselves with objective standards, or they make "temporal comparisons," in which they compare themselves with their own selves at an earlier time or with a possible "future self" (Taylor, Neter, & Wayment, 1995; Wood & Wilson, 2003). Comparison processes have been studied by cognitive psychologists in connection with processes of similarity and analogical reasoning (e. g. Gentner and Markman, 1994), and the idea that people evaluate themselves by comparing themselves with standards—such as expectations, goals, and values—is prevalent in social psychology. According to Higgins's (1987) theory of self-discrepancies, people harbor standards of an "ideal" self and of an "ought" self—how they hope and aspire to be and how they feel they "should" or are obligated to be. Falling short of one's ideal self leads to sadness, whereas falling short of "ought" standards leads to anxiety and guilt. Another example involves self-regulation (to be discussed in more detail later). Across a

variety of domains, people are thought to measure their behavior against a standard; if it fails to meet the standard, they adjust their behaviour, compare again with standard, adjust again if necessary, and so on (Carver, 2003). So prominent is the idea of standards that Carver (2003) has speculated:

One theme [in the self-regulation literature]...is the principle of attaining consistency between elements of the self and between the self and the actions that it engages in. The human mind appears to have a mechanism that operates to compare mental elements with each other (self and experience, goal and behavior) and to bring them into greater consistency, if that can be done without too much difficulty (p. 194).

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.

Moreover, the standards people employ for self-evaluation often come from the social world. Parents can be a source of ideal and ought guides (Moretti & Higgins, 1999). Graduate students subliminally primed with the scowling face of a faculty member evaluated their own research more harshly (Baldwin, Carrell, & Lopez, 1990), suggesting that standards of significant others can be incorporated into the self.

Other people also can act as standards. In his theory of social comparison, Festinger (1954) posited that people self-evaluate by comparing their opinions and abilities with those of others. Social comparisons may be even more important than Festinger (1959) implied. He said that people make social comparisons when no objective standard is available, yet when both types of standards are available, people sometimes prefer information about how they stand relative to other people over objective standards, and social standards often have more impact than objective standards (see Wood & Wilson, 2003, for a review).

Besides providing standards, the social world also provides feedback that is critical to self-evaluation processes. Feedback may arrive in the form of test scores, parents' praise, and others' disapproving looks. Research suggests that after receiving feedback, people fairly automatically engage in a self-comparison process in which they compare the feedback to their self-conceptions (e.g., Eisenstadt & Lieppe, 1994; Jussim, Yen, & Aiello, 1995). If the feedback fits, the feedback is accepted, but if it does not, it may be only partially accepted, or even rejected. Such self-comparison processes may contribute to the finding that the self-perceptions of people with low self-esteem are more readily influenced than those of people with high self-esteem by unfavorable feedback; such feedback is more consistent with lows' self-views (e.g., Shrauger & Rosenberg, 1970).

People are constrained in the process of self-evaluation, then, by what they already know about themselves. Motivations also limit or steer self-evaluation processes. When one's self-esteem is threatened, motives to self-enhance often become paramount; one may be less open to diagnostic information or even dismiss unfavorable feedback as

unbelievable (e.g., Blaine & Crocker, 1993). A second motivation is self-verification; people also resist changing their self-views (Kwang & Swann, in press). We discuss self-enhancement and self-verification more in later sections.

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. Self-esteem refers to one's overall evaluation of and liking for oneself. People differ from each other in their characteristic levels of self-esteem, which remain quite stable over time, yet people also fluctuate in their self-esteem around their own average levels. "State self-esteem" refers to one's feelings about oneself at the moment.

At the individual 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. When people focus on positive aspects of themselves, their state self-esteem increases (e.g., McGuire & McGuire, 1996).

Considerable evidence indicates that social experiences are central to both trait and state self-esteem. According to attachment theory, people begin to learn about their self-worth as infants, in their interactions with caregivers. If the caregiver is loving and responsive to the infant's needs, the infant develops a model of the self that is worthy of love and responsiveness. If not, the child will develop negative self-models and be anxious in relationships (e.g., Holmes, Cameron, and Baldwin, 2005; Shaver, Collins, & Clark, 1996). We have already discussed how social comparisons can influence one's self-concept; comparisons with other people also can boost or deflate one's self-esteem (Wood, 1989).

Social acceptance may be the chief determinant of self-esteem. Leary's sociometer theory proposed that the very existence of self-esteem is due to the need to monitor the degree to which one is accepted and included by other people (Leary & Baumeister, 2000). Indeed, the more people feel included by other people in general, as



well as accepted and loved by specific people in their lives, the higher their trait self-esteem (Leary & Baumeister, 2000). Numerous experimental studies indicate that rejection leads to drops in state self-esteem (e.g., Leary, Haupt, Strausser, & Chokel, 1998; Wood, Heimpel, Manwell, & Whittington, 2008). Interpersonal stressors in the everyday lives of university students are associated with declines in state self-esteem (Stinson, Logel, Zanna, Holmes, Cameron, Wood, & Spencer, 2008). In contrast, being in a long-term relationship with a loving partner can raise the self-esteem of people with low self-esteem (Murray, Holmes, & Griffin, 1996).

The connection between the individual and social levels of self is also evident in research on how individuals' self-esteem-related goals influence their social lives. A vast social psychological literature reveals that motivations to maintain, protect, or improve self-esteem can, for example, guide how people present themselves to others (e.g., Baumeister, Tice, & Hutton, 1989), lead people to compare themselves with others who are less fortunate so as to boost their own spirits (Wood, Taylor, & Lichtman, 1985), and lead them to stereotype other people in order to feel better about themselves (Fein & Spencer, 1997; Sinclair & Kunda, 2000). Three self-esteem related goals—self-enhancement, self-protection, and self-verification—will be discussed in more detail later.

We 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. Self-esteem researchers have not examined links between the individual and neural levels of self, but self-esteem is connected with depression, which has been

examined at the neural level. Depression and self-esteem are substantially inversely correlated (e.g., with  $r$ s reaching  $-.60$ s and  $-.70$ s; Watson, Suls, & Haig, 2002); low self-esteem is even one of the symptoms of depression. Depression is well known to have neurotransmitter correlates and to be associated with brain changes.

Evidence is mounting that social acceptance and rejection are accompanied by changes at the neural level (Lieberman, 2010). Neuroimaging studies suggest that the social pain caused by rejection involve the same brain areas as does physical pain (namely, dorsal ACC activity), and signs of social acceptance have been associated with ventral striatum activity (reviewed by Lieberman, 2010).

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. There is also evidence of how the operation of the molecular level underlies how people who vary in self-esteem differ in their responses to stress. Taylor et al. (2003b) found that people who had positive self-appraisals had lower cardiovascular responses to stress, more rapid cardiovascular recovery, and lower baseline cortisol levels than people with negative self-appraisals. Furthermore, additional research by Taylor and her colleagues links these findings with the neural level. Participants with greater psychosocial resources, including higher self-esteem along with other variables such as optimism and extraversion, exhibited lower amygdala activity during threat regulation, which appeared to account for their lower cortisol reactivity (Taylor, Burklund, Eisenberger, Lehman, Hilmert, & Lieberman, 2008).

Further evidence suggestive of the interplay of three levels—social, individual, and molecular—has emerged recently. Stinson et al. (2008) proposed that lower self-esteem (individual level) predicts health problems (molecular level), and that poor-quality social bonds (social level) explain this association. Two studies of university students yielded evidence consistent with this view that low self-esteem led to interpersonal problems, which in turn led to acute drops in self-esteem, which then exacerbated the interpersonal problems, which in turn resulted in health problems (e.g., missed classes due to illness and visits to the physician).

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 (e.g., Baumeister, Campbell, Krueger, & Vohs, 2003), and the exceptions may arise from underlying neural and molecular differences that the individual level does not capture. The next section discusses attempts to foster self-esteem through self-enhancement.

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 in which one attempts to foster positive aspects of oneself, and self-limiting cases in

which one attempts to prevent the behavioral expression of negative aspects of oneself. We will discuss self-enhancement as an important kind of self-facilitation, and self-regulation as an important kind of self-limitation.

### **Self-enhancement**

Self-enhancement, “the motive to develop and maintain a favorable self-image” (Alicke & Sedikides, in press, p. \*), has been a dominant topic in the social psychological literature for decades. Evidence that people are motivated to maintain a positive self-view is “overwhelming” (Tesser, 2003, p. 279). Self-enhancement has been seen as a motivation guiding all facets of human behaviour, with some researchers concluding that it is *the* paramount motive, overriding other goals such as self-accuracy and self-consistency (e.g., Baumeister, 1998; but see Kwang & Swann, 2009).

Research has identified many strategies of self-enhancement. To cope with failure, for example, people may attribute the failure externally (e.g., say the test is unfair), minimize the failure, focus on other positive aspects of themselves, derogate other people, or make “downward comparisons”—compare themselves with others who are inferior (e.g., Blaine & Crocker, 1993; Dodgson & Wood, 1998). Over and over again, research has found that the people who engage in such self-enhancement strategies are dispositionally high in self-esteem, rather than low in self-esteem (e.g., Blaine & Crocker, 1993). This self-esteem difference may occur because people with high self-esteem (HSEs) are more motivated than people with low self-esteem (LSEs) to repair unhappy moods (Heimpel, Wood, Marshall, & Brown, 2005); or because HSEs are more motivated than LSEs to feel good about themselves (e.g., Baumeister et al., 1989); or

because LSEs are equally motivated to self-enhance, but cannot as readily claim or defend a positive view of themselves (e.g., Blaine & Crocker, 1993).

One self-enhancement strategy deserves special mention because it is a mainstay of self-help books and the popular press: “positive self-statements.” People facing a stressor, cancer patients, and people chronically low in self-esteem are encouraged to say to themselves such things as, “I am a beautiful person” and “I can do this!” Despite the popularity of positive self-statements and the widespread assumption that they work, their effectiveness was not subjected to scientific scrutiny until recently. Wood, Perunovic, and Lee (2009) found that repeating the statement, “I am a lovable person” improved people’s moods only for those who already had high self-esteem. For people with low self-esteem, the statement actually backfired, worsening their moods and their feelings about themselves.

A strikingly different self-enhancement strategy is “self-affirmation” (Steele, 1988). As studied by social psychologists, self-affirmation does not refer to saying positive things to oneself, but to much more subtle methods involving the expression of one’s values. Self-affirmation strategies have included writing a paragraph concerning a value one cherishes (e.g., politics, social connections), or even merely completing a scale highlighting such values. Such strategies seem to be self-enhancing in that they reduce defensiveness (e.g., Crocker, Niiya, & Mischkowski, 2008; Spencer, Josephs, & Steele, 1993), reduce stereotyping (Fein & Spencer, 1997), make people more open to self-evaluation (Spencer, Fein, & Lomore, 2001), and can substitute for other methods of self-enhancement (e.g., Wood, Giordano-Beech, & DuCharme, 1999).

Although self-enhancement may seem to be a private matter, operating at the individual level, the social level is clearly influential. Most threats to self-esteem arise in social contexts when feedback from others or others' behavior of others leads people to doubt their preferred view of themselves, or to feel devalued or rejected. Hence self-enhancement results from the process of self-evaluation, whose social causes and context we have already discussed. In addition, self-enhancement processes may enlist the social level. Some of the self-enhancement strategies identified above, such as downward comparisons and derogating other people, involve using the social realm to boost oneself at the individual level. Another example comes from research on the triggers of stereotyping. Fein and Spencer (1997) showed that after failing, people were especially likely to seize on a stereotype of Jewish women. Similarly, after receiving negative feedback, people derogated the person who delivered the feedback, if that person was a woman rather than a man (Sinclair & Kunda, 2000). Other social strategies of self-enhancement can include being boastful and overconfident (e.g., Colvin, Block, & Funder, 1995), helping others (e.g., Brown & Smart, 1991), and aggressing against others (Twenge & Campbell, 2003). People may also enhance themselves through their group memberships and social identities (Banaji & Prentice, 1994). Self-enhancement research, then, reveals links between the individual and social levels of self because the social world often elicits the need for self-enhancement, and certain self-enhancement strategies involve the interpersonal realm. In addition, because self-enhancement can encourage or diminish stereotyping, aggression, and prosocial behavior, self-enhancement clearly has many potential social consequences.

Self-enhancement also operates at the molecular level, as shown by a study of self-affirmation. Participants who engaged in a values-affirmation task before they faced a stressor had lower cortisol responses to stress than did participants who had not engaged in values-affirmation (Creswell, Welch, Taylor, Sherman, Gruenewald, & Mann, 2005).

Self-enhancement also operates at the neural level as it involves applications of concepts such as *loveable* which, as we argued early, can be understood as patterns of activation in populations of neurons. The study of Wood, Heimpel, Manwell, and Whittington (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 positive self-statements can have negative effects on people with low self-esteem.

### **Self-protection**

Social psychologists studying self-esteem distinguish between self-enhancement and self-protection: Self-enhancement involves striving for or enhancing positive self-views, whereas self-protection involves avoiding or minimizing negative self-views (Alicke & Sedikides, in press). Baumeister, Tice, and Hutton (1989) proposed that people with high self-esteem focus on self-enhancement—taking risks to draw attention to their skills and talents—whereas people with low self-esteem focus on self-protection—seeking to *avoid* drawing attention to their deficiencies. LSEs "focus not on their good points but on trying to minimize their weaknesses" (Schlenker, Weigold, & Hallam, 1990, p. 856). This view of HSEs as self-enhancing and of LSEs as risk-averse and self-protective is supported by research on decision-making (Josephs, Larrick, Steele, &

Nisbett, 1992), achievement (Tice, 1991), and social comparison (Wood, Giordano-Beech, & DuCharme, 1999; Wood, Giordano-Beech, Taylor, Michela, & Gaus, 1994).

The interplay between the individual and social levels is highlighted in evidence that LSEs' tendency toward self-protection characterizes their approach to interpersonal life (see Forest & Wood, 2010, for a review). For example, LSEs tend to use indirect strategies to strike up romantic relationships; whereas HSEs may boldly ask a stranger for a date, LSEs are more likely to wait for the other to show interest. LSEs also tend to be more inhibited than HSEs in revealing themselves to others through expressing their emotions or self-disclosing. And considerable research by Murray and Holmes (2009) indicates that, even in their close relationships, when they feel insecure, LSEs distance themselves from their romantic partners, whereas HSEs seek closeness or engage in other relationship-promoting behaviors.

### **Self-verification**

People are motivated not only to feel good about themselves or to avoid feeling bad about themselves; they also wish to hold self-views that are stable. Swann's self-verification theory proposes that people try to confirm and maintain their self-views, even if those self-views are unfavorable, because stable self-views lend clarity and predictability to their lives (Swann & Schroeder, 1995). Stable self-views are also thought to smooth social life, by making individuals more predictable to each other (Swann, in press).

Once again, considerable evidence links the individual and social levels. Self-verification motives guide social behavior. People communicate their self-views to others through their clothes, possessions, and actions (Swann, in press). Many studies indicate



that people seek information from others that confirms their self-views and resist information that does not (reviewed by Swann & Schroeder, 1995). For example, depressed university students are more likely to solicit unfavorable feedback from their roommates than are nondepressed students (Swann, Wenzlaff, Krull, & Pelham, 1992). Given a choice, people also want interaction partners and roommates (e.g., Swann & Pelham, 2002; Swann, Stein-Seroussi, & Giesler, 1992), and even may be more committed to spouses (see Swann, Chang-Schneider, & Angulo, 2008; but see Murray et al, 1996), whose views of them match their own self-views.

### **Self-regulation**

Although self researchers were long preoccupied with the topics of self-concept and self-esteem, “during the 1980s they came to appreciate that self-regulation is one of the most important functions of the self” (Gailliot, Mead, & Baumeister, 2008,, p. 474). Self-regulation concerns how people pursue their goals or try to control their own behavior, thoughts, or feelings. An idea discussed earlier in the section on self-evaluation—that people continually compare themselves with standards—is central to many theories of self-regulation (e.g., Carver & Scheier, 1981). Such theories posit that when people experience a discrepancy between a standard and their own standing (behavior, thoughts, or feelings) on the relevant dimension, they deliberately or even automatically attempt to reduce that discrepancy. They can reduce the discrepancy in one of three ways. They can try to adjust their behavior (or thoughts or feelings) so that it meets the standard, change their standards, or exit the situation. Self-regulation is successful when the discrepancy is eliminated or reduced (e.g., Carver & Scheier, 1990).

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. We 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).

Next we discuss three main foci of self-regulation research in recent years: goal pursuit, emotion regulation, and “ego-depletion”—how exercising self-control in one domain diminishes one’s capacity to do so in a second domain.

*Goal pursuit.* Research on “social comparison” establishes a basic connection between the individual and social levels; to meet such goals as self-evaluation, self-improvement, and self-enhancement, individuals compare themselves with other people (Wood, 1989). In this case, other people serve as the standards for meeting one’s goal progress.

Other people can even influence which goals we adopt. Fitzsimons and her colleagues have found that observing a stranger’s goal-directed behavior can lead people to pursue the same goals themselves, or to synchronize their goal pursuits with others, with interesting consequences. For example, people who observe others fail work harder, and people who observe others succeed take it easy (McCullough, Fitzsimons, & Chua, 2010). Even being in the presence of someone who was a stranger a few minutes before, but who shares similarities such as tastes in movies, can lead one to adopt the other’s

goals as one's own (Walton, Cohen, Cwir, & Spencer, 2010). Such effects can even occur subconsciously. For example, when participants who had a goal to achieve to please their mother were primed with their mother, they outperformed control participants on an achievement task (Fitzsimons & Bargh, 2003).

One's own goals also affect one's relationships with others. People draw closer to others who are instrumental in helping them to progress toward their goals, and distance themselves from others who do not promote such progress (Fitzsimons & Shah, 2008). People seem to cultivate a social environment for themselves that promotes their goals, especially when their progress toward their goals is poor (Fitzsimons & Fishbach, 2010).

*Emotion regulation.* Research on emotion regulation—which concerns how people try to manage their emotional states—has amply demonstrated the interplay between the individual and social levels. For example, people try to adjust their moods in preparation for an upcoming social interaction, according to the social requirements expected (Erber & Erber, 2000). In addition, social events affect one's emotion regulation; rejection experiences appear to lead people with low self-esteem to feel less deserving of a good mood, which in turn dampens their motivation to improve a sad mood (Wood, Heimpel, Manwell, & Whittington, 2009).

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, individual, 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, meditation, 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 antidepressants 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.

*Self-control.* “Ego-depletion” studies demonstrate that when people override their emotions, thoughts, impulses, or automatic or habitual behaviors, they have trouble doing so a second time (Baumeister, Vohs, & Tice, 2007). For example, in one study, research participants had to resist freshly-baked chocolate-chip cookies and could eat only radishes instead. When they then faced an impossible puzzle, they gave up more rapidly than participants who had not been required to resist the tempting cookies (Baumeister, Bratslavsky, Muraven, & Tice, 1998). In another study, participants who were asked to suppress certain thoughts later had more trouble resisting free beer than did control participants, even when they expected to take a driving test (Muraven, Collins, & Nienhaus, 2002).

Ego-depletion research has shown connections between the individual and social levels in two ways. First, difficult social interactions can deplete one’s self-regulatory resources (Vohs, Baumeister, & Ciarocco, 2005). Interracial interactions, for example, can be taxing if one tries not to appear prejudiced. Richeson and Shelton (2003) found that after prejudiced white participants interacted with a black participant, they performed more poorly on a cognitive control task, compared to participants who interacted with a white participant or participants scoring low in prejudice. Social interactions also can be depleting if one is required to engage in atypical self-presentation, such as being boastful to strangers (Vohs et al., 2005). And in yet another example of the harmful consequences of social rejection, studies have indicated that it too can impair self-regulation (see Gailliot et al., 2008, for references).

Second, ego-depletion, in turn, makes it difficult to navigate social interactions. Participants who had engaged in previous acts of unrelated effortful self-regulation later were more egotistical in their self-descriptions and less able to choose topics for discussion with a stranger that were appropriate in their level of intimacy (Vohs et al., 2005). Engaging in self-control behaviors, such as dieting, seems to make it harder for people to respond constructively to a romantic partner's bad behavior (Finkel & Campbell, 2001). Self-regulatory depletion also may encourage sexual infidelity and acts of discrimination (see Gailliot, 2008, for references). Successful self-regulation, then, may smooth one's interpersonal interactions and make one's close relationships more harmonious.

*Self-regulation and the neural level.* People who have sustained damage to the prefrontal cortex exhibit various self-regulatory deficits, such as impulsivity and poor judgment (see Gailliot, 2008, for references). The anterior cingulate is involved in tasks that deplete self-regulatory resources via the coordination of divided attention, and the dorsolateral prefrontal cortex affects the activation, maintenance, and modification of goal-directed responses (Baumeister, Vohs et al. (2003). Attempts at self-control recruit a network of brain regions including the lateral and posterior dorsomedial prefrontal cortex (Lieberman, 2010). The consensus across thirty neuroimaging studies of emotion regulation in particular is that right ventrolateral PFC and left ventrolateral PFC activity are involved. Other areas also are implicated, including the presupplementary motor area, the posterior dorsomedial PFC, left dorsolateral PFC, and rostral ACC, and their involvement appears to depend on whether the emotion regulation is intentional or incidental to the participants' task (Lieberman, 2010).

Research by Richeson and her colleagues (Richeson, Baird, Gordon, Heatherton, Wyland, Trawalter, & Shelton, 2003) elegantly links the neural, individual, and social levels of self-regulation. They found that for White participants who held especially negative unconscious attitudes toward Blacks, interacting with a Black person led them to perform poorly on a subsequent self-regulatory task. This effect was mediated by the extent to which these White participants' dorsolateral prefrontal cortex was activated while they viewed Black faces (in a separate session).

*Self-regulation and the molecular level.* Glucose from the bloodstream fuels the brain, and the brain regions involved in self-regulation may be especially susceptible to fluctuations in glucose (Gailliot, Baumeister, DeWall, Maner, Plant, Tice, & Schmeichel, 2000). Gailliot, Mead, and Baumeister (2008) summarized several lines of research that suggest that:

Problems with glucose or low glucose have been linked with [criminal behavior and] several other signs of poor self-regulation, including aggression, impulsivity, deficits in attention control and emotion regulation, troubles coping with stress, and difficulties quitting smoking (p. 481).

In a series of studies, Gailliot and his colleagues (2007) established that blood glucose levels underlie ego-depletion phenomena: An act of effortful self-regulation, such as controlling one's emotions or interacting with someone of a different race, reduced the amount of available glucose; low glucose after an initial task of self-regulation predicted poorer performance on later attempts at self-regulation; and restoring glucose (via sweetened lemonade) restored participants' self-regulation ability. This research, then,

links three levels: The individual level (an act of self-control), the social level (e.g., an interracial interaction), and the molecular level (blood glucose).

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 (Brinthaup and Lipka, 1994). We cover two change phenomena: self-expansion and self-development.

### **Self-expansion**

According to Aron's self-expansion theory, human beings have a central desire to expand the self—to acquire resources, perspectives, and identities that enhance their ability to accomplish goals, providing a motivation to enhance potential efficacy (Aron, McLaughlin-Volpe, Mashek, Lewandowski, Wright, & Aron, 2004, p. 105). This motivation to self-expand, at the individual level, influences the social level: Aron and his colleagues argue that self-expansion motives lead people to want to enter and maintain close relationships with others. In close relationships, each partner “includes the other in the self,” meaning that each takes on the other's resources, perspectives, and identities to some extent. Evidence for such processes is illustrated by findings of a study by Aron, Paris, and Aron (1995), who asked university students to respond to the open-ended question “Who are you today?” every two weeks for ten weeks. When respondents had fallen in love during the preceding two weeks, their answers to this question revealed increases in the diversity of their self-concept, compared to periods when they had not fallen in love and compared to other respondents who had not fallen in love. They also



showed increased self-efficacy and self-esteem. These results remained significant even after mood changes were controlled statistically.

Falling in love also seems to be accompanied by changes in the brain. fMRI studies show that when people who have recently fallen intensely in love look at a photo of or think about their beloved, they have increased activity in the caudate nucleus, which is a central part of the brain's "reward system," as well as in the right ventral tegmental area, a region associated with the production and distribution of dopamine to other brain regions (Aron, Fisher, Mashek, Strong, Li, & Brown, 2004). Even subliminal priming with a beloved's name has similar effects (Ortigue, Bianchi-Demicheli, Hamilton, & Grafton, 2007). These results suggest that passionate romantic love is associated with dopamine pathways in the reward system of the brain.

Recent research offers exciting evidence of possible brain changes with self-expansion. Ortigue and Bianchi-Demicheli (2010) found that when people were primed with their romantic partner's name (and not a friend's), they showed more intense activation of the left angular gyrus, the same region that is activated when people think of themselves.

### **Self-development**

Self-development refers to the changes that people naturally undergo over the course of their lives. Major developmental periods include early years when infants and toddlers begin to acquire identities (Bloom, 2004; Rochat, 2009), adolescence when teenagers establish increasing independence from parents (Sylwester, 2007), and old age when physical decline imposes new limitations on the self. Each of these periods

involves extensive social, individual, neural, and molecular changes, but we 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 individual 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-affecting 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, individual, 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.

## **OBJECTIONS**

The multilevel view of the self is open to objections from many directions. Some philosophers will think that we have slighted phenomenological aspects of the self - what it feels like to be you (e.g. Gallagher and Zahavi, 2008; Strawson, 2009; Zahavi 2005). Perhaps selfhood is more a matter of ongoing lived experience than the result of multiple mechanisms. Our 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 encourages hanging on to transcendental views of the self as soul.

A second objection is that our 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. Gallagher, 2006; Gibbs, 2005), but the multilevel approach is compatible with these insights (Thagard, forthcoming-b). 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 our account of the self with views that understand cognition as intimately coupled with bodies and the physical and social worlds with which they interact (Legrand and Ruby, 2009). We 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. Understanding the self as a dynamic system situated in physical and social worlds requires attention to internal representational models (Ismael, 2007)

A final objection is that the account of selves as multilevel systems is terminally obscure, bereft of explanatory power. We 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 individual 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. There is even a start on explaining such neuropathologies of the self as anosognosia, asomatognosia, delusional misidentification, depersonalization, and Capgras and Fregoli syndromes (Feinberg, 2009). Psychologists and philosophers have tended to ignore the molecular level, but increased focus on neural mechanisms is inevitably leading also to increased attention to molecular mechanisms.

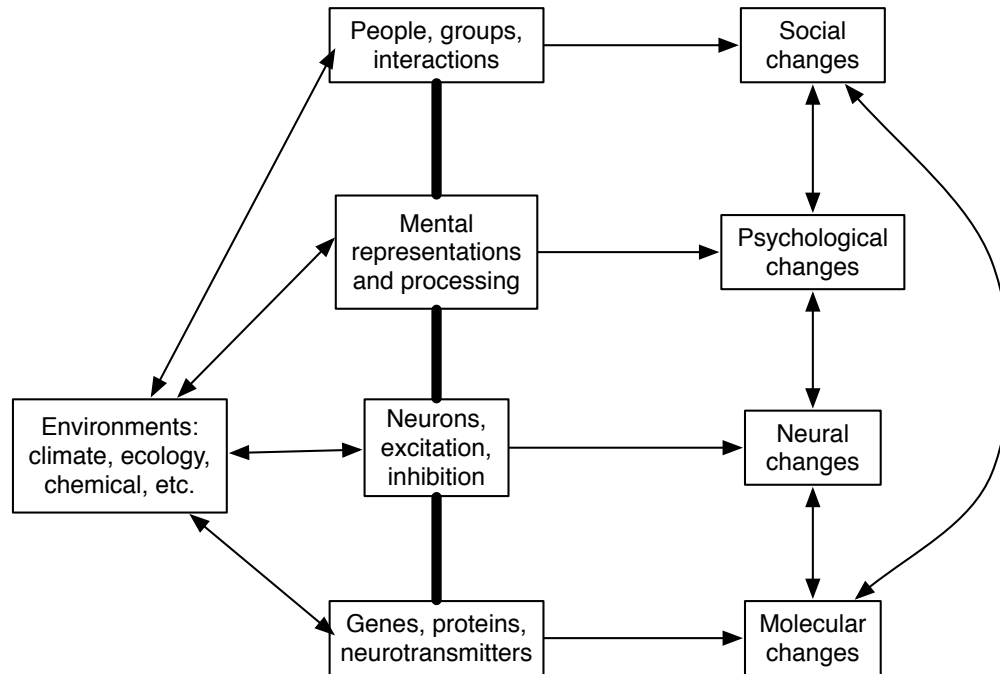
What is most obviously lacking in current discussions is much of an understanding of the relations among levels. We have maintained that there are interlevel feedback loops that account for much of the richness and unpredictability of human behavior. But much more research needs to be done to better comprehend the relations among the social, individual, 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? Our 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. Our multilevel account is not a theory of the self, but rather a framework for developing specific theories that describe mechanisms that operate within and between levels.

The justification for adopting the multilevel system view of the self is that it is superior to alternative accounts in potentially 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, ecologies, 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 we 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 we 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), economics (Thagard, forthcoming-a), and culture (Thagard, forthcoming-c).

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.

Our concern in this paper has been narrower: to make sense of the self by considering it as a multilevel system consisting of interacting social, individual, neural, and molecular mechanisms. We have shown the relevance of all of these levels to many important phenomena, including: self-concepts, self-consciousness, self-deception, self-presentation, self-disclosure, self-criticism, self-esteem, self-enhancement, self-protection, self-verification, self-regulation, self-expansion, and self-development. These thirteen are representative of three general classes (self-representing, self-efficacy, and self-changing) that cover more than seventy 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. Our 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. Our terminology is very similar: Composition=Parts, Environment=Environment, Structure=Interconnections, and Mechanism=Changes. The major difference is that we 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), our 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. We 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 we 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. Our 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 claim 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**

- Alicke, M. D., Dunning, D. A., & Krueger, J. I. (Eds.). (2005). *The self in social judgment*. New York: Psychology Press.
- Alicke, M. D., & Sedikides, C. (in press). Self-enhancement and self-protection: Historical overview and conceptual framework. In M. D. Alicke & C. Sedikides (Eds.), *The handbook of self-enhancement and self-protection*. New York, NY: Guilford Press.
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108(4), 441-485.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

- Anderson, J. R. (2007). *How can the mind occur in the physical universe?* Oxford: Oxford University Press.
- Aron, A. (2003). Self and close relationships. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity*. (pp. 442-461). New York, NY, US: Guilford Press.
- Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, Motivation, and Emotion Systems Associated With Early-Stage Intense Romantic Love. *Journal of Neurophysiology*, 94(1), 327-337.
- Aron, A., McLaughlin-Volpe, T., Mashek, D., Lewandowski, G., Wright, S. C., & Aron, E. N. (2004). Including others in the self. *European Review of Social Psychology*, 15, 101-132.
- Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of personality and social psychology*, 69(6), 1102-1112.
- Aronson, E. (2007). *The social animal*. New York: Worth Publishers.
- Baldwin, M. W., Carrell, S. E., & Lopez, D. F. (1990). Priming relationship schemas: My advisor and the Pope are watching me from the back of my mind. *Journal of experimental social psychology*, 26(5), 435-454.
- Banaji, M. R., & Prentice, D. A. (1994). The self in social contexts. *Annual Review of Psychology*, 45, 297-332.
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7, 84-91.
- Baumeister, R. (Ed.). (1999). *The self in social psychology*. Philadelphia: Psychology Press.
- Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology, Vols. 1 and 2 (4th ed.)*. (pp. 680-740). New York, NY, US: McGraw-Hill.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of personality and social psychology*, 74(5), 1252-1265.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, 4(1), 1-44.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497-529.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3), 497-529.
- Baumeister, R. F., Tice, D. M., & Hutton, D. G. (1989). Self-presentational motivations and personality differences in self-esteem. *Journal of personality*, 57(3), 547-579.
- Baumeister, R. F., Vohs, K. D., Leary, M. R., & Tangney, J. P. (2003). Self-regulation and the executive function of the self *Handbook of self and identity*. (pp. 197-217). New York, NY, US: Guilford Press.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16(6), 351-355.

- Beaman, A. L., Klentz, B., Diener, E., & Svanum, S. (1979). Self-awareness and transgression in children: Two field studies. *Journal of personality and social psychology*, 37(10), 1835-1846.
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity*. Princeton: Princeton University Press.
- Benjamin, J., Li, L., Patterson, C., Greenberg, B. D., Murphy, D. L., & Hamer, D. H. (1996). Population and familial association between the D4 dopamine receptor gene and measures of Novelty Seeking. *Nature Genetics*, 12(1), 81-84.
- Berscheid, E., & Regan, P. (2005). *The psychology of interpersonal relationships*. Mahwah, NJ: Prentice-Hall.
- Blaine, B., & Crocker, J. (1993). Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review. In R. F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 55-86). New York: Plenum.
- Blaine, B., Crocker, J., & Baumeister, R. F. (1993). Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review *Self-esteem: The puzzle of low self-regard*. (pp. 55-85). New York, NY, US: Plenum Press.
- Bloom, P. (2004). *Descartes's baby*. New York: Basic Books.
- Bonanno, G. A., Field, N. P., Kovacevic, A., & Kaltman, S. (2002). Self-enhancement as a buffer against extreme adversity: Civil war in Bosnia and traumatic loss in the United States. *Personality and Social Psychology Bulletin*, 28(2), 184-196.
- Bond, M. H., Kwan, V. S. Y., & Li, C. (2000). Decomposing a sense of superiority: The differential social impact of self-regard and regard for others. *Journal of Research in Personality*, 34(4), 537-553.
- Breytspraak, L. M. (1984). *The development of self in later life*. Boston: Little, Brown.
- Brinthaup, T. M., & Lipka, R. P. (Eds.). (1994). *Changing the self: Philosophies, techniques, and experiences*. Albany: State University of New York Press.
- Brown, J. D., & Smart, S. A. (1991). The self and social conduct: Linking self-representations to prosocial behavior. *Journal of personality and social psychology*, 60(3), 368-375.
- Bunge, M. (1998). *Social science under debate: A philosophical perspective*. Toronto: University of Toronto Press.
- Bunge, M. (2003). *Emergence and convergence: Qualitative novelty and the unity of knowledge*. Toronto: University of Toronto Press.
- Campbell, J. D. (1990). Self-esteem and clarity of the self-concept. *Journal of personality and social psychology*, 59(3), 538-549.
- Carver, C. S. (2003). Self-awareness. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity*. (pp. 179-196). New York, NY, US: Guilford Press.
- Carver, C. S., & Scheier, M. F. (1978). Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of personality and social psychology*, 36(3), 324-332.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York: Springer-Verlag.
- Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97(1), 19-35.

- Churchland, P. M. (2007). *Neurophilosophy at work*. Cambridge: Cambridge University Press.
- Churchland, P. S., & Sejnowski, T. (1992). *The computational brain*. Cambridge, MA: MIT Press.
- Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of personality and social psychology*, 68(6), 1152-1162.
- Cooley, C. H. (1902). *Human nature and the social order*.
- Craver, C. F. (2007). *Explaining the brain*. Oxford: Oxford University Press.
- Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22, 547-663.
- Creswell, J. D., Welch, W. T., Taylor, S. E., Sherman, D. K., Gruenewald, T. L., & Mann, T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress responses. *Psychological Science*, 16(11), 846-851.
- Crocker, J., Niiya, Y., & Mischkowski, D. (2008). Why does writing about important values reduce defensiveness? Self-affirmation and the role of positive other-directed feelings. *Psychological Science*, 19(7), 740-747.
- Curtis, R. C., & Miller, K. (1986). Believing another likes or dislikes you: Behaviors making the beliefs come true. *Journal of personality and social psychology*, 51(2), 284-290.
- D'Argembeau, A., Collette, F., Van der Linden, M., Laureys, S., Del Fiore, G., Degueldre, C., et al. (2005). Self-referential reflective activity and its relationship with rest: A PET study. *NeuroImage*, 25, 616-624.
- Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: Cambridge University Press.
- Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, 35(10), 867-881.
- Davis, J. A. (1966). The Campus as a Frog Pond: an Application of the Theory of Relative Deprivation to Career Decisions of College Men. *American Journal of Sociology*, 72(1), 17-31.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268.
- Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown.
- Dodgson, P. G., & Wood, J. V. (1998). Self-esteem and the cognitive accessibility of strengths and weaknesses after failure. *Journal of personality and social psychology*, 75(1), 178-197.
- Dunning, D. (2003). The relation of self to social perception. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 421-441). New York: Guilford Press.
- Eisenstadt, D., & Leippe, M. R. (1994). The self-comparison process and self-discrepant feedback: Consequences of learning you are what you thought you were not. *Journal of personality and social psychology*, 67(4), 611-626.

- Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.
- Eliasmith, C., & Stewart, T. C. (forthcoming). A new biologically implemented cognitive architecture. *unpublished manuscript, University of Waterloo Centre for Theoretical Neuroscience*.
- Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science*, 25, 245-286.
- Erber, R., & Erber, M. W. (2000). The self-regulation of moods: Second thoughts on the importance of happiness in everyday life. *Psychological Inquiry*, 11(3), 142-148.
- Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of personality and social psychology*, 41(2), 232-242.
- Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of personality and social psychology*, 73(1), 31-44.
- Feinberg, T. E. (2009). *From axons to identity: Neurological explorations of the nature of the self*. New York: Norton.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Finkel, E. J., & Campbell, W. K. (2001). Self-control and accommodation in close relationships: An interdependence analysis. *Journal of personality and social psychology*, 81(2), 263-277.
- Fitzsimons, G. M., & Bargh, J. A. (2003). Thinking of you: Nonconscious pursuit of interpersonal goals associated with relationship partners. *Annual Meeting of the Society for Personality and Social Psychology*, Feb 2002, Savannah, GA, US; Portions of this research were presented as a poster at the aforementioned meeting., 84(1), 148-164.
- Fitzsimons, G. M., & Fishbach, A. (2010). Shifting closeness: Interpersonal effects of personal goal progress. *Journal of personality and social psychology*, 98(4), 535-549.
- Fitzsimons, G. M., & Shah, J. Y. (2008). How goal instrumentality shapes relationship evaluations. *Journal of personality and social psychology*, 95(2), 319-337.
- Forest, A. L., & Wood, J. V. (2010). Self-protection versus self-optimization: Managing conflicting self-esteem-related motives in interpersonal relationships. *Unpublished Manuscript, Department of Psychology, University of Waterloo, Canada*.
- Gailliot, M. T., Mead, N. L., & Baumeister, R. F. (2008). Self-regulation. In O. P. John, R. W. Robbins & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 472-491). New York: Guilford Press.
- Gailliot, M. T., Mead, N. L., Baumeister, R. F., John, O. P., Robins, R. W., & Pervin, L. A. (2008). Self-regulation *Handbook of personality psychology: Theory and research* (3rd ed.). (pp. 472-491). New York, NY, US: Guilford Press.
- Gallagher, S. (2006). *How the body shapes the mind*. Oxford: Oxford University Press.
- Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. London: Routledge.



- Gaucher, D., Wood, J. V., Stinson, D. A., Holmes, J. G., Logel, C., & Forest, A. L. (2010). Testing a risk regulation model of emotional expressivity: How feeling secure in a partner's regard promotes expression. *Unpublished manuscript, University of Waterloo, Canada.*
- Gentner, D., & Markman, A. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, 5, 152-158.
- Gibbs, R. W. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
- Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York, NY, US: Free Press.
- Glimcher, P. W., Camerer, C., Poldrack, R. A., & Fehr, E. (Eds.). (2009). *Neuroeconomics: Decision making and the brain*. Amsterdam: Elsevier.
- Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.
- Goleman, D. (1985). *Vital lies, simple truths*. New York: Simon & Schuster.
- Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of personality and social psychology*, 37(2), 147-169.
- Heimpel, S. A., Wood, J. V., Marshall, M. A., & Brown, J. D. (2002). Do people with low self-esteem really want to feel better? Self-esteem differences in motivation to repair negative moods. *Journal of personality and social psychology*, 82(1), 128-147.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319-340.
- Holmes, J. G., Cameron, J., & Baldwin, M. W. (2005). An integrative review of theories of interpersonal cognition: An interdependence theory perspective *Interpersonal cognition*. (pp. 415-447). New York, NY, US: Guilford Press.
- Hume, D. (1888). *A treatise of human nature*. Oxford: Clarendon Press.
- Hummel, J. E., & Holyoak, K., J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220-264.
- Ingram, R. E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological bulletin*, 107(2), 156-176.
- Ismael, J. T. (2007). *The situated self*. Oxford: Oxford University Press.
- Johnson, M. L. (Ed.). (2005). *The Cambridge handbook of age and ageing*. Cambridge: Cambridge University Press.
- Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the self from the negative consequences of risky decisions. *Journal of personality and social psychology*, 62(1), 26-37.
- Jussim, L., Yen, H., & Aiello, J. R. (1995). Self-consistency, self-enhancement, and accuracy in reactions to feedback. *Journal of experimental social psychology*, 31(4), 322-356.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Kwang, T., & Swann, W. B., Jr. (2009). Will the real self-enhancement motive please stand up? Other motives may underlie the quest for positivity. *Unpublished manuscript, University of Texas at Austin, TX.*

- Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, Vol. 32. (pp. 1-62). San Diego, CA, US: Academic Press.
- Leary, M. R., Haupt, A. L., Strausser, K. S., & Chokel, J. T. (1998). Calibrating the sociometer: The relationship between interpersonal appraisals and the state self-esteem. *Journal of personality and social psychology*, 74(5), 1290-1299.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34-47.
- Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, 116, 252-282.
- Lieberman, M. D. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annual Review of Psychology*, 58, 259-289.
- Lieberman, M. D. (2010). Social cognitive neuroscience. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., pp. 143-193). New York: McGraw-Hill.
- Macdonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131, 202-223.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1-25.
- Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of personality and social psychology*, 51(4), 858-866.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224-253.
- McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and the heuristic identity theory. *Theory & Psychology*, 11, 736-760.
- McCullough, K. C., Fitzsimons, G. M., & Chua, S. N. (2010). Vicarious goal satisfaction. *Unpublished manuscript, University of Idaho, Moscow, ID*.
- McGuire, W. J., & McGuire, C. V. (1996). Enhancing self-esteem by directed-thinking tasks: Cognitive and affective positivity asymmetries. *Journal of personality and social psychology*, 70(6), 1117-1125.
- McGuire, W. J., McGuire, C. V., Child, P., & Fujioka, T. (1978). Salience of ethnicity in the spontaneous self-concept as a function of one's ethnic distinctiveness in the social environment. *Journal of personality and social psychology*, 36(5), 511-520.
- Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.
- Metzinger, T. (2009). *The ego tunnel: The science of the mind and the myth of the self*. New York: Basic Books.
- Moretti, M. M., & Higgins, E. T. (1999). Own versus other standpoints in self-regulation: Developmental antecedents and functional consequences. *Review of General Psychology*, 3(3), 188-223.
- Muraven, M., Collins, R. L., & Neinhuis, K. (2002). Self-control and alcohol restraint: An initial application of the Self-Control Strength Model. *Psychology of Addictive Behaviors*, 16(2), 113-120.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Murray, S. L., & Holmes, J. G. (2009). ADD.

- Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The self-fulfilling nature of positive illusions in romantic relationships: Love is not blind, but prescient. *Journal of personality and social psychology*, 71(6), 1155-1180.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.
- Organ, T. W. (1987). *Philosophy and the self: East and west*. Selinsgrove, PA: Susquehanna University Press.
- Ortigue, S., & Bianchi-Demicheli, F. (2010). *Overlapping neural systems mediating self-consciousness, mirror neurons, and love: A combined fMRI and 4D EEG neuroimaging approach*. Paper presented at the Society for Personality and Social Psychology conference, .
- Ortigue, S., Bianchi-Demicheli, F., de C. Hamilton, A. F., & Grafton, S. T. (2007). The neural basis of love as a subliminal prime: An event-related functional magnetic resonance imaging study. *Journal of cognitive neuroscience*, 19(7), 1218-1230.
- Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of personality and social psychology*, 74(5), 1197-1208.
- Paulhus, D. L. (2007). Self-deception. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 189-190). Thousand Oaks, CA: : Sage.
- Pettigrew, T. F. (1967). Social evaluation theory: Convergences and applications. *Nebraska Symposium on Motivation*, 15, 241-311.
- Prior, H., Schwarz, A., & Güntürkün, O. (2008). Mirror-induced representation in magpies: Evidence of self-recognition. *PLoS Biology*, 6(8), e202.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102-1107.
- Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. In S. Duck (Ed.), *Handbook of personal relationships* (pp. 367-389). Oxford, England: John Wiley & Sons.
- Richards, E. J. (2006). Inherited epigenetic variation--revisiting soft inheritance. *Nature Reviews Genetics*, 7(5), 395-401.
- Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, 6(12), 1323-1328.
- Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, 14(3), 287-290.
- Rochat, P. (2009). *Others in mind: Social origins of self-consciousness*. Cambridge: Cambridge University Press.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York, NY, US: Holt, Rinehart & Winston.
- Ross, M., Xun, W. Q. E., & Wilson, A. E. (2002). Language and the bicultural self. *Personality and Social Psychology Bulletin*, 28(8), 1040-1050.

- Roy, M. A., Neale, M. C., & Kindler, K. S. (1995). The genetic epidemiology of self-esteem. *The British Journal of Psychiatry*, 166, 813-820.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- Scheier, M. F. (1976). Self-awareness, self-consciousness, and angry aggression. *Journal of personality*, 44(4), 627-644.
- Scheier, M. F., & Carver, C. S. (1977). Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of personality and social psychology*, 35(9), 625-636.
- Scheier, M. F., & Carver, C. S. (1983). Self-directed attention and the comparison of self with standards. *Journal of experimental social psychology*, 19(3), 205-222.
- Schlenker, B. R. (2003). Self-presentation. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity*. (pp. 492-518). New York, NY, US: Guilford Press.
- Schlenker, B. R., Weigold, M. F., & Hallam, J. R. (1990). Self-serving attributions in social context: Effects of self-esteem and social pressure. *Journal of personality and social psychology*, 58(5), 855-863.
- Shaver, P. R., Collins, N., & Clark, C. L. (1996). Attachment styles and internal working models of self and relationship partners. In G. J. O. Fletcher & J. Fitness (Eds.), *Knowledge structures in close Relationships: A social psychological approach* (pp. 25-61). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist*, 48(11), 1117-1131.
- Shrauger, J. S., & Rosenberg, S. E. (1970). Self-esteem and the effects of success and failure feedback on performance. *Journal of personality*, 38(3), 404-417.
- Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467-482.
- Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, 26(11), 1329-1342.
- Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive psychology: Mind and brain*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of personality and social psychology*, 35(9), 656-666.
- Spencer, S. J., Fein, S., & Lomore, C. D. (2001). Maintaining one's self-image vis-à-vis others: The role of self-affirmation in the social evaluation of the self. *Motivation and Emotion*, 25(1), 41-65.
- Spencer, S. J., Josephs, R. A., Steele, C. M., & Baumeister, R. F. (1993). Low self-esteem: The uphill struggle for self-integrity. *Self-esteem: The puzzle of low self-regard*. (pp. 21-36). New York, NY, US: Plenum Press.
- Stapel, D. A., & Tesser, A. (2001). Self-activation increases social comparison. *Journal of personality and social psychology*, 81(4), 742-750.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). New York: Academic Press.

- Stinson, D. A., Cameron, J. J., Wood, J. V., Gaucher, D., & Holmes, J. G. (2009). Deconstructing the reign of error: Interpersonal warmth explains the self-fulfilling prophecy of anticipated acceptance. *Personality and Social Psychology Bulletin*, 35(9), 1165-1178.
- Stinson, D. A., Logel, C., Holmes, J. G., Wood, J. V., Hogle, J. V., Gaucher, D., et al. (in press). The regulatory function of self-esteem: Testing to the acceptance and epistemic signaling systems. *Journal of Personality and Social Psychology*.
- Stinson, D. A., Logel, C., Zanna, M. P., Holmes, J. G., Cameron, J. J., Wood, J. V., et al. (2008). The cost of lower self-esteem: Testing a self- and social-bonds model of health. *Journal of personality and social psychology*, 94(3), 412-428.
- Stinson, D. A., Wood, J. V., & Doxey, J. R. (2008). In search of clarity: Self-esteem and domains of confidence and confusion. *Personality and Social Psychology Bulletin*, 34(11), 1541-1555.
- Strawson, G. (2009). *Selves: An essay in revisionary metaphysics*. Oxford: Clarendon Press.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2), 143-148.
- Swann, W. B., Chang-Schneider, C., Angulo, S., Wood, J. V., Tesser, A., & Holmes, J. G. (2008). Self-verification in relationships as an adaptive process *The self and social relationships*. (pp. 49-72). New York, NY, US: Psychology Press.
- Swann, W. B., Jr. (in press). Self-verification theory. In A. K. P. Van Lang, & E. T. Higgins (Ed.), *Handbook of theories of social psychology*. New York: Sage.
- Swann, W. B., Jr., & Pelham, B. (2002). Who wants out when the going gets good? Psychological investment and preference for self-verifying college roommates. *Self and Identity*, 1(3), 219-233.
- Swann, W. B., & Schroeder, D. G. (1995). The search for beauty and truth: A framework for understanding reactions to evaluations. *Personality and Social Psychology Bulletin*, 21(12), 1307-1318.
- Swann, W. B., Stein-Seroussi, A., & Giesler, R. B. (1992). Why people self-verify. *Journal of personality and social psychology*, 62(3), 392-401.
- Swann, W. B., Wenzlaff, R. M., Krull, D. S., & Pelham, B. W. (1992). Allure of negative feedback: Self-verification strivings among depressed persons. *Journal of abnormal psychology*, 101(2), 293-306.
- Sylwester, P. (2007). *The adolescent brain: Reaching for autonomy*. Thousand Oaks, CA: Corwin Press.
- Tajfel, H. (1974). Social identity and intergroup behavior. *Social science information*, 13, 65-93.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological bulletin*, 103(2), 193-210.
- Taylor, S. E., Burklund, L. J., Eisenberger, N. I., Lehman, B. J., Hilmert, C. J., & Lieberman, M. D. (2008). Neural bases of moderation of cortisol stress responses by psychosocial resources. *Journal of personality and social psychology*, 95(1), 197-211.
- Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003a). Are self-enhancing cognitions associated with healthy or unhealthy biological profiles? *Journal of personality and social psychology*, 85(4), 605-615.

- Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003b). Portrait of the self-enhancer: Well adjusted and well liked or maladjusted and friendless? *Journal of personality and social psychology*, 84(1), 165-176.
- Taylor, S. E., Neter, E., & Wayment, H. A. (1995). Self-evaluation processes. *Personality and Social Psychology Bulletin*, 21(12), 1278-1287.
- Tesser, A., Leary, M. R., & Tangney, J. P. (2003). Self-evaluation (pp. 275-290). New York, NY, US: Guilford Press.
- Thagard, P. (1999). *How scientists explain disease*. Princeton, N.J.: Princeton University Press.
- Thagard, P. (2005). *Mind: Introduction to cognitive science* (2nd ed.). Cambridge, Mass.: MIT Press.
- Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.
- Thagard, P. (2007). The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy*, 24, 367-380.
- Thagard, P. (2008). How cognition meets emotion: Beliefs, desires, and feelings as neural activity. In G. Brun, U. Doguoglu & D. Kuenzle (Eds.), *Epistemology and emotions* (pp. 167-184). Aldershot: Ashgate.
- Thagard, P. (2010). *The brain and the meaning of life*. Princeton, NJ: Princeton University Press.
- Thagard, P. (forthcoming-a). Explaining economic crises: Are there collective representations? *Episteme*.
- Thagard, P. (forthcoming-b). How brains make mental models. In L. Magnani, W. Carnielli & C. Pizzi (Eds.), *Model-based reasoning in science and technology. Abduction, logic, and computational discovery*. Berlin: Springer.
- Thagard, P. (forthcoming-c). Mapping minds across cultures. In R. Sun (Ed.), *Grounding social sciences in cognitive sciences*. Cambridge, MA: MIT Press.
- Thagard, P., & Aubie, B. (2008). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition*, 17, 811-834.
- Thagard, P., & Stewart, T. C. (forthcoming). The Aha! experience: Creativity through emergent binding in neural networks. *Cognitive Science*.
- Tice, D. M. (1991). Esteem protection or enhancement? Self-handicapping motives and attributions differ by trait self-esteem. *Journal of personality and social psychology*, 60(5), 711-725.
- Tice, D. M., Butler, J. L., Muraven, M. B., & Stillwell, A. M. (1995). When modesty prevails: Differential favorability of self-presentation to friends and strangers. *Journal of personality and social psychology*, 69(6), 1120-1138.
- Tice, D. M., & Wallace, H. M. (2003). The reflected self: Creating yourself as (you think) others see you. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity*. (pp. 91-105). New York, NY, US: Guilford Press.
- Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, 1124, 239-261.
- Twenge, J. M., & Campbell, W. K. (2003). "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin*, 29(2), 261-272.

- Vandekerckhove, M., von Scheve, C., Ismer, S., Jung, S., & Kronast, S. (Eds.). (2008). *Regulating emotions: Culture, social necessity, and biological inheritance*. Malden, MA: Blackwell.
- Vohs, K. D., Baumeister, R. F., & Ciarocco, N. J. (2005). Self-Regulation and Self-Presentation: Regulatory Resource Depletion Impairs Impression Management and Effortful Self-Presentation Depletes Regulatory Resources. *Journal of personality and social psychology*, 88(4), 632-657.
- Walton, Cohen, Cwir, & Spencer (2010). ADD.
- Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of personality and social psychology*, 83(1), 185-197.
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings*. Cambridge, MA: Harvard University Press.
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological bulletin*, 106(2), 231-248.
- Wood, J. V., Giordano-Beech, M., & Ducharme, M. J. (1999). Compensating for failure through social comparison. *Personality and Social Psychology Bulletin*, 25(11), 1370-1386.
- Wood, J. V., Giordano-Beech, M., Taylor, K. L., Michela, J. L., & Gaus, V. (1994). Strategies of social comparison among people with low self-esteem: Self-protection and self-enhancement. *Journal of personality and social psychology*, 67(4), 713-731.
- Wood, J. V., Heimpel, S. A., Manwell, L. A., & Whittington, E. J. (2009). This mood is familiar and I don't deserve to feel better anyway: Mechanisms underlying self-esteem differences in motivation to repair sad moods. *Journal of personality and social psychology*, 96(2), 363-380.
- Wood, J. V., Perunovic, W. Q. E., & Lee, J. W. (2009). Positive self-statements: Power for some, peril for others. *Psychological Science*, 20, 860-866.
- Wood, J. V., Saltzberg, J. A., & Goldsamt, L. A. (1990). Does affect induce self-focused attention? *Journal of personality and social psychology*, 58(5), 899-908.
- Wood, J. V., Taylor, S. E., & Lichtman, R. R. (1985). Social comparison in adjustment to breast cancer. *Journal of personality and social psychology*, 49(5), 1169-1183.
- Wood, J. V., & Wilson, A. E. (2003). How important is social comparison? In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity*. (pp. 344-366). New York, NY, US: Guilford Press.
- Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge, MA: MIT Press.