

# The Future of Psychology

## Connecting Mind to Brain

Lisa Feldman Barrett

*Boston College, Massachusetts General Hospital, and Harvard Medical School*

---

**ABSTRACT**—*Psychological states such as thoughts and feelings are real. Brain states are real. The problem is that the two are not real in the same way, creating the mind–brain correspondence problem. In this article, I present a possible solution to this problem that involves two suggestions. First, complex psychological states such as emotion and cognition can be thought of as constructed events that can be causally reduced to a set of more basic, psychologically primitive ingredients that are more clearly respected by the brain. Second, complex psychological categories like emotion and cognition are the phenomena that require explanation in psychology, and, therefore, they cannot be abandoned by science. Describing the content and structure of these categories is a necessary and valuable scientific activity.*

---

*Physical concepts are free creations of the human mind, and are not, however it may seem, uniquely determined by the external world.*

—Einstein & Infeld (1938, p. 33)

*The cardinal passions of our life, anger, love, fear, hate, hope, and the most comprehensive divisions of our intellectual activity, to remember, expect, think, know, dream, with the broadest genera of aesthetic feeling, joy, sorrow, pleasure, pain, are the only facts of a subjective order which this vocabulary deigns to note by special words.*

—James (1890, p. 195)

From its inception in the early 18th century (as an amalgam of philosophy, neurology, and physiology), psychology has always been in a bit of an identity crisis, trying to be both a social and a natural science.<sup>1</sup> Psychologists attempt to bridge the social and natural worlds using the conceptual tools of their time. Throughout our history, the link between the two has felt less like a solid

---

Address correspondence to Lisa Feldman Barrett, Department of Psychology, Boston College, Chestnut Hill, MA 02467; e-mail: barretli@bc.edu.

<sup>1</sup>Of course, psychology is not now, nor has it ever been, a unified discipline with a single approach to science. It encompasses many different topics, methods, and assumptions, as well as much diversity in its level of interest toward and treatment of the brain.

footbridge and more like a tightrope requiring lightness of foot and a really strong safety net. Mind–brain, and relatedly, behavior–brain, correspondence continue to be central issues in psychology, and they remain the largest challenge in 21st-century psychology.

The difficulty in linking the human mind to behavior on the one hand and to the brain on the other is rooted, ironically enough, in the way the human brain itself works. Human brains categorize continuously, effortlessly, and relentlessly. Categorization plays a fundamental role in every human activity, including science. Categorizing functions like a chisel, dividing up the sensory world into figure and ground, leading us to attend to certain features and to ignore others. Via the process of categorization, the brain transforms only some sensory stimulation into information. Only some of the wavelengths of light striking our retinas are transformed into seen objects, and only some of the changes in air pressure registered in our ears are heard as words or music. To categorize something is to render it meaningful. It then becomes possible to make reasonable inferences about that thing, to predict what to do with it, and to communicate our experience of it to others. There are ongoing debates about how categorization works, but the fact that it works is not in question.

The brain's compulsion to categorize presents certain unavoidable challenges to what can be learned about the natural world from human observation. Psychologists know that people don't contribute to their perceptions of the world in a neutral way. Human brains do not dispassionately look upon the world and carve nature at its joints. We make self-interested observations about the world in all manner of speaking. And what holds true for people in general certainly holds for scientists in particular. Scientists are active perceivers, and like all perceivers, we see the world from a particular point of view (that is not always shared by other scientists). We parse the world into bits and pieces using the conceptual tools that are available at a particular point in time and with a particular goal in mind (often inextricably linked to said conceptual tools). This is not a failing of the scientific method per se—it is a natural consequence of how the human brain sees and hears and feels . . . and does science.

An example of how categorization shapes science comes from the study of genetics. When molecular biologists first began to study the units of inheritance, they (inspired by Mendel) searched

for and found *genes*: bits of DNA that make the proteins needed to constitute the human body. Yet, only a small proportion of human DNA (somewhere between 2% to 5%, depending on which paper you read) are genes; the rest of the stuff (that does not directly produce proteins) was labeled “junk” on the assumption that it was largely irrelevant to the biological understanding of life. As it turns out, however, “junk DNA” has some rather important functions, including regulating gene expression (i.e., turning on and off protein production) in a contextually sensitive fashion (for a generally accessible review, see Gibbs, 2003). Scientists have discovered that much of what makes us human and makes one person different from another lurks in this junk. The result has been nothing short of a revolution in molecular genetics. Genes do not, in and of themselves, provide a sufficient recipe for life. The unit of selection is not the gene, but the individual, who, for the purposes of molecular genetics, can be thought of as a bundle of genes that are turned on and off by the rest of our DNA that regulates the epigenetic context. And, the more they learned about junk DNA, the more scientists realized that it is not so easy to define what is a gene and what is not. Some molecular geneticists now try to avoid the word “gene” altogether. Instead they use the more mechanistic term *transcriptional unit*.

In this article, I argue that perhaps psychology needs to reconsider its vocabulary of categories. Like any young science, psychology has been practicing a very sophisticated form of phenomenology, observing the psychological world using categories derived from our own experiences. We then use common sense words to name these categories, leading us to reify them as entities. We then search for the counterparts of these categories within the brain. These two practices—carving and naming—have a far-reaching consequence: Psychology may more or less accept the Kantian idea that the knowledge stored in a human brain contributes to thoughts, feelings, memories, and perceptions in a top-down fashion, but at the same time we accept without question that *emotions, thoughts, memories, the self*, and the other psychological categories in folk psychology reflect the basic building blocks of the mind. We do this in much the same way that Aristotle assumed that fire, earth, air, and water were the basic elements of the material universe; as if the categories themselves are not constructed out of something else more basic. In our causal explanations, psychologists talk about psychological facts as if they are physical facts.

But what if psychological facts are not physical facts? What if the phenomena we want to explain—emotions, cognitions, the self, behaviors—are not just the subject matter of the human mind, but are also the creations of that mind? What if the boundaries for these categories are not respected in the very brain that creates them?

Such a state of things might lead some scientists to conclude that psychological categories are not real, or that psychology as a science can be dispensed with. That all scientists need to do is understand the brain. But nothing can be farther from the truth. The main point I make in this article is that, in psychology, we

simultaneously take our phenomenology too seriously and not seriously enough: too seriously when trying to understand how the mind corresponds to the brain and not seriously enough when we want to understand psychological phenomena as real and scientifically valuable, even in the face of spectacular and unrelenting progress in neuroscience. Changing this state of affairs is a central task for the future.

## THE METAPHYSICS OF WHAT EXISTS

### Scientific Ontology

The vocabulary of categories in any field of science segregates the phenomena to be explained and in so doing makes them real. Yet not all categories are real in the same way.

Natural sciences like physics deal with scientific categories that are assumed to be *observer independent* (they are real in the natural sense and can be discovered by humans). These kinds of categories function like an archaeologist’s chisel—we need a mind to discover and experience instances of these categories, but they do not depend on our minds for their existence. In 2007, I discovered that philosopher John Searle (in writing about social institutions and social power; Searle, 1995) called these *ontologically objective categories* or “brute facts.” They are also called natural kind categories. At times, human experience can lull scientists into using the wrong categories at first (e.g., genes and junk DNA), but the phenomena themselves (e.g., DNA, RNA, proteins, and so on) exist outside of human experience and can have a corrective influence on which categories are used. Here, the presumption is that the scientific method takes us on what can seem like a long, slow, carnival ride toward discovering and explaining the material world.

Social sciences like sociology or economics deal with categories that are *observer dependent* (and are real because they are invented and shared by humans). Observer-dependent categories function like a sculptor’s chisel—they constitute what is real and what is not. Searle calls these ontologically subjective categories because they exist only by virtue of collective intentionality (which is a fancy way of saying that they are real by virtue of the fact that everybody largely agrees on their content). Humans create observer-dependent categories to serve some function. Their validity, and their very existence, in fact, comes from consensual agreement. Little pieces of paper have value to procure goods not because of their molecular structure, but because we all agree that they do, and these little pieces of paper would cease to have value if many people changed their mind (and refused to accept them in lieu of material goods). Only certain types of pair bonds between humans are defined as “marriages” and confer real social and monetary benefits. People are citizens of the same country (e.g., Canada, Yugoslavia, or the Soviet Union) only as long as they all agree that the country exists and this membership becomes part of a person’s identity and often confers social and economic

advantage.<sup>2</sup> We might also think of them as nominal kind categories,<sup>3</sup> as artifact categories, or as cognitive tools for getting along and getting ahead.

Psychology, in walking a tightrope between the social world and the natural world, tries to map observer-dependent categories to observer-independent categories. The trick, of course, is to be clear about which is which and to never mistake one for the other. Once psychology more successfully distinguishes between the two, I predict that we will be left with the more tractable but never simple task of understanding how to map mind and behavior to the human brain.

Beginning in 1992, I began to craft the position that emotion categories labeled as *anger* and *sadness* and *fear* are mistakenly assumed to be observer independent, when in fact they are dependent on human (particularly Western) perceivers for their existence. I published the first sketch of these ideas about a decade later (Barrett, 2005, 2006b), spurred on by my discovery of *the emotion paradox*: In the blink of an eye, perceivers experience anger or sadness or fear and see these emotions in other people (and in animals, or even simple moving shapes) as effortlessly as they read words on a page,<sup>4</sup> yet perceiver-independent measurements of faces, voices, bodies, and brains do not clearly and consistently reveal evidence of these categories (Barrett, 2006a; Barrett & Wager, 2006; also see Barrett, Lindquist, et al., 2007). Some studies of cardiovascular measurements, electromyographic activity of facial muscles, acoustical analyses of vocal cues, and blood-flow changes within the brain are consistent with the traditional idea that emotions are observer-independent categories, but the larger body of evidence disconfirms their status as ontologically objective entities. There is not a complete absence of statistical regularities across these measures during the events that we name as the same emotion (e.g., *anger*), but the variance observed within any single category is not all that different from the variability across different categories.

One solution to the emotion paradox suggests that *anger*, *sadness*, *fear*, and so on are observer-dependent psychological categories and that instances of these emotions live in the head of the perceiver (Barrett, 2006b). This is not to say that emotions exist only in the head of the perceiver. Rather, it is more correct to say that they cannot exist without a perceiver. I experience myself as angry or I see your face as angry or I experience the rat's behavior as angry, but anger does not exist independent of someone's perception of it. Without a perceiver, there are only internal sensations and a stream of physical actions.

<sup>2</sup>If enough people withdraw this agreement, the country splinters into two or more new countries. If people cannot agree to disagree, as is the case of those living in Quebec, which is still a province in Canada (the country where I am from), then the country remains intact.

<sup>3</sup>A nominal kind is a category, denoted by a word, that is a combination of more fundamental properties (Frawley, 1992).

<sup>4</sup>People don't always agree with each other in these perceptions, but that is a different story.

In 2007, based largely on neuroanatomical grounds, I extended this line of reasoning, arguing that the categories labeled as *emotion* and *cognition* are not observer-independent categories (Duncan & Barrett, 2007; for a similar view, see Pessoa, 2008). *Thinking* (e.g., sensing and categorizing an object, or deliberating on an object) is experienced as a fundamentally different sort of mental activity than *feeling* (i.e., representing how the object influences one's internal state). As a result, psychologists have believed for some time that cognitions and emotions are separate and distinctive processes in the mind that interact like the bit and parts of a machine. But the brain does not really respect these categories, and thus mental states cannot be said to be categorically one or the other. Nor can behavior be caused by their interaction.

In this article, I am extending this reasoning even further by proposing that many—perhaps even the majority—of the categories with modern psychological currency are like money, marriage, nationality, or any of the ontologically subjective categories that Searle writes about. The complex psychological categories we refer to as *thoughts*, *memories*, *emotions*, and *beliefs*, or *automatic processing*, *controlled processing*, or *the self*, and so on, are observer dependent. They are collections of mental states that are products of the brain, but they do not correspond to brain organization in a one-to-one fashion. These categories exist because a group of people agreed (for phenomenological and social reasons) that this is a functional way to parse the ongoing mental activity that is realized in the brain. Some of the categories are cross-culturally stable (because they function to address certain universal human concerns that stem from living in large, complex groups), whereas others are culturally relative. The distinction between categories like *emotion* and *cognition*, for example, is relative and can vary with cultural context (e.g., Wikan, 1990), thus calling into question the fact that they are universal, observer-independent categories of the mind.

Even the most basic categories in psychology appear to be observer dependent. Take, for example, *behaviors* (which are intentional, bounded events) and *actions* (which are descriptions of physical movements). People easily and effortlessly see behaviors in other people and in nonhuman animals. We typically believe that behaviors exist independent of an observer, and that they are detected, but not created, by the human brain. But this is not quite true. Behaviors are actions with a meaning that is inferred by an observer. Social psychology has accumulated a large and nuanced body of research on how people come to see the physical actions of others as meaningful behaviors by inferring the causes for those actions (usually by imputing an intention to the actor; for a review, see Gilbert, 1998; Vallacher & Wegner, 1987). People and animals are constantly moving and doing things—that is, they are constantly engaging in a flow of actions. A perceiver automatically and effortlessly partitions continuous movements into recognizable, meaningful, discrete behavioral acts using category knowledge about people and animals (Vallacher & Wegner, 1987). In emotion research, a rat

that kicks up bedding at a threat is said to be *defensive treading* or in a state of *fear*. Similarly, a rat in a small spare box who becomes still (except for respiration) in response to a tone that sometimes predicts an electric shock is said to be *freezing* or in a state of *fear*. Two very different actions are referred to as the same behavior if the same intention is inferred. But a freezing rat might also be referred to as *vigilant*—in an alert, behavioral stance that allows a martialing of attentional resources to quickly learn more about a stimulus when its predictive value is uncertain (cf. Barrett, Lindquist, et al., 2007). The category used depends on the inference made by the observer.

The same point can be made about situations. Physical surroundings exist separately from observers, but situations do not (for a discussion, see Barrett, 2006c).

A similar point can even be made about what are typically assumed to be the observer-independent phenomena measured during functional magnetic resonance functional imaging. Areas of the brain that show increased activity during *cognition*, *perception*, or *emotion* (or whatever the researcher is interested in measuring) are assumed to reflect changes in blood flow caused by neuronal firing at those locations. But just as behavioral scientists separate the variance in a measured behavior into *effect* (i.e., the measured variance of interest) and *error* (i.e., the measured variance that is not of interest), so do cognitive neuroscientists routinely separate blood oxygen level dependent changes into *signal* (the strong changes that they believe to be task dependent) and *noise* (the weaker changes that they don't care about). This separation is guided by the neuropsychological assumption that psychological functions are localized to modules in particular brain areas, like islands on a topographical map, because lesions in particular areas appear to disrupt specific psychological functions. In recent years, however, it has become clear (using multivariate voxel pattern analysis procedures) that the so-called noise carries meaningful psychological information (e.g., Haynes & Rees, 2006; Kay, Naselaris, Prenger, & Gallant, 2008; Norman, Polyn, Detre, & Haxby, 2006), just as junk DNA is not junk at all. This turn of events makes brain mapping less like cartography (mapping stationary masses of land) and more like meteorology (mapping changing weather patterns or “brainstorms”).

Let me be clear about what I am saying here—it is a brute fact that the brain contains neurons that fire to create mental states or cause actions and this occurs independent of human experience and measurement. It is not a brute fact, however, that this neuronal activity can be easily classified as *automatic processing* or *controlled processing*; that some “islands” in the brain realize *cognitions* whereas others realize *emotion*; or even that *the self*, or *goals*, or *memories* live in specific parts of the brain (whether in a local or distributed specific, unchanging network). We use categories to separate ongoing mental activity into discrete mental states (such as, in this culture, *anger*, an *attitude*, a *memory*, or *self-esteem*), to classify a stream of physical movements into behaviors (such as *lying*, *stealing*, or

*joking*), or to classify parts of the physical surroundings as situations. These categories come from and constitute human experience. The category instances are real, but they derive their reality from the human mind (in the context of other human minds). Mental activity is classified this way for reasons having to do with collective intentionality, communication, and even self-regulation, but not because this is the best way to understand how the brain mechanistically creates the mind and behavior. *Emotion* and *cognition* make up the Western psychological and social reality, and they must be explained by the brute fact of how the human brain works, but emotion and cognition are not mechanisms that are necessarily respected by the human brain or categories that are required by the human brain. Brain states are observer-independent facts. The existence of mental states is also an observer-independent fact. *Cognitions*, *emotions*, *memories*, *self-esteem*, *beliefs*, and so on are not observer-independent events, however. They are categories that have been formed and named by the human mind to represent and explain the human mind.

### What's in a Name?

Words are powerful in science. When dealing with observer-independent categories, words set the ground rules for what to look for in the world. To the extent that scientists understand and use the word in a similar way, they agree on what to search for. They assume, for the moment, that genetic material really is segregated into genes and junk, and they then go about searching for the deep properties that ground these categories in the material world, with the hope either that they are right or that their observations will lead them to formulate better, more accurate categories. When dealing with observer-dependent categories that populate psychology, however, words are ontologically powerful. They set the ground rules for what exists.

Words can also be dangerous. They present scientists with a Faustian bargain. We need words to do the work of science, but words can lead us to mistake observer-dependent categories (or nominal kinds) for observer-independent categories (or natural kinds). By naming both defensive treading and freezing as *fear*, for example, scientists are lulled into thinking these behaviors share a deep property, and they will spend years searching for it, even when it may not exist. This is because a word doesn't only name a category, it also encourages a very basic form of essentialism that Paul Bloom (2004) argues is already present in how people think about the events and objects in their everyday lives. A word functions like an “essence placeholder” that encourages people to engage in psychological essentialism—it convinces the perceiver that there is some deep reality to the category in the material world (Medin & Ortony, 1989). This is true even in young children (e.g., Xu, Cote, & Baker, 2005). William James (1890) described the danger of referring to psychological categories with words when he wrote, “Whenever we have made a word . . . to denote a certain group of phenomena, we are prone to suppose a substantive entity existing beyond the phenomena, of which the word shall be the name” (p. 195). In

psychology’s active and ongoing attempt to knit the social and natural worlds together into one seamless universe, words cause us to take phenomenology inspired categories—Western categories no less—and search for years (often in vain) for the specific brain areas, genes, hormones, or some other biological product that they correspond to. Then we end up arguing about whether the amygdala is the brain locus of *fear*, whether dopamine is the hormone for *reward*, or whether the serotonin transporter gene (5-HTTLPR) is the cause of *depression*.

### FORWARD INTO THE PAST

Thus far, I have suggested that psychology is populated by a set of observer-dependent categories that do not directly correspond (in a one-to-one fashion) to the observer-independent facts of neurons firing in the brain. If this claim is true, then psychology’s current vocabulary of phenomenologically grounded categories won’t interface very well with neuroscience to effectively weave social and natural sciences together into one cozy blanket and solve the problem of mind–brain (or behavior–brain) correspondence. Psychology may need a different set of psychological categories—categories that more closely describe the brain’s activities in creating the mind and causing action. That being said, if emotion, cognition, memory, the self, and so on, exist—they are real by virtue of the fact that everyone within a culture experiences them, talks about them, uses them as reasons for actions—then they cannot be discarded or ontologically reduced to (or merely redefined as nothing but) neurons firing. Psychology must explain the existence of cognition and emotion because they are part of the world that we (in the Western hemisphere) live in (even if they are a part that we, ourselves, created).<sup>5</sup> Can psychology both describe what emotions and cognitions (or whatever the mental categories within a given cultural context) are and also explain how they are caused? I think the answer is yes. And as with most things psychological, the answer begins with William James.

Over a century ago, William James wrote about the psychologist’s fallacy. “The great snare of the psychologist,” James wrote, “is the confusion of his own standpoint with that of the mental fact about which he is making his report” (James, 1890, p. 196). This is pretty much the same thing as saying that psychologists confuse observer-dependent (or ontologically subjective) distinctions with observer-independent (or ontologically objective) ones. The solution to the psychologist’s fallacy, according to James, is to take a psychological constructionist approach. “A science of the relations of mind and brain,” James wrote, “must show how the elementary ingredients of the former correspond to the elementary functions of the latter” (James, 1890, p. 28).

Psychological constructionist models of the mind were developed in early years of psychology and have appeared

consistently throughout the history of our science, although they have tended not to dominate (for a review, see Gendron & Barrett, in press). They are grounded in the assumption that experienced psychological states are not the elemental units of the mind or the brain, just as fire, water, air, and earth are not the basic elements of the universe. Instead, psychological states are products that emerge from the interplay of more basic, all-purpose components. The importance of distinguishing between the function of a mechanism (or process) and the products that it creates (what the functions are in the service of or what they allow to emerge) is inherent to a psychological constructionist approach. The contents of a psychological state reveal nothing about the processes that realize it, in much the same way that a loaf of bread does not reveal the ingredients that constitute it.

### A RECIPE FOR PSYCHOLOGY IN THE 21ST CENTURY

The modern constructionist approach that I envision for psychology in the 21st century is grounded in a simple observation. Every moment of waking life, the human brain realizes mental states and actions by combining three sources of stimulation: sensory stimulation made available by and captured from the world outside the skin (the exteroceptive sensory array of light, vibrations, chemicals, etc.), sensory signals captured from within the body that holds the brain (somatovisceral stimulation, also called the interoceptive sensory array or the *internal milieu*), and prior experience that the brain makes available by the reactivation and reinhibition of sensory and motor neurons (i.e., memory). These three sources—sensations from the world, sensations from the body, and prior experience—are continually available, and they form three of the fundamental aspects of all mental life. Different combinations and weights of these three ingredients (plus others) produce the myriad of mental events that constitute the mind. Depending on the focus of attention and proclivities of the scientist, this stream of brain activity is parsed into discrete psychological moments that we call by different names: *feeling*, *thinking*, *remembering*, or even *seeing*.

*Perception* is the name for psychological moments in which the focus is on understanding what externally driven sensations refer to in the world. Researchers who are interested in understanding perception (“What is the object?”) and behavior (“How do I act on it?”) might ask how sensory stimulation from the body and prior experience with an external sensory array keep track of and impart meaning to the immediate sensations of light, air, vibrations, chemicals, and so on. Said another way, scientists are asking how the brain makes predictions about the meaning of the current sensory array from the world (Bar, 2007; Barrett & Bar, 2009), allowing us to know our relation to the immediate surroundings in a moment-to-moment way and act accordingly.

*Cognition* is the name for psychological moments in which the focus is on understanding how prior experiences are reinstated

<sup>5</sup>The same point can be made about other mental categories that exist in other cultural contexts.

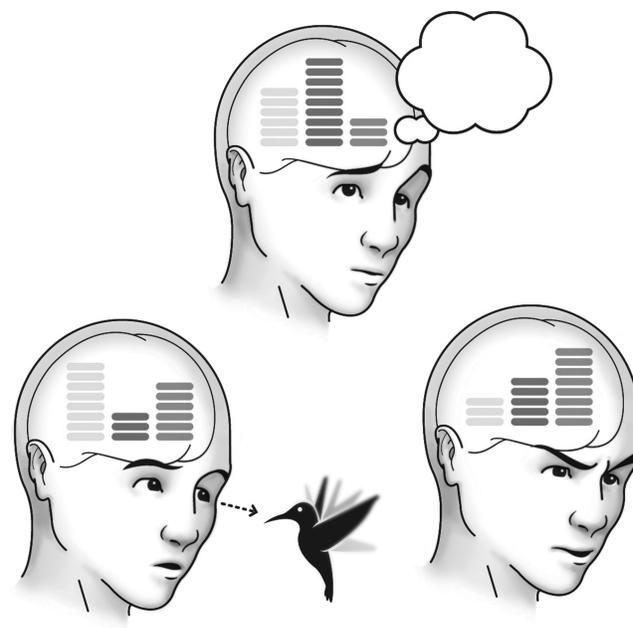
in the brain. When a person experiences the act of remembering, this mental activity is called *memory*. When they do not, it is called *thinking*. When the mental activity refers to the future, it is called *imagining*. And this mental activity provides a sense of self that continues through time. Researchers who are interested in understanding cognition (“How is the past reconstituted?”) typically ask how prior instances of sensory stimulation from the world, and from the body, are encoded and associatively recombined or reinstated for future use.

*Emotion* is the name for psychological moments in which the focus is on understanding what the internal sensations from the body represent. Researchers who are interested in understanding emotion experience (“How do I feel?”) examine how sensory information from the world and prior experience in the form of conceptual knowledge about emotion together create a context for what internal bodily sensations stand for in psychological terms.<sup>6</sup> Together, these three sources of input create the mental states named with emotion words. These conceptualized states are the mental tools that the human brain uses to regulate itself and the body’s internal state either directly or by acting on the world. This last piece is essentially a restatement of the model of emotion that I proposed in Barrett (2006b).

This very general description of mental life can be developed into a psychological constructionist approach that consists of five principles: (a) the mind is realized by the continual interplay of more basic primitives that can be described in psychological terms; (b) all mental states (however categorized) can be mapped to these more basic psychological primitives; (c) these basic psychological primitives correspond closely to distributed networks in the brain; (d) the mind is more like a set of recipes than like a machine; and (e) mental events are probabilistically, not mechanistically, causal.

### Principle 1: Psychological Primitives

The basic processes that constitute complex psychological categories can be described as psychologically primitive (to borrow a phrase from Ortony & Turner, 1990), meaning that they are psychologically irreducible and cannot be redescribed as anything else psychological. These psychological primitives are the ingredients in a recipe that produces a psychological moment—what we call an emotion, or memory, or thought, and so on—although they are not specific to any one kind of moment (see Fig. 1). Unlike the culturally relative complex psychological categories that they realize, psychological primitives might be universal to all human beings. (This is not the same as proposing that there are broad, general laws for psychology or for domains of psychology like



**Fig. 1.** Depictions of three brain states comprised of different combinations of the same three psychological primitives (represented in yellow, pink, and blue). Depending on the recipe (the combination and relative weighting of psychological primitives in a given instance) and a psychologist’s interest and theoretical proclivities, mental states are called *seeing* or *thinking* or *feeling*.

emotion or memory.) It might be possible to describe the operations that the brain is performing to create psychological primitives, but these operations would be identified in terms of the psychological primitives that they constitute.

Although identifying specific psychological primitives is beyond the scope of this article, elsewhere, my lab and I nominated three phenomena as psychological primitives. One psychological primitive might be what has been termed valuation, salience, or affect (producing a change in a person’s internal physical state that can be consciously experienced as pleasant or unpleasant, and arousing to some degree). Another might be categorization (determining what something is, why it is, and what to do about it). And a third might be a matrix consisting of different sources of attention (where attention is defined as anything that can change the rate of neuronal firing; for discussion, see Barrett, 2006b; Barrett, Tugade, & Engle, 2004; Duncan & Barrett, 2007).<sup>7</sup> And of course there are other psychological primitives.

<sup>6</sup>Researchers who are interested in emotion perception (“Is the rat fearful?”; “Is my dog sad?”; “Is my friend angry?”) focus their attention more like perception researchers, by asking how the perceiver’s bodily sensations and prior knowledge create the momentary experience of another creature’s behavior as emotional.

<sup>7</sup>There are a number of different sources of attention in the human brain, which neuroanatomist Marcel Mesulam (2000) refers to as an attentional matrix. In my reading of the neuroanatomy, this matrix is made up of at least five interconnected sources that can apply attention to a sensory neuron: (a) sensory stimulation from the world (what scientists term a bottom-up or sensory-driven or exogenous source of attention), (b) projections from lateral prefrontal cortex (what scientists term a top-down or goal-directed or endogenous source of attention), (c) projections from association areas that are involved with encoding prior experience (what might be called a memory-based source of attention), (d) projections from limbic areas and paralimbic cortex (which I have called an affective source of attention); and (e) projections from the ascending arousal systems. The lateral prefrontal cortex, the nuclei that originate the ascending

### Principle 2: An Ontology of Levels, Not Kinds

Complex psychological categories refer to the contents of the mind that can be redescribed as the psychological primitives that are themselves the products of neuronal firing. What psychology needs in the 21st century is a toolbox filled with categories for representing both the products and the processes at the various levels. Like David Marr's (1982) famous computational framework for vision (which has been oft-discussed in mind–brain correspondence; e.g., Mitchell, 2006; Ochsner & Lieberman, 2001), the categories at each level of the scientific ontology capture something different from what their component parts capture, and each must be described in its own terms and with its own vocabulary.<sup>8</sup> Unlike Marr's framework, as well as other recent treatments of mind–brain correspondence that explicitly discuss the need for a multilevel approach, it is assumed here that each level of the ontology must stand in relation to (and help set the boundaries for) the other levels. That is, there must be an explicit accounting (i.e., a mapping) of how categories at each level relate to one another. One such ontology of categories to describe mind–brain correspondence is suggested in Table 1 and is discussed as Principle 3.

### Principle 3: Networks, Not Locations

At the top of the ontology, complex psychological categories, such as *anger*, correspond to a collection of brain states that can be summarized as a broadly distributed *neural reference space*. A *neural reference space*, according to neuroscientist Gerald Edelman, refers to the neuronal workspace that implements the brain states that correspond to a class of mental events. A specific instance of a category (e.g., a specific instance of *anger*) corresponds to a brain state within this neural reference space. The individual brain states transcend anatomical boundaries and are coded as a flexible, distributed assembly of neurons. For example, the brain states corresponding to two different instances of *anger* may not be stable across people or even within a person over time.

Each mental state can be redescribed as a combination of psychological primitives. In this ontology, psychological primitives are functional abstractions for brain networks that contribute to the formation of neuronal assemblies that make up each brain state. They are psychologically based, network-level descriptions.<sup>9</sup> These networks are distributed across brain areas. They are

---

arousal systems in the brainstem and forebrain, and the selected thalamic nuclei all receive projections from affective circuitry and, in a way, can be considered indirect avenues for affective attention.

<sup>8</sup>Marr's abstract *computational level* of analysis specifies what processes do, without specifying how they do it. His *algorithmic middle level* provides a description of the logical steps that are needed to implement the computational level. The most basic *implementation level* of analysis specifies how to build something to carry out the sequence of steps outlined by the algorithms.

<sup>9</sup>Flour, stock, and butter are the basic ingredients of gravy, but flour and butter are first mixed together first to prepare a roux, which is then mixed into the stock to give a gravy its silkiness. So, a roux would not be considered a basic ingredient of the recipe, but it is a necessary stage or step for gravy without lumps. Similarly, large-scale brain networks contribute to circuits that have

**TABLE 1**  
*Mind–Brain Correspondence*

Psychology	Example	Brain
Complex psychological category	Emotion (e.g., anger, sadness, fear, etc.), cognition (i.e., thoughts, memories and beliefs), the self	Neural reference space
Psychological primitive	Core affect, categorization, executive attention	Distributed network
Momentary mental state	Specific instance of anger	Neural assembly

not necessarily segregated (meaning that they can partially overlap). Each network exists within a context of connections to other networks, all of which run in parallel, each shaping the activity in the others.

All psychological states (including behaviors) emerge from the interplay of networks that work together, influencing and constraining one another in a sort of tug-of-war as they create the mind. From instance to instance, networks might be differentially constituted, configured, and recruited. This means that instances of a complex psychological category (e.g., different instances of *anger*) will be constituted as different neuronal assemblies within a person at different times (which means that there is considerable intraindividual variability in addition to interindividual and cultural variability). It also means that phenomena that bear no subjective resemblance would be constituted from many of the same brain areas.

This scientific ontology has a family resemblance to other discussions of how psychology might map to brain function (e.g., Henson, 2005; Price & Friston, 2005). Like these other scientific ontologies, this one takes its inspiration from a number of notable neuroscience findings that together appear to constitute something of a paradigm shift in the field of cognitive neuroscience away from attempting to localize psychological functions to one spot or in a segregated network and toward more distributed approaches to understanding how the brain constitutes mental content. Specifically, the proposed ontology is consistent with: (a) research on large-scale distributed networks in the human brain (e.g., Friston, 2002; Fuster, 2006; Mesulam, 1998; Seeley et al., 2007); (b) neuroanatomical evidence of pervasive feedback connections within the primate brain (e.g., Barbas, 2007) that are further enhanced in the human brain, as well as evidence on the functional importance of this feedback (e.g., Ghuman, Bar, Dobbins, & Schnyer, 2008); (c) population-based coding and multivoxel pattern analysis, in which information is contained in spatial patterns of neuronal activity (e.g., Haynes & Rees, 2006; Kay et al., 2008; Norman et al., 2006); (d) studies that demonstrate considerable degeneracy in brain pro-

---

different levels of complexity and stand in relation to the network in a hierarchical fashion.

cessing (the idea being that there are multiple neuronal assemblies that can produce the same output; cf. Edelman 1987; Noppeney, Friston, & Price, 2004); (e) evidence that psychological states require temporally synchronized neuronal firing across different brain areas (e.g., Axmacher, Mormanna, Fernández, Elgera, & Fell, 2006; Dan Gläuser & Scherer, 2008) so that the local field potentials that are associated with neuronal synchronization are strongly correlated to the hemodynamic signals that are measured in functional neuroimaging (Niessing et al., 2005); and (f) the idea that a single neuron can code for more than one feature, depending on the assembly it is a part of (i.e., even individual neurons in the primary visual cortex might not be “feature detectors” in the strict sense of the term; e.g., Basole, White, & Fitzpatrick, 2003). By combining these novel approaches, it becomes clear that psychological states are emergent phenomena that result from a complex system of dynamically interacting neurons within the human brain at multiple levels of description. Neither the complex psychological categories nor the psychological primitives that realize them correspond to particular locations in the brain per se, and thus do not reconcile well with the kind of localization approach to brain function that was inspired by neuropsychology, remained popular in neuroscience throughout much of the 20th century, and continues to prevail today.<sup>10</sup>

The scientific ontology proposed here is also distinct from other scientific ontologies in three important ways. First, and perhaps most important, it deals with the existence of two domains of reality (one that is subjective and one that is objective) and their relation to one another.

Second, it helps solve the puzzle of why different sorts of behavioral tasks are associated with similar patterns of neural activity. For example, the so-called “default network” (which includes the ventral medial prefrontal cortex, dorsomedial prefrontal cortex, posterior cingulate and retrosplenial cortex, inferior parietal cortex, and at times, medial temporal structures like the hippocampus and lateral temporal cortex) shows increased activity not only during the spontaneous, highly associative “default” mental activity that is without an external stimulus, but also when one remembers the autobiographical past, envisions the future, or infers mental states in others; during self-referential processing and moral decision making; while imagining fictitious experiences (Buckner, Andrews-Hanna, & Schacter, 2008); during scene construction and contextual framing (Bar, 2007); and during the experience and perception of emotion (Barrett & Lindquist, 2008; Kober et al., 2008; Wager et al., 2008). Many functions have been proposed

for this circuitry, but one approach is to ask what all these tasks have in common: They associatively recombine bits and pieces of stored, prior experience to construct episodic projection or mental simulation. Sometimes this process constructs a memory of the past. Sometimes it constructs an imagined future. Sometimes it constructs the present in a manner that is reminiscent of Edelman’s notion of the “remembered present.” When constructing the present, this circuitry’s more general purpose might be to perform a meaning analysis of sensory input both from the body and from the world to indicate what those sensations stand for. When working together as a functional network, this circuitry’s more general purpose may be to impart meaning to the current sensory array based on prior, episodic experience. It allows the brain to predict what the current sensory information means based on that last time something like it was encountered and to formulate an appropriate response. A similar view is discussed in Bar (2007), who suggested that this circuitry functions to connect sensory input with memory to create predictions about what the sensory input refers to.

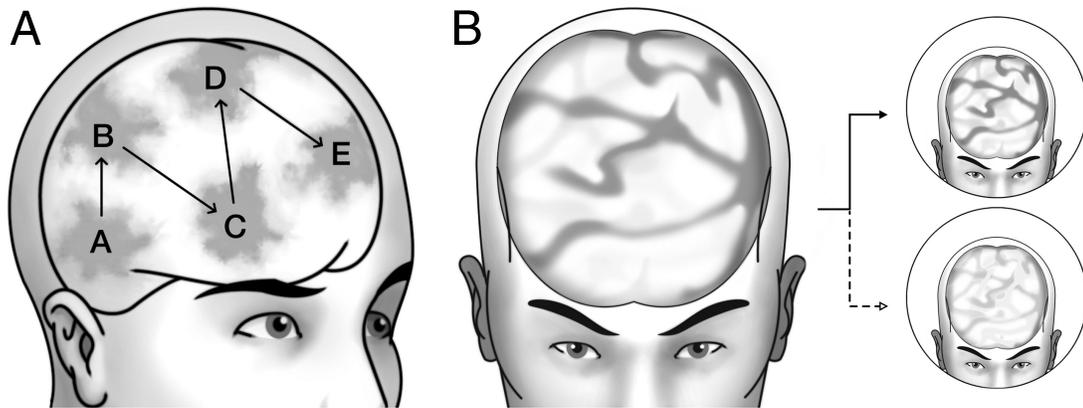
Finally, this psychological constructionist ontology also unifies a number of smaller scientific paradoxes with one solution. For example, it helps us to understand how perceptual memory can influence declarative memory tasks (even though implicit and explicit memory are supposed to be mechanistically different; e.g., Voss, Baym, & Paller, 2008), as well as how the same subjective feeling of remembering (Phelps & Sharot, 2008) or mental imagery (Kosslyn, Thompson, Sukel, & Alpert, 2005) can be produced in different ways (or, as discussed in Principle 4, with different recipes).

#### Principle 4: Recipes, Not Machines

In the psychological constructionist ontology proposed here, the metaphor for the mind in the 21st century is not a machine, but a recipe book. Psychological primitives are not separate, interacting bits and pieces of the mind that have no causal relation to one another like the cogs and wheels of a machine. Instead, they are more like the basic ingredients in a well-stocked pantry that can be used to make any number of different recipes (which make the mental states that people experience and give names to).<sup>11</sup> The products of the various recipes are not universal, although they are not infinitely variable or arbitrary either (e.g., bread can be baked with or without eggs, but you need some kind of grain to make bread what it is). The recipes are not universal. The recipe for *anger* will differ from instance to instance (with a context) within a person, and even if there is a modal recipe, it might differ across persons within a particular cultural context, as well as across cultural contexts. At the psychological level, however, the ingredients that make up the recipes might be universal (although how they function in conjunction with one another might not be). And as with all recipes, the amount of each ingredient is only one

<sup>10</sup>Even Brodmann, the originator of the much relied on Brodmann areas of cortical topography, did not believe in the idea that psychological functions are localized to specific, discrete brain areas. In 1909, in describing his views on function–location correspondence, he wrote, “one cannot think of their taking place in any other way than through an infinitely complex and involved interaction and cooperation of numerous elementary activities [...] we are dealing with a physiological process extending widely over the whole cortical surface and not a localised function within a specific region” (Brodmann, 1909/1994, p. 255).

<sup>11</sup>The metaphor of a recipe works for describing any emergent phenomenon, such as the interplay of genes and epigenetic factors that together produce observed phenotypic behaviors (Bateson, 1976).



**Fig. 2.** Two models of mental causation. A: In the mechanistic linear model, Psychological Process A localized in Brain Area 1 causes the separate and distinct Psychological Process B localized in Brain Area 2, and so on. B: In the probabilistic model, Brain State A at Time 1 (left panel) makes it easier to enter Brain State B (right top) than Brain State C (right bottom) at Time 2.

factor that is important to making the end product what it is. The process of combining ingredients is also important (e.g., are the dry ingredients added to the wet or vice versa, and are they whipped in, stirred in, or cut in?). As a result, it is not enough to just identify what the ingredients are, but also how they coordinate and shape one another during the process of construction.

The recipe analogy also helps us to see the scientific utility of distinguishing between complex psychological categories, psychological primitives, and neuronal firing, and to understand the relation between them. The category *bread* differs for a food critic, a chef, and a chemist. The job of a food critic is to communicate about bread. Food critics don't need to know the recipe for two different breads to describe which one has the preferred flavor and texture for a particular meal. That being said, it helps a food critic to know the recipe to explain why a particular bread tastes better for a particular meal. Also, it is more efficient (and less costly in both the economic and caloric sense) to change the taste of bread by modifying the recipe than by slathering a slice in butter and jam. A chef must know the recipe for bread to make one (e.g., a chef must know that flour and water are key ingredients, and that one must add yeast for bread to rise), but he or she need not know that flour and water interact to produce an interconnected network of coiled proteins (called *gluten*) that trap and hold the gases made by the yeast when bread is baking. That being said, every good chef knows that it helps to have some knowledge of chemistry, otherwise experimenting with changing the recipe can feel like shots in the dark. In a similar way, scientists must understand that the category *anger* differs for the social scientist, the psychologist, and the neuroscientist, but each of them would be better off knowing something about the category at the other levels of analysis.

#### **Principle 5: Probabilistic Causation, Not Linear Causation**

If mental events are constructed like recipes, then *goals* or *anger* or *memories* or *attitudes* do not cause *behavior* in the typical mechanistic way that psychologists now think about causation,

where “Psychological Process A” localized in “Brain Area 1” causes the separate and distinct “Psychological Process B” localized in “Brain Area 2”, and so on (see Fig. 2a). Instead a psychological-construction approach conceives of mental causation differently. Saying “anger caused aggressive behavior” might translate into the claim that a constructed mental state corresponding to “Brain State A at Time 1” (categorized as *anger*) increases the probability of occurrence of a second constructed mental state corresponding to “Brain State B at Time 2” (“slamming one’s fist against the table”; see Fig. 2b). This is similar to what connectionist modelers like Spivey (2007) and computational neuroscientists like O’Reilly and Munakata (2000) might argue. Alternatively, saying that “anger causes aggressive behavior” might translate into one brain state that is being constructed over time but, for reasons having to do with inferred intention, is experienced as two separate mental events (anger on the one hand, and the behavior of slamming one’s fist against the table on the other). Your internal bodily sensations might be categorizing this as *anger* because you experience these sensations in conjunction with your goals being blocked, and you clench your fingers to strike a table because you perform this action as you attend to a desire to cause harm. The point is that your subjective parsing of neuronal responses across space and time may not correspond to discrete bursts of neuronal activity. Either way, to say that “a person pounds a table because he is angry” is to give a reason for a behavior. Reasons are not causes for behavior and therefore do not constitute an explanation of it (for a discussion, see Searle, 2007).

Either option points to the implication that psychologists must abandon the linear logic of an experiment as a metaphor for how the mind works. In the classic experiment, we present a participant (be it a human or some nonhuman animal) with some sensory stimulation (i.e., a stimulus) that provokes some brain activity, and then we measure some response. Correspondingly, psychological models of the mind (and brain) almost always follow a similar ordering (stimulus → organism → response). The

relevant neurons are presumed to generally lie quiet until stimulated by a source from the external world. Scientists talk about independent variables because we assume that they exist separate from the participant.

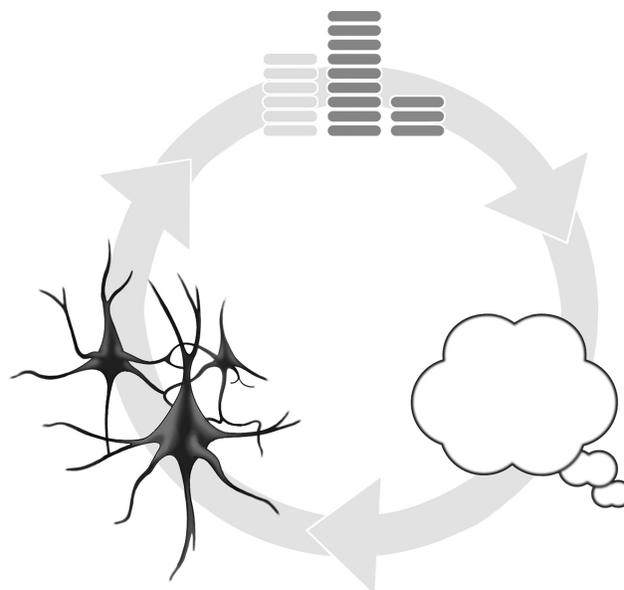
In real life, however, there are no independent variables. Our brains (not an experimenter) help to determine what is a stimulus and what is not, in part by predicting what will be important in the future (Bar, 2007). Said another way, the current state of the human brain makes some sensory stimulation into *information* and relegates the rest to the psychologically impotent category of “background noise.” In this way, sensory stimulation only modulates preexisting neuronal activity, but does not cause it outright (Llinas, Ribary, Contreras, & Pedroarena, 1998), and our brain contributes to every mental moment whether we experience a sense of agency or not (and usually we do not). This means that the simple linear models of psychological phenomena that psychologists often construct (stimulus → organism → response) may not really offer true explanations of psychological events.

The implication, then, is that mental events are not independent of one another. They occur in a context of what came before and what is predicted in the future. This kind of model building is easy for a human brain to accomplish, but difficult for a human mind to discover, because we have a tendency to think about ingredients in separate and sequential rather than emergent terms (e.g., Hegarty, 1992).

### LOSING YOUR MIND?

If complex categories such as *emotions*, *memories*, *goals*, and *the self* are collections of mental states that are created from a more basic set of psychological ingredients, it might be tempting to assume that psychology can dispense with the complex categories altogether. After all, a complex psychological category like *anger* will not easily support the accumulation of knowledge about how anger is caused if varieties of anger are constituted by many different recipes. This was certainly William James’s position when describing his constructionist approach: “Having the goose which lays the golden eggs, the description of each egg already laid is a minor matter” (James, 1890, p. 449). When it came to emotion, James was both a constructionist and a material reductionist. He espoused a token–token identity model of emotion, in which every instance of emotion that feels different can be ontologically reduced to a distinctive physical state, even when they are all members of the same common-sense category. (This approach stands in contrast to a type–type identity model in which every kind of emotion can be reduced to one and only one type of physical state.) Like James, some scientists believe that once we understand how such psychological events are implemented in the brain, we won’t need a science of psychology at all. Mental states will be reduced to brain states, and psychology will disappear.

But even William James can be wrong. In the constructionist account proposed here, a process should not be confused with the mental content it produces, but neither can it replace the



**Fig. 3.** Causal relations among levels. Networks of neurons realize psychological primitives that in turn are the basic ingredients of the mind. These basic ingredients construct instances of complex psychological categories like *the self*, *attitudes*, *controlled processing*, *emotion*, and so on.

need for describing that content. Said another way, the kind of material reductionism that James advocated should be avoided if for no other than the very pragmatic reason that complex psychological categories are the targets of explanation in psychology. You have to know what you are explaining in order to have something to explain. You have to be able to identify it and describe it well. A scientific approach to understanding any psychological phenomenon requires both description (“What is it?”) and explanation (“How was it constructed or made?”).

But the more important reason to avoid material reduction is that the various phenomena we are discussing (complex psychological categories, psychological primitives, and neuronal firing) each exist at different levels of scientific inquiry and do not exist at others. Complex psychological categories like *cognition* (*memories*, *beliefs*, *imaginings*, *thoughts*), *emotions* (*anger*, *fear*, *happiness*), and other varieties of psychological categories (*the self*, *attitudes*, and so on) are phenomena that fall squarely in the social science camp. They dwell at the boundary between sociology and anthropology on the one hand, and psychology on the other. Being observer-dependent categories that exist by virtue of collective intention (a group of human minds agree that anger exists and so it does), they are phenomenological distinctions. To understand them is to understand the nature, causes, and functions of these phenomenological distinctions (or the distinctions between whatever categories exist in your cultural context). They may not correspond to the brute facts of neuronal firing, but they are real in a relational way. If I categorize my mental state as a *thought* (instead of a *feeling*) and communicate this to you, you will understand something about

the degree to which I feel responsible for that state and the degree to which I feel compelled to act on it, as long as you belong to a culture where the emotion–cognition distinction exists (because in some cultures it does not). Furthermore, from a descriptive standpoint, we have to understand the function of these categories, both for the collective (which could be a dyad or a group of people) and for the individual. They can be epistemologically objective (i.e., studied with the methods of science) because they exist by consensus (in fact, consensus equals recognition in the science of emotion). And these categories may even have a biologically constructive quality of their own (see Fig. 3). As many neuroscientists have pointed out, humans are not born with the genetic material to provide a sufficient blueprint for the synaptic complexity that characterizes our brains. Instead, our genetic make up requires plasticity. Evolution has endowed us with the capacity to shape the microstructure of our own brains, perhaps in part via the complex categories that we transmit to one another as we create the social and cultural context. From this standpoint, psychological construction can also be understood in relation to social construction and neuroconstruction approaches.

At the other end of the continuum, there are brain states that are made up of collections of neurons firing with some frequency. Brain states are phenomena that fall squarely in the natural science camp. Brain states are observer independent—they do not require the mind they create to recognize them. In realizing the mind, they change from moment to moment within a person, and they certainly vary across people. But understanding how a neuron fires is not the same as understanding why it fires, and the latter question cannot be answered without appealing to something psychological.

In between are psychological primitives—the basic ingredients of the mind that are informed by both the categories above and below them. They are not completely observer independent, but neither are they free from the objective fact of the workings of the brain. Psychological primitives are caused by physical and chemical processes in the brain, but understanding these causes alone will never provide a sufficient scientific understanding of what psychological primitives are. They, too, have content that must be described for a complete understanding of what they are. That being said, when discussing psychological primitives, the structure of the brain cannot be ignored either. Psychological primitives will not necessarily replace complex psychological categories in the science of psychology, although sometimes they should. Whether complex psychological categories can be ontologically reduced to psychological primitives depends on the question that a scientist is trying to answer.

As a result of all this, it is possible to causally reduce complex psychological events to brain states and psychological primitives to distributed neuronal activity (what Searle, 1992, calls *causal reduction*) without redefining the mental in terms of the physical (what Searle calls *ontological reduction*). Just as knowing that a car is made of atoms (or quarks) will not help a

mechanic understand what happens when the motor stops working (Searle’s example), the firing of neurons alone is not sufficient for a scientific understanding of why a book is enjoyable, whether you enjoyed the book the last time you read it, why you like to read, or what joy feels like.

Now, it may be possible that the scientific need for psychological primitives is merely the result of the rudimentary state of neuroscience methods, and that even these psychological categories can be dispensed with once we have methods that can better measure cortical columns, which are (by conventional accounts) the smallest unit of functional specialization in the cortex (whose size is measured in microns).<sup>12</sup> It is possible that once we can measure columns in a human cortex while it is realizing some psychological moment (e.g., see Kamitani & Tong, 2005), “what” may finally correspond to “where” (i.e., process might correspond to one specific place in the brain), and it will be possible to ontologically reduce psychological primitives into the functioning of these units.<sup>13</sup>

But I suspect this will not happen, for four reasons. First of all, there is some debate over whether columns are, in fact, the most basic functional units of cortical organization. Dendrites and axons of the neurons within a column extend beyond those columns (DeFilpe et al., 2007; Douglas & Martin, 2007), which suggests that the functional units of the cortex may be somewhat larger than a column itself. Second, a single neuron within a column can participate in a number of different neuronal assemblies, depending on the frequency and timing of its firing (Izhikevich, Desai, Walcott, & Hoppensteadt, 2003), which suggests that a given neuron can potentially participate in a variety of different psychological primitives (meaning it is selective, rather than specific, for a function). Third, recent evidence suggests that specific neurons do not necessarily code for single features of a stimulus. A recent study in ferrets suggests that individual neurons (when participating in neuronal assemblies) appear to respond to more than one type of sensory cue, even in primary sensory areas where receptive fields for neurons are supposed to be well defined (as in primary visual cortex or V1; Basole et al., 2003). In addition, a recent study with rats demonstrates that there is a functional remapping of cells in the nucleus accumbens (part of the ventral striatum)—sometimes they code for reward and other times for threat, depending on the context (Reynolds & Berridge, 2008).

Finally, and perhaps most controversially, it may be a bit of an overstatement to assume that all humans have exactly the same

<sup>12</sup>Neurons with common functional properties lay stacked in a column, from white matter to cortical surface (Mountcastle, 1997). Minicolumns are between 25 and 80  $\mu\text{m}$  (about 11 neurons wide; Buldyrev et al., 2000) and columns (also called hypercolumns or macrocolumns) are aggregates of minicolumns ranging in size from about 300–900  $\mu\text{m}$  wide (Goldman & Nauta, 1977; Mountcastle, 1997). Columns vary in width, neuron makeup, as well as density and connectivity across species (Elston, 2007), with humans having the greatest connectivity between neurons both within a column and across columns.

<sup>13</sup>I do not mean to imply that the cortex alone is important to psychology; it goes without saying that subcortical areas are important.

nervous system. Human brains continue to expand at a rapid rate after birth (Clancy, Darlington, & Finlay, 2001), with most of the size increase being due to changes in connectivity with other neurons (Schoenemann, Sheehan, & Glotzer, 2005; for a review, see Schoenemann, 2006; but see Schenker, Desgouttes, & Semendeferi, 2005), including an increase in the size dendritic trees and density of dendritic spines (Mai & Ashwell, 2004). This means that although all humans may have the same brain at a gross anatomical level, the connections between neurons are exceptionally plastic and responsive to experience and environmental influence, producing considerable variability in brains at the micro level. The implication is that the neuronal networks that constitute psychological primitives will be molded by experience or epigenetic influences and that they may not be isometric across people.

If these kinds of findings forecast the future of neuroscience, then they suggest even more strongly that psychological primitives may be the best categories for consistently describing what the brain is doing when it realizes the mind. If one accepts this reasoning, then psychology will never disappear in the face of neuroscience.

## CONCLUSIONS

As a science of the mind, psychology is equipped with the ability to analyze how being human affects the process of doing science. We are in a better position than most to see how scientists make unintentionally biased observations of the world and have the capacity to correct for this all too common mistake. For the last century, psychology has largely used phenomenological categories to ground our scientific investigations into the mind and behavior. These categories influence the questions we ask, the experiments we design, and the interpretation of our data. We have spent the last century differentiating among psychological phenomena, improving on their labels, and searching for their correspondence in the natural world (i.e., locations in the brain).

Science always begins with common-sense categories. As I mentioned at the outset of this article, Aristotle assumed that fire, earth, air, and water were the basic elements of matter because these are the substances that he experienced. When modern physicists first looked at the world to discover the building blocks of matter, they saw discrete particles, like atoms. Later they identified parts of atoms, like electrons, protons, and neutrons. Then it was discovered that electrons were not really physical particles at all but that they are instead more like probabilistic energy states. Eventually physicists proposed the existence of something even smaller—particles they could not see and had to create new names for (e.g., quarks and leptons). Now, amidst much debate, many physicists believe that the universe is constituted of little strings vibrating in various modes across 11 different dimensions.

Time and space are experienced as separate phenomena and were once used by physicists to guide questions about the

material universe. That is, until Einstein changed the terms of the questions entirely with his theories of relativity. We now know that time and space are not rigidly independent categories—they are different ways of experiencing the same phenomenon. Psychology, of course, has studied time and space as subjective experiences for many years. Perhaps we should start approach *emotion* and *cognition* in the same way.

This is not to say that psychology's work in the 20th century was all for naught. All the work describing psychological categories delineates the phenomena to be explained, even if the categories themselves don't do the actual explaining. A major task of 21st-century psychology, however, is to link description to explanation.

In this article, I have argued that psychology has a crucial role in understanding how the observer-dependent mental phenomena (our human way of parsing the ongoing stream of consciousness into distinct psychological events) are created by a set of observer-independent neural phenomena. To accomplish this, psychology will need a major revision of concepts and framework. Specifically, in the psychology that I envision, we will have a hierarchy of categories: complex psychological categories that are observer-dependent and refer to collections of mental states that correspond to broadly distributed neural reference spaces in the brain; psychological primitives or building blocks of the mind that correspond to distributed networks and that combine to make various mental states; and the (more or less) observer-independent categories describing the anatomy and dynamics of the brain.

The psychological science that I envision for the 21st century is not purely a social constructionist science. Instead, I am suggesting that psychology is a young science, and, like any young science, we must divest ourselves of the assumption that human experience reveals the way the world (in this case, the brain) works. That being said, phenomenology has a place in psychology, even if it is not a causal place in the way that we typically understand cause. Complex psychological categories may be the targets of explanation, but this does not completely strip them of their scientific utility.

When it comes to understanding mind–brain correspondence, perhaps the empiricists, the rationalists, and the Kantians were all a little bit correct. Knowledge about the human mind is achieved from data captured by observing the natural world but not independently of our conceptual understanding of what those data mean for a human living in large, complex groups of other humans. Believing that a psychological phenomenon exists and is real can, in a certain sense, make it so.

*Acknowledgments*—The ideas contained in this essay were developed, in part, from discussions with many colleagues over the past several years. I especially thank Mahzarin Banaji, Moshe Bar, Larry Barsalou, Brad Dickerson, Dave DeSteno, James Gross, Dae-Shik Kim, Kristen Lindquist, Batja Mesquita, and Wendy Wood. Many thanks again to Moshe, Batja, Kristen,

and Dave and also to Jerry Clore, Jennifer Fugate, Paul Gade, Yang-Ming Huang, Kevin Quinn, Debbie Prentice, Mike Ross, and Maya Tamir, who commented on an earlier draft of this article. Preparation of this manuscript was supported by the National Institutes of Health Director's Pioneer Award (DP1OD003312), a National Institute of Mental Health's Independent Scientist Research Award (K02 MH001981), grants from the National Institute of Aging (AG030311) and the National Science Foundation (BCS 0721260, BCS 0527440), a contract with the Army Research Institute (W91WAW-08-C-0018), and by a James McKeen Cattell Award and a Sabbatical Fellowship from the American Philosophical Society. The views, opinions, and/or findings contained in this article are solely those of the author and should not be construed as an official Department of the Army or Department of Defense position, policy, or decision.

## REFERENCES

- Axmacher, N., Mormanna, F., Fernández, G., Elgera, C.E., & Fell, J. (2006). Memory formation by neuronal synchronization. *Brain Research Reviews*, *52*, 170–182.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–289.
- Barbas, H. (2007). Specialized elements of orbitofrontal cortex in primates. *Annals of the New York Academy of Sciences*, *1121*, 10–32.
- Barrett, L.F. (2005). Feeling is perceiving: Core affect and conceptualization in the experience of emotion. In L.F. Barrett, P.M. Niedenthal, & P. Winkielman (Eds.), *Emotions: Conscious and unconscious* (pp. 255–284). New York: Guilford.
- Barrett, L.F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science*, *1*, 28–58.
- Barrett, L.F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*, 20–46.
- Barrett, L.F. (2006c). Valence as a basic building block of emotional life. *Journal of Research in Personality*, *40*, 35–55.
- Barrett, L.F., & Bar, M. (2009). See it with feeling: Affective predictions in the human brain. *Philosophical Transactions of the Royal Society of London: Series B. Biological Sciences*, *364*, 1325–1334.
- Barrett, L.F., & Lindquist, K.A. (2008). *The emotion enigma*. Unpublished manuscript, Boston College.
- Barrett, L.F., Lindquist, K., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., & Brennan, L. (2007). Of mice and men: Natural kinds of emotion in the mammalian brain? *Perspectives on Psychological Science*, *2*, 297–312.
- Barrett, L.F., Mesquita, B., Ochsner, K.N., & Gross, J.J. (2007). The experience of emotion. *Annual Review of Psychology*, *58*, 373–403.
- Barrett, L.F., Tugade, M.M., & Engle, R.W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553–573.
- Barrett, L.F., & Wager, T.D. (2006). The structure of emotion: Evidence from the neuroimaging of emotion. *Current Directions in Psychological Science*, *15*, 79–85.
- Basole, A., White, L.E., & Fitzpatrick, D. (2003). Mapping multiple features in the population response of visual cortex. *Nature*, *423*, 986–990.
- Bateson, P. (1976). Specificity and the origins of behaviour. In J. Rosenblatt, R.A. Hinde, & C. Beer (Eds.), *Advances in the study of behavior* (Vol. 6, pp. 1–20). New York: Academic Press.
- Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic.
- Brodmann, K. (1994). *Localization in the cerebral cortex*. (L.J. Garey, Trans.). London: Smith-Gordon. (Original work published 1909)
- Buckner, R.L., Andrews-Hanna, J.R., & Schacter, D.L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.
- Buldryev, S.V., Cruz, L., Gomez-Isla, T., Gomez-Tortosa, E., Havlin, S., Le, R., et al. (2000). Description of microcolumnar ensembles in association cortex and their disruption in Alzheimer and Lewy body dementias. *Proceedings of the National Academy of Sciences, USA*, *97*, 5039–5043.
- Clancy, B., Darlington, R.B., & Finlay, B.L. (2001). Translating developmental time across mammalian species. *Neuroscience*, *105*, 7–17.
- Dan Clouser, E.S., & Scherer, K.R. (2008). Neuronal processes involved in subjective feeling emergence: Oscillatory activity during an emotional monitoring task. *Brain Topography*, *20*, 224–231.
- DeFelipe, J., Alonso-Nanclares, L., Arellano, J., Ballesteros-Yanez, I., Benavides-Piccione, R., & Munoz, A. (2007). Specializations of cortical microstructure of humans. In J.H. Kass & T.M. Preuss (Eds.), *Evolution of nervous systems: A comprehensive reference* (Vol. 4, pp. 168–190). New York: Elsevier.
- Douglas, R.J., & Martin, K.A.C. (2007). Mapping the matrix: The ways of neocortex. *Neuron*, *56*, 226–238.
- Duncan, S., & Barrett, L.F. (2007). Affect as a form of cognition: A neurobiological analysis. *Cognition and Emotion*, *21*, 1184–1211.
- Edelman, G.M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic.
- Einstein, A., & Infeld, L. (1938). *Evolution of physics*. Cambridge, United Kingdom: Cambridge University Press.
- Elston, G.N. (2007). Specialization of the neocortical pyramidal cell during primate evolution. In J.H. Kass & T.M. Preuss (Eds.), *Evolution of nervous systems: A comprehensive reference* (Vol. 4, pp. 191–242). New York: Elsevier.
- Frawley, W. (1992). *Linguistic semantics*. New York: Erlbaum.
- Friston, K.J. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, *25*, 221–250.
- Fuster, J.M. (2006). The cognit: A network model of cortical representation. *International Journal of Psychophysiology*, *60*, 125–132.
- Gendron, M., & Barrett, L.F. (in press). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review*.
- Ghuman, A., Bar, M., Dobbins, I.G., & Schnyer, D. (2008). The effects of priming on frontal-temporal communication. *Proceedings of the National Academy of Science*, *105*, 8405–8409.
- Gibbs, W.W. (2003). The unseen genome: Gems among the junk. *Scientific American*, *289*, 47–53.
- Goldman, P.S., & Nauta, W.J. (1977). Columnar distribution of cortico-cortical fibers in the frontal association, limbic, and motor cortex of the developing rhesus monkey. *Brain Research*, *122*, 393–413.
- Haynes, J.D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*, 523–534.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1084–1102.

- Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology*, *58A*, 193–233.
- Izhikevich, E.M., Desai, N.S., Walcott, E.C., & Hoppensteadt, F.C. (2003). Bursts as a unit of neural information: Selective communication via resonance. *Trends in Neurosciences*, *26*, 161–167.
- James, W. (1890). *The principles of psychology*. New York: Holt.
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*, 679–685.
- Kay, K.N., Naselaris, T., Prenger, R.J., & Gallant, J.L. (2008). Identifying natural images from human brain activity. *Nature*, *452*, 352–355.
- Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K.A., & Wager, T.D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, *42*, 998–1031.
- Kosslyn, S.M., Thompson, W.L., Sukel, K.E., & Alpert, N.M. (2005). Two types of image generation: Evidence from PET. *Cognitive, Affective, and Behavioral Neuroscience*, *5*, 41–53.
- Llinas, R., Ribary, U., Contreras, D., & Pedraarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *353*, 1841–1849.
- Mai, J.K., & Ashwell, K.W.S. (2004). Fetal development of the nervous system. In G. Paxinos & J.K. Mai (Eds.), *The human nervous system* (Vol. 1, pp. 49–94). New York: Elsevier.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.
- Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.
- Mesulam, M.M. (1998). From sensation to cognition. *Brain*, *121*, 1013–1052.
- Mesulam, M.M. (2000). Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. In M.M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 1–120). New York: Oxford University Press.
- Mitchell, J.P. (2006). Mentalizing and Marr: An information processing approach to the study of social cognition. *Brain Research*, *1079*, 66–75.
- Mountcastle, V.B. (1997). The columnar organization of the neocortex. *Brain*, *120*, 701–722.
- Niessing, J., Ebisch, B., Schmidt, K.E., Niesseng, M., Singer, W., & Galsuke, R.A.W. (2005). Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science*, *309*, 948–951.
- Noppeney, U., Friston, K.J., & Price, C.J. (2004). Degenerate neuronal systems sustaining cognitive functions. *Journal of Anatomy*, *205*, 433–442.
- Norman, K.A., Polyn, S.M., Detre, G.J., & Haxby, J.V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430.
- Ochsner, K.N., & Lieberman, M.D. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, *56*, 717–734.
- O'Reilly, R.C., & Munakata, J.A. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- Ortony, A., & Turner, T.J. (1990). What's basic about basic emotions? *Psychological Review*, *97*, 315–331.
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *2*, 148–158.
- Phelps, E.A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science*, *17*, 147–152.
- Price, C.J., & Friston, K.J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*, 262–275.
- Reynolds, S.M., & Berridge, K.C. (2008). Emotional environments return the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*, 423–425.
- Schenker, N.M., Desgouttes, A.M., & Semendeferi, K. (2005). Neural connectivity and cortical substrates of cognition in hominoids. *Journal of Human Evolution*, *49*, 547–569.
- Schoenemann, P.T. (2006). Evolution of the size and function of the human brain. *Annual Review of Anthropology*, *35*, 379–406.
- Schoenemann, P.T., Sheehan, M.J., & Glotzer, L.D. (2005). Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nature Neuroscience*, *8*, 242–252.
- Searle, J.R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Searle, J.R. (1995). *The construction of social reality*. New York: Free Press.
- Searle, J.R. (2007). *Freedom and neurobiology: Reflections on free will, language, and political power*. New York: Columbia University Press.
- Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*, 2349–2356.
- Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.
- Vallacher, R.R., & Wegner, D.M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*, 3–15.
- Voss, J.L., Baym, C.L., & Paller, K.A. (2008). Accurate forced-choice recognition without awareness of memory retrieval. *Learning and Memory*, *15*, 454–459.
- Wager, T.D., Barrett, L.F., Bliss-Moreau, E., Lindquist, K., Duncan, S., Kober, H., et al. (2008). The neuroimaging of emotion. In M. Lewis, J.M. Haviland-Jones, & L.F. Barrett (Eds.), *The handbook of emotion* (3rd ed., pp. 249–271). New York: Guilford.
- Wikan, U. (1990). *Managing turbulent hearts: A Balinese formula for living*. Chicago: University of Chicago Press.
- Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science*, *316*, 372–377.