

BOSTON COLLEGE O'NEILL LIBRARY  
DOCUMENT SERVICES - REQUEST FORM

DEPARTMENT CHARGE

Name Michael Ring Date 1/31/07  
E-Mail Address psych off @ bc.edu  
Campus Mail Address McGuinn 301  
Phone 552-4100 Department to be charged Psychology

NOTICE WARNING CONCERNING  
COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specific conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Signature Michael Ring

Delivery Methods:

Campus Mail  Pick-up  E-Mail as a:  PDF  JPEG  TIFF

Book/Journal Title Social Psychology and the Unconscious

Chapter/Article Title Chapter 4: On the Automaticity of Emotion

Author Lisa Feldman Barrett

Volume/Date \_\_\_\_\_ Pages 173-217

Hold originals for pick-up by patron Michael Ring

Return originals via campus mail

Special Instructions:

Do not write below this line

# 4

## On the Automaticity of Emotion

LISA FELDMAN BARRETT, KEVIN N. OCHSNER, and  
JAMES J. GROSS

Any emotion, if it is sincere, is involuntary

Mark Twain

**T**he year is 1846. You are a stagecoach driver. All is quiet and peaceful as the red sun sets beyond the horizon, the coach rumbles contentedly along, and sagebrush cast long shadows across the lonesome prairie. Hidden in the shadows, however, is a rattlesnake. Disturbed from its slumber by the horses, the rattler suddenly strikes out, scaring the horses into a fearful, frenzied sprint. Out of control, the stagecoach careens towards the edge of a sheer cliff. First you try to soothe your steeds, but they cannot hear you. Then you try to forcibly rein them in, but their strength is too great. Life itself hangs in the balance as you grimly struggle to control the careening stagecoach.

The distinction between wild stagecoach steed and wily stagecoach driver in many ways mirrors the distinction between feeling and thinking embedded within Western Culture. Emotions are assumed to be primitive, automatic, animalistic entities dwelling within us that the more developed human part of our minds come to know about and control. The notion that feeling is first, fast, and feral traces back to biblical stories of the First Family and their misbegotten emotional impulses to taste that tempting fruit. These ideas about emotion continue in modern-day stories of lovers driven mad with jealousy, businessmen blinded by greed, and widows overcome with grief. As Plato suggested long ago in the *Phaedrus*, in each of these cases, our emotions, like wild horses, drive us to emotional places we do not deliberately choose to visit and thus must be harnessed and restrained.

In the first section of this chapter, we outline the commonsense view that emotions automatically play themselves out when we encounter certain situations. We describe how this commonsense view – with varying degrees of elaboration and complexity – forms the basis of a consensual view of emotion that pervades much of the scientific inquiry into emotion. We refer to this consensual view as “the modal model” of emotion. In the second section, we argue that although the

“modal model” has much to recommend it, mounting evidence suggests that it has several important limitations. In the third section, we argue that the field needs to move beyond a search for entities that conform neatly to our intuitions about what automatic emotions “must” be like. We call for a richer examination of the bottom-up and top-down processes that together give rise to emotion, and suggest that a constraint satisfaction approach may provide the conceptual framework that is needed in order to move beyond the modal model.

### FROM COMMON SENSE TO A “MODAL MODEL” OF EMOTION

#### *The Commonsense Approach to Emotion*

As common sense has it, emotions are triggered automatically, overcome us, and cause us to act. We yell because we are angry. We cry because we are sad. We jump because we are afraid. Anger, sadness, and fear cause and explain our behavior, just as lightning causes and explains thunder. As the pent up electrical energy of an emotion is discharged, the result is a largely inescapable set of stereotyped outputs that occur in a rapid, involuntary, and automated fashion. Prior knowledge, expectations, beliefs, or any other cognitive input have little impact on the process. You might know that lightning is about to strike, but you can't stop it from happening, and plugging your ears will not keep you from hearing the thunderous blast. As a consequence, emotions such as anger happen to you, and overcome you, rapidly overriding whatever else you might have been doing, thinking, and feeling. From a first-person perspective, the conscious experience of emotion (the feeling) is taken as clear evidence that the causal mechanism – the “emotion” – was triggered. Feeling angry is evidence that the anger mechanism has fired. What is more, anger seems to overtake others in much the same way. When observing others, expressive behavior is seen as evidence that an emotion is triggered. The given quality of our own experience, and the way that emotion seems to control others without their awareness, seems to provide proof for the automaticity of emotional responses.

This folk conception of emotion underlies our everyday construal of emotions in self and others. Because we experience instances of anger (or sadness or fear) in ourselves and in others as having a rapid onset and a more or less ballistic trajectory, we believe anger (or sadness or fear) must exist as a primitive entity lying in wait within the brain or body, ready to spring forth automatically and at a moment's notice once the appropriate triggers are present. The commonsense idea of emotions as automatic eruptions is even evident in the criminal justice code, where “passion” is seen as a justifiable defense for violent crime. In the US, the “sudden heat of passion” constitutes adequate provocation that reduces an act of intentional homicide to an offense of voluntary manslaughter (Dressler, 2001, p. 527).

In this view, our emotions are rarely, if ever, the product of controlled, deliberate, and conscious thought. Although it is possible to “think ourselves” into an

emotional state, controlled processes typically serve to control, rather than to elicit, emotional responses. Any number of aphorisms admonish us to regulate our feelings either by deliberately thinking differently – “looking on the bright side,” “making a silk purse out of a sow’s ear,” “finding the silver lining in every dark cloud,” – or by acting differently – “putting on a happy face,” “putting a lid on it,” “showing some restraint,” “never letting them see you sweat,” “getting a grip,” or more generally by keeping ourselves from expressing the anger, sadness, or fear that we might feel inside. The very fact that we have to take control to regulate our emotional responses is further evidence for us that they are automatic, and it is precisely because we experience our emotions interfering with our more reasoned responses (that we identify as more essentially “human” in nature) that we experience our emotions as automatic, animalistic, and foreign.

More generally, our experiences of emotions erupting outside of our control, and our sense of agency and effort in shaping them, strongly supports our intuition that there are two fundamentally distinct forms of processing that characterize the human mind: automatic processing (which we share with other animals) and controlled processing (which is most developed in humans). As we will see, virtually every major scientific account of emotion incorporates our intuitions about dual-processing modes in the mind, and specifically our commonsense distinction between automatic elicitation and controlled regulation of emotion.

### *Two Major Scientific Approaches to Emotion*

Dual-process models pervade contemporary psychology (e.g., Barrett, Tugade, & Engle, 2004; Chaiken & Trope, 1999; Devine, 1989; Gilbert, 1991, 1998; Power & Dalgleish, 1997; Schacter, 1997; Sloman, 1996; Smith & DeCoster, 2000; Trope, 1986). A central tenet of such models is that behavior is determined by the interplay of automatic and controlled processing. Models vary in their specifics and emphasis, but all hold that responses to an event begin with the automatic (sometimes called nonconscious, implicit, or heuristic) processing of information. Such processing is assumed to be a default mode, initiated by the simple registration of sensory inputs, which in turn passively activate knowledge structures (called schemas, scripts, or concepts, or even internal goal states) that shape perception and action. Controlled (sometimes called conscious, explicit, or systematic) processing can determine, to a large extent, the degree to which automatic processing is expressed in thoughts, feelings, and behaviors. Controlled processing requires attentional resources, is volitional, is largely goal-driven, and can be used to modulate automatic processes when the outputs they produce conflict with valued goals.

As in other domains in psychology, emotion has proven hospitable ground for dual-process logic (Smith & Neumann, 2005). Indeed, in our view, a dual-process model lies at the heart of much of the scientific theorizing and research dealing with emotion for the past century. This dual-process framework can be summarized in a very simple way in Figure 4.1. Some event or goal-relevant stimulus (usually external, although it could be internal) triggers an emotion mechanism (or set of mechanisms), which, in turn, automatically produces a complex sequence of coordinated changes in the brain and body that constitute an emotional response.

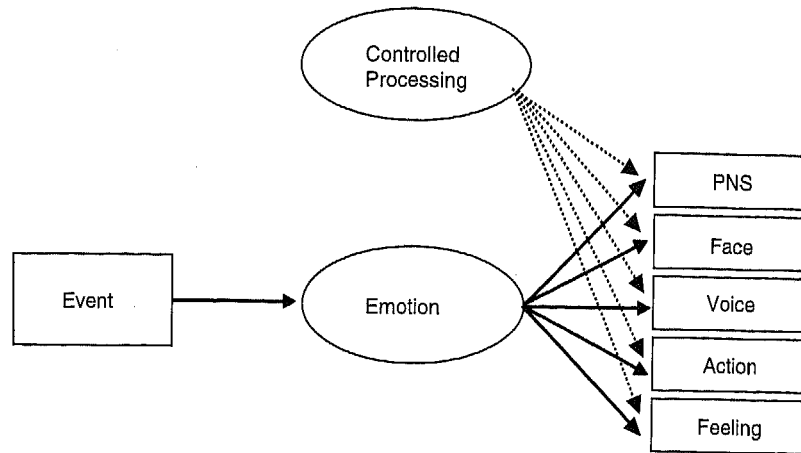


FIGURE 4.1 A dual-process view of emotion. PNS = peripheral nervous system activation.

Substitute for “emotion” any referent for a specific emotion category (“anger”, “sadness”, “fear”, and so on, in English), and you have a model of that emotion. Controlled regulation is thought to occur separately and modulate the extent to which a coordinated emotional response actually manifests in observed behavior.

Within this dual-process framework, two historically distinct (but often complementary) approaches to the study of emotion can be distinguished. One approach has focused on the output side of the emotion-generative process, namely the coordinated expression of complex patterns of behavior that comprise the observable, tangible, and socially impactful component of an emotional response. This has been referred to as the *basic emotion* approach. A second approach has focused on the input side of the emotion-generative process, namely the processing of environmental stimuli that gives rise to the emotional response depicted in Figure 4.1. This has been referred to as the *appraisal* approach. Despite differences in their surface features (for a review, see Scherer et al., 2001), these two approaches share two central assumptions. First, each of these approaches assumes that there are definable kinds of emotion (defined by the brain, or by the deep structure of situations). Second, these approaches assume that emotion generation is dominated by automatic processing (with regulation usually occurring after the fact). Given its ubiquity in the field, we refer to Figure 4.1 as the “modal model” [and in the past have referred to it as the “natural kind model” (Barrett, 2006b) or the “consensual model” of emotional responding (Gross, 1998)]. In the following sections, we describe the role the “modal model” has played in each of these two major approaches.

**The Basic Emotion Approach** One of the earliest modern examples of the basic emotion approach can be found in Darwin’s (1859/1965) “The Expression of the Emotions in Man and Animals.” Darwin’s ideas about emotion were infused

with commonsense beliefs about how mental states seek expression in, and therefore automatically cause, behaviors. Facial and other behaviors (such as baring the teeth in anger) were seen as vestigial expressions of emotion mechanisms that are homologous in human and nonhuman animals. Darwin focused on a small number of emotions (many of which are now referred to as “basic” emotions), stressing the universality as well as the phylogenetic and ontogenetic continuity of their expressions.

William James (1884, 1890, 1894) famously disagreed with Darwin (and commonsense) as to the correct temporal ordering of the elements in Figure 4.1. Instead of the sequence depicted in Figure 4.1, James argued that an individual’s emotional response was best characterized by the sequence: stimulus → physiological response → experience. Notwithstanding this notable difference, James clearly retained the core notion that emotion involved an automatic release of a coordinated set of responses to relevant stimuli (with the embodiment of those responses producing the experience of emotion). As William James himself put it:

The nervous system of every living thing is but a bundle of predispositions to react in particular ways upon the contact of particular features of the environment . . . The neural machinery is but a hyphen between determinate arrangements of matter outside the body and determinate impulses to inhibition or discharge within its organs . . . Every living creature is in fact a sort of lock, whose wards and springs presuppose special forms of key – which keys however are not born attached to the locks, but are sure to be found in the world near by as life goes on. And the locks are indifferent to any but their own keys. (1884, pp. 190–191).

This quote nicely captures the modern idea of cognitive impenetrability – meaning that emotion circuits fire in an obligatory way once triggered by sensory information about a stimulus, and occur regardless of the context. Onset is rapid, involuntary, and requires little or no attention. Although James concentrated on developing the automatic elicitation side of the emotion equation, he also incorporated the commonsense notion that controlled processes may come into play, such as when we regulate emotion after-the-fact, by limiting its expression. As James saw it: “refuse to express a passion, and it dies.” (James, 1884, p. 197).

Later models built more directly on this elicitation-regulation distinction, and further developed the dual-process metaphor for emotion processing. Cannon (1927, 1931) and Bard (1928; Bard & Rioch, 1937), who proposed one of the earliest psychological models of emotion localization in the brain, argued that the emotional part of the brain (in their view, the hypothalamic circuit including the thalamus and hypothalamus) produced responses that could be down-regulated by evolutionarily more recent neocortical regions. Papez (1937) similarly argued for top-down cortical regulation of subcortical emotional responses, and MacLean (1949) continued this tradition, positing that the newer “mammalian” part of his triune brain architecture exerted top-down regulatory control of the emotional responses that issued from the older and more primitive “reptilian” and “old mammalian” parts of the brain.

Although concepts such as reptilian and mammalian are no longer part of

contemporary basic emotion models, these models have nonetheless retained an emphasis on subcortical structures in the generation of emotion (e.g., Panksepp, 1998). Perhaps the best-known example has been offered by LeDoux (1996), who demonstrated that links between sensory systems and the amygdala are necessary and sufficient for the expression of conditioned fear, but who, along with others, has shown that cortical areas (particularly medial prefrontal cortex) are involved in expressing contextual learning that inhibits the conditioned fear response (Milad & Quirk, 2002; Morgan, Romanski & LeDoux, 1993; Quirk & Gehlert, 2003; Quirk, Likhtik, Pelletier, & Paré, 2003).

A dual-process metaphor can also be clearly seen in the family of models that comprise the modern-day “basic emotion” approach to emotion. Like Darwin and James, basic emotion models focus more on the ways in which emotional responses are automatically elicited. Boiled down to their essential ingredients, these views posit that each kind of “basic” emotion issues from a neural program or circuit, hardwired at birth, homologous with circuits found in nonhuman mammals, that is responsible for an automatic syndrome of hormonal, muscular, and autonomic effects that constitutes the distinctive signature of an emotional response. Although the specific set of “basic” emotions varies somewhat across emotion theorists, there is agreement that specific “basic” emotion mechanisms correspond to English emotion categories (e.g., “fear,” “sadness,” “disgust,” “anger,” or happiness”). This small set of stereotyped, automatic emotion responses are regulated after the fact, usually by means of controlled attentional processes shaped by epigenetic influences, such as context and learning history.

Perhaps the best-known example is Ekman’s neurocultural model (Ekman, 1972), which describes emotions as issuing from “affect programs” (Tomkins, 1962) that, once triggered by an eliciting stimulus, direct a complex pattern of coordinated outputs to produce a stereotyped emotional response. Panksepp’s (1998; Panksepp et al., 2000) neurobiological model takes its lead more directly from MacLean’s (1949) triune brain concept, but is similar, in principle, to the neurocultural model. Panksepp argues for different “basic” emotion systems (seeking/expectancy, rage/anger, fear/anxiety, lust/sexuality, care/nurturance, panic/separation, and play/joy). Each kind of emotion is a separate, inherited, complex reflex that is hardwired at birth and causes a distinctive syndrome of hormonal, muscular, and autonomic effects.

At the core of both Ekman’s (1972) and Panksepp’s (1998; Panksepp et al., 2000) models is the idea that there is a hardwired set of emotion-specific mechanisms that fire automatically and thereby generate a suite of emotional responses. Over time, however, there has been softening of the emphasis on fixed, hardwired programs that govern emotions from birth. For example, both Ekman and Panksepp acknowledge that there is a greater range of human emotional responding than can be accounted for by a set of basic emotions. Recent developments in the neurocultural model have attempted to account for the complexity and subtlety of emotional life by arguing for families of emotion response (Ekman, 1992), or by suggesting a distinction between “primordial” and “elaborated” emotions (Keltner & Haidt, 2001), where the former are “basic” emotions that produce a stereotyped response signature, and the latter are responses that are more shaped by the

norms and social practices within a culture. Both also allow controlled processing to enter the picture, although primarily as a way of regulating the emotion-generative process. In the neurocultural model, culture not only influences the stimuli that trigger emotion programs, it also specifies display rules and regulatory outcomes, so that observed emotional responses display considerable cultural variation, even as the causal mechanisms are hardwired into the brain. Panksepp (1998) also allows for the environment to modulate emotional outputs in the form of cortical control of the basic emotion systems once they have been triggered.

**The Appraisal Approach** The dual-process metaphor can also be found in a second family of emotion models, which we refer to collectively as the *appraisal* approach. Appraisal models have been concerned with patterns of cognition that trigger an emotional response. Like the basic emotion approach, many models within the appraisal approach retain the commonsense distinction between automatic elicitation and controlled regulation, although they also incorporate the idea that automatic and controlled cognitive processes (or steps) can interact and give rise to emotional responses. Thus, when appraisal models unpack the input side of Figure 4.1, they typically describe a cognitive logic that involves both automatic and controlled components, although once the emotion is elicited, it is assumed to run automatically to completion.

In these models, emotions are a consequence of how people construe situations. Frijda (1988), one of the best-known and most influential appraisal theorists, calls this the “law of situational meaning.” Instead of assuming that a stimulus situation automatically triggers or releases a fixed emotional response (as William James had), appraisal models hold that intervening cognitive processes automatically elicit and determine the quality and intensity of emotional responses. Input an event with a particular meaning, and the result is an emotion of a particular kind. The cognitive processes that compute this meaning, formally known as appraisals, link the external world (an individual’s immediate situation) to the internal world (the individual’s goals, needs, and concerns). Appraisals diagnose whether the situation in question is relevant to the person’s well-being, and if so, identify the nature of that relevance, and trigger an emotion that will maximize the likelihood of producing a functionally effective response consistent with the organism’s most central concerns (Kappas, 2001; Smith & Kirby, 2001).

Appraisal models vary in terms of the logic and content of the cognitive judgments that are held to be necessary and sufficient to produce emotional responses. In some models, appraisals describe the way that a situation is experienced, and constitute a descriptive structure of which emotions are felt when; they do not, in and of themselves, indicate the processes by which the meaning is made or arrived at (e.g., Ortony, Clore, & Collins, 1988; Smith & Ellsworth, 1985). In this view, sadness occurs when a situation’s meaning involves loss, and fear when the meaning involves danger. Situational analysis, or appraisals, can be determined in any number of ways. A situation’s meaning can be determined with associative processing by reinstating an appraisal that derives from a similar situation experienced

in the past, or it can be computed on the spot using a rule-based analysis driven by features of the situation and the goals of the person (Clore & Ortony, 2000). Both types of processing can be automatic (Smith & DeCoster, 2000), although on-line computation allows the possibility for more controlled processing.

In other models, appraisals do more than describe the meaning of situations – they are a set of cognitive processes that literally generate an emotional response (e.g., Roseman, Antoniou, & Jose, 1996; Scherer, 1984). Even the appraisal-as-mechanism models differ in which appraisals are seen as necessary and sufficient, which combinations of appraisals elicit particular emotional responses, which categories of emotion are explained, and so on (Roseman & Smith, 2001). They also vary in the fixity of the appraisal process. Some assume that appraisals are made in a specific fixed sequence (e.g., Scherer, 1984, 2001), while others argue for more flexible ordering in appraisal processing (e.g., Frijda, 1986; Roseman et al., 1996), although often it is assumed that particular appraisals (whether a stimulus is pleasant or unpleasant, novel or familiar) come before others that can be more flexibly deployed (e.g., Ellsworth, 1991; Lazarus, 1991). They differ in the extent to which they relax the assumption of cognitive impenetrability (the idea that emotion elicitation is not influenced by factors such as prior knowledge, expectations, beliefs, or any other cognitive input). Appraisal models also differ in whether they consider these rule-based computations to *cause* an emotional response, to *constitute* the response, or to be a *consequence* of emotion processing (cf. Ellsworth & Scherer, 2003; Frijda & Zeelenberg, 2001).

Despite their differences, virtually all appraisal theorists hold that people are continually assessing situations for personal relevance, beginning with an evaluation of whether or not the stimulus is “good for me/bad for me” (Arnold, 1960; Lazarus, 1966; Mandler, 1984; Ortony et al., 1988; Roseman, 1984; Scherer, 1984; Smith & Ellsworth, 1985). The primacy of such an evaluation is consistent with the general idea that we automatically evaluate stimuli (e.g., Bargh, Chaiken, Gollwitzer, & Pratto, 1992; Fazio, 2001; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Ferguson & Bargh, 2004), as well as the specific notion that some aspects of emotional responding – at least those that are related to computing affective valence – are generated automatically (e.g., Berridge & Winkielman, 2003; Cacioppo et al., 1999). Appraisal theorists also hold that different situations that evoke the same appraisal pattern produce the same emotional episode. Each emotion is elicited by a distinctive pattern of appraisals (e.g., Arnold, 1960; Frijda, 1986; Lazarus, 1991; Ortony, Clore, & Collins, 1988; Roseman, 1984; Scherer, 1984; Smith & Ellsworth, 1985), and the pattern of appraisals, rather than a dedicated neural circuit, is responsible for generating the emotional response.

Appraisal models, like basic emotion models, initially retained the automatic elicitation – controlled regulation distinction. Arnold (1960), who was the first contemporary appraisal theorist, assumed that appraisals are an “intuitive and immediate” assessment of the stimulus situation (p. 182). An explicit debate about the importance of the automatic versus controlled processing in emotion generation (Lazarus, 1982; Lazarus & Folkman, 1984; Zajonc, 1980, 1984) refocused attention to the idea that the conceptual processing engendered by appraisal processes can occur automatically, and now most appraisal models incorporate a role

for simple, nonconscious appraisals in emotion generation. Even when appraisals occur in response to remembered or imagined events, the cognitive processing involved in the appraisal process is thought to proceed automatically. Particularly in models that propose appraisals as preconditions for activating emotional responses (rather than constituting the responses themselves), appraisals are seen as inputs to the neural circuit view characterized by basic emotion models. Once appraisals have been computed, an emotion is triggered in a way that is very similar to Figure 4.1.

Over time, however, appraisal theorists have shown increasing appreciation for the role of controlled processing in emotion generation, and models have more explicitly relied on the dual-process metaphor that is implicit in basic emotion and early appraisal approaches. For example, Leventhal & Scherer (1987) argued that two kinds of automatic processes generate emotional responses that are modified by a third type of controlled process. An initial sensory-motor level of processing implements a form of biologically prepared perceptual processing driven by innate, unconditioned, hard-wired feature detectors that give rise to reflex-like responses. A second level (thought to mediate the majority of emotional responses) implements schematic processes that automatically match current stimulus patterns to learned stimulus patterns to generate coordinated emotional responses. When these first two types of processes generate a response that is sufficiently intense to enter awareness, consciously guided conceptual-level processes come into play, deploying propositional knowledge to refine emotional responses. Conceptual processing is thought to become ever more automatized (like skill learning) with practice. Many other models similarly suggest that emotions can be generated by some combination of automatic and controlled processing (e.g., Clore & Ortony, 2000; Dalgleish, 2004; Power & Dalgleish, 1997; Smith & Kirby, 2001; Teasdale, 1999; Teasdale & Barnard, 1993; Wells & Matthews, 1994).

In relaxing the assumption of cognitive impenetrability, appraisal models usually allow for the possibility that controlled processing can implement appraisal logic (i.e., the rule-based evaluations that cause the resulting emotional response). The common idea in these models is that various forms of automatic processing (including low-level perceptual processing of stimuli and prior knowledge in the form of schemas) interact with more effortful processing to produce emotional responses (Clark & Isen, 1982; Frijda & Zeelenberg, 2001; Lazarus, 1991). In this way, most appraisal theorists seem to agree that humans play an active role in shaping their information processing, and can exert some control over emotion-generative appraisal processes.

### *The “Modal Model”*

In the past, basic emotion and appraisal approaches have been treated as opposing explanations for emotional responding (Ortony & Turner, 1990; Turner & Ortony, 1992). We believe that despite their differences, both approaches share two intuitively appealing assumptions that can be found in our commonsense ideas about emotion. These common assumptions comprise the core of what we refer to as

the “modal model” of emotion. The first of these assumptions is the idea that emotional responses can be characterized as belonging to a small set of discrete categories. The second is the assumption that there is a boundary between the automatic generation of emotion and its controlled regulation after the emotion itself has been triggered.

First, both basic emotion and appraisal accounts focus their attention on explaining a small set of discrete emotions. Although appraisal models acknowledge (at least in principle) the enormous variety in emotional responding and do not assume that particular emotions are basic in any biological way, most models organize emotional responding into the familiar set of discrete categories used by basic emotion theorists. Major research efforts have been directed at identifying the profile of appraisals for a fixed number of discrete emotions (anger, sadness, fear, and so on) that are very similar to the list discussed by basic emotion approaches (e.g., Roseman, 1984, 1991; Smith & Ellsworth, 1985). For example, although Lazarus (1966) initially rejected the idea that there are a limited number of categorically distinct “basic” emotions, he came to view that there are a small number of “relational themes” that correspond to discrete emotions (Lazarus 1991; Smith & Lazarus, 2001). Similarly, Ortony et al. (1988) discussed a large range of emotional responses, but proposed a hierarchical structure in which some emotions are just more differentiated versions of other emotions. Scherer (2001) suggested that we tend to use basic level categories like “anger” to refer to qualitatively different emotional states, some of which may be “modal” emotions whose appraisal profiles recur with some frequency (Scherer, 1994), whereas other emotions may follow from fluctuations in appraisal profiles that may yield a large number of different emotional responses (e.g., Scherer, 2001).

Second, both basic emotion and appraisal accounts rely heavily on the notion that emotions are generated automatically. Thus, both types of models posit that emotional responses act as an organizing force, “hijacking” the entire system (i.e. disrupting whatever other processes are operative at the time) to deal with the circumstances that elicited the emotion in the first place. Oatley and Johnson Laird (1987), for example, suggested that basic emotions are internal signals that disrupt ongoing cognitive processing and reset it into specific modes to deal with basic biosocial challenges. Once an emotion is triggered (whether computed by an emotion program or a set of appraisals), there results an inescapable, involuntary and automated set of synchronized changes in response systems that produce the signature emotional response (like the output side of Figure 4.1).

Both assumptions embodied by the modal model – that there are a small set of different kinds of emotion, and that emotions issue more or less automatically – have guided emotion research for the past century. They have guided the questions that researchers ask, the way that emotional responses are measured, and the interpretation of the data. The resulting research programs have been productive and important. No model is perfect, however, and as we shall see in the next section, there are both theoretical and empirical reasons for a course correction in the way that science approaches the study of emotion.

## A CRITICAL EVALUATION OF THE MODAL MODEL

The modal model has been a tremendously valuable organizing force in the field of emotion. The basic emotion approach has helped to define emotion as a topic worthy of study in its own right, facilitating the development of empirical methods for examining facial (e.g., Ekman & Friesen, 1978), vocal (e.g., Scherer, 1986), autonomic (e.g., Cacioppo, Klein, Berntson, & Hatfield, 1993), and central (e.g., Davidson & Irwin, 1999) aspects of emotional responding. It has served as the de facto yardstick against which competing accounts of emotion are evaluated. The appraisal approach has helped to establish the importance of personal relevance and meaning in triggering emotional responses, and has attempted to unpack the notion of ballistic, automatic action programs into a more complicated set of both automatic and controlled processes that together contribute to the generation of an emotional response.

Useful as the modal model has been, however, it is limited and limiting in at least two ways. First, as we describe in more detail below, the modal model privileges a relatively limited number of emotions, leaving large gaps in our understanding of the full spectrum of emotional phenomena in need of explanation. Second, as we see it, the available empirical evidence, guided by a research agenda defined by the modal model, does not uniformly support the core tenets of this model. Specifically, the modal view posits automatic processing mechanisms that do not dovetail neatly with our emerging understanding of the behavioral and biological bases of emotion. As a result, a comprehensive functional architecture for emotion that considers a complete scope of emotional phenomena, and specifies a testable set of functional and neural mechanisms, has, to date, failed to coalesce from this model. In the following sections, we critically consider these two important limitations of the modal model.

### *Are There a Limited Number of Discrete Kinds of Emotion?*

An account of emotion, according to Clore and Ortony (2000, p 32), “needs to do justice to the full richness and range of emotions that comprise human emotional life.” The modal model, however, focuses attention on just one part of the larger emotional landscape by considering a small number of kinds of emotion about which we can make inductive discoveries, and which conform to the event → automatic response pattern. In this way, the modal model leads us to restrict our scientific inquiry to characterizing only a fraction of our emotional life. This practice is consistent with the viewpoint held by many emotion researchers that emotions should be defined by species-general aspects. As a field, we ask questions like, “How many emotions are there?” “Which specific pattern of antecedent events, neural activity, physiology, and motor behavior defines each emotion?” and “How do we evoke pure instances of emotion, uncontaminated by contextual influences?” Guided by the modal model, we assume that *kinds* of emotions would reveal themselves if only we could find the right eliciting stimulus or measures (cf. Barrett, 2006b).

Although it is clear that people have experiences that they refer to as anger,

sadness, fear, and so on, there is also good evidence that they can experience many other varieties of emotional response as well – responses not readily classifiable as fitting one of the canonical emotion kinds or forms. Variability may be the norm, rather than the exception, and according to some evolutionary biologists, variability is the thing to be explained (Lewinsohn, 2000). Although movies and novels are replete with examples of full-blown canonical emotion responses, emotion scientists have yet to take an empirical tally of how often these occur in everyday life. It is just assumed that they occur often enough to justify an almost exclusive focus on them empirically. Certainly these stereotyped responses are rarely, if ever, seen in all their glory in the laboratory. What is more, as we describe below, behavior, as well as the physiology that supports it, is more context-sensitive, and linked to the requirements of the situation, than the modal model of emotion would lead us to expect (cf. Barrett, 2006b; Bradley, 2000; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Davidson, 1994; Lang, Bradley, & Cuthbert, 1990). Functional demands vary with situations, making it likely that instances of the same emotion can be associated with a range of behaviors (e.g., behaviors associated with fear can range from freezing to vigilance to flight). This observation suggests that there is considerable heterogeneity in emotional responses that might be called fearful (or angry, etc.).

Even putting aside the issue of whether important aspects of human emotional life are neglected by the modal model, there remains the question as to the success of the search for definable patterns of coordinated responses that characterize each kind of emotion. Despite a century of effort, and much to everyone's surprise, there has been little accumulation of evidence to support the hypothesis that emotions represent clearly defined kinds (for a review, see Barrett, 2006b). To appreciate this unexpected outcome, consider a key prediction of the modal model and one of the most compelling ideas in the psychology of emotion: the notion that emotional states have specific and unique patterns of somatovisceral changes.

Although individual studies have reported distinct autonomic correlates for different emotion categories (e.g., Christie & Friedman, 2004; Ekman, Levenson & Friesen, 1983; Levenson, Ekman, & Friesen, 1990), meta-analytic summaries generally fail to find distinct patterns of peripheral nervous system responses for each basic discrete emotion (Cacioppo et al., 2000). Instead, peripheral nervous system responses configure for conditions of threat and challenge (Quigley, Barrett, & Weinstein, 2002; Tomaka et al., 1993; Tomaka, Blascovich, Kibler, & Ernst, 1997), and for positive versus negative affect (Cacioppo et al., 2000; Lang, Greenwald, Bradley, & Hamm, 1993). What is puzzling here is the gap between common sense (of course different emotions affect my body differently!) and the available physiological findings. Studying the physiological bases of emotion is fraught with challenges (Levenson, 1988), and it is certainly possible that methodological problems are responsible for the impoverished empirical evidence for physiological distinctions among emotions. But it is important to recognize that an equally viable alternative explanation for the lack of consistent findings is that there are, in fact, no clearly demarcated and discrete kinds of emotion that are unambiguously distinguished by patterns of peripheral physiological responses.

This sentiment is reinforced by a similar gap between common intuitions that

emotions cause distinctive patterns of expressive behaviors and the relevant empirical data. One of the major research efforts in the basic emotion approach has been concerned with detecting and describing universal facial expressions of emotion (Ekman & Friesen, 1978; Ekman, Friesen, & Ellsworth, 1972; Ekman, Sorenson, & Friesen, 1969; Izard, 1971). This research has yielded evidence that posed facial expressions can be judged with considerably greater than chance accuracy by individuals from a wide range of Western and non-Western cultures (cf. Ekman, 1994; but see Elfenbein & Ambady, 2002). Although perceivers can reliably assign posed facial configurations to discrete emotion categories, alternative explanations for these findings have been offered (cf., Russell, 1994; Russell, Bachorowski, & Fernandez-Dols, 2003), including the idea that perceivers are imposing, rather than detecting, true categorical distinctions in the facial configurations that they rate (Barrett, 2006a). More important, however, is the fact that very limited systematic data exist concerning the *production* (as opposed to the perception) of emotion expressions across cultures. Even within a culture, facial electromyography measurements coordinate around positive versus negative affect (Cacioppo et al., 2000) or intensity of affect (Messinger, 2002), rather than discrete emotion categories per se. More generally, it has been suggested that expressive behaviors in mammals rarely broadcast fixed, encoded messages about the sender's internal state (Fridlund, 1994; Seyfarth & Cheney, 2003), suggesting facial movements and vocal signals may not necessarily "display" information about the sender's emotional state (cf. Russell et al., 2003), even though we perceive them as coordinated "expressions."

Neither is it clear that vocal sounds carry specific information about discrete kinds of emotion (for a review, again see Russell et al., 2003). Listeners also do better than chance at classifying acted portrayals of emotion in vocal cues (Banse & Scherer, 1986; Hess et al., 1988; Johnstone & Scherer 2000; Juslin & Laukka, 2003; Wallbott & Scherer, 1986), but these portrayals do not necessarily have the same acoustic characteristics that are observed in naturally produced vocal expressions (for a discussion, see Bachorowski & Owren, 2003). Furthermore, the acoustic properties of produced vocal cues give clues to a speaker's identity (Edmonson, 1987), indicate his or her arousal level (e.g., Bachorowski 1999; Bachorowski & Owren 1995; Kappas et al., 1991), and are thought to elicit affective responses in listeners (Bachorowski & Owren 2001) more than they give evidence about kinds of emotion per se.

In like fashion, a given instrumental behavioral response need not express a specific kind of emotion. For example, although fear may be associated with freezing, fear is associated with a number of other behaviors, ranging from vigilance to attack (for a review, see Bouton, 2005). The threat (or defense) system is organized so that an animal will engage in different behaviors depending on its psychological distance from a predator (Fanselow, 1994; Fanselow & Lester, 1988). Not only are different behaviors associated with the same emotion category, but one type of behavior can be associated with many categories. For example, varieties of aggressive behavior (e.g., defensive, offensive, predatory) are associated with different types of stimulus situations and are caused by different neural circuitry (Blanchard & Blanchard, 2003).

Evidence from reports of subjective experience also calls into question whether or not there are bounded and distinct kinds of experiences. Not only do people vary greatly in whether or not they distinguish between feelings of anger, sadness, and so on, in reports of subjective experience (Barrett, 1998, 2004; Feldman, 1995), but these reports can be decomposed into more elemental psychological properties, such as valence and arousal. In revealing valence, and to a lesser extent arousal properties, self-reports of experienced emotion produce a similar structure to that which is observed for psychophysiological and behavioral measures of emotion.

Perhaps most important, however, is the finding that physiological, behavioral, and experiential outputs for each emotion category are not as highly intercorrelated as one might expect based on the modal model (Mauss, Wilhelm, & Gross, 2004; for a review, see Bradley & Lang, 2000). Psychophysicists have long observed weak correlations across response systems (e.g., Weinstein, Averill, Opton, & Lazarus, 1968) and even within the same "response system" (e.g., Lacey, 1967). Recent studies similarly have found modest correlations among measures of emotional responding in the context of fear (Hodgson & Rachman, 1974; Lang, 1988; Rachman, 1984), exhilaration (Ruch, 1995), and surprise (Reisenzein, 2000). Although links between emotion experience and facial behavior have tended to be the strongest (e.g., Adelman & Zajonc, 1989; Blumberg & Izard, 1991; Rosenberg & Ekman, 1994), even these links are often modest (e.g., Fernández-Dols, Sánchez, Carrera, & Ruiz-Belda, 1997; Fridlund, 1994) and inconsistent (e.g., Casey, 1993; Chovil, 1991; Gross, John, & Richards, 2000; for a meta-analytic review, see Cacioppo et al., 2000). Despite much effort, then, there has been surprisingly little evidence generated for the modal model's prediction of strong response coupling in emotional responding.

As this brief review indicates, physiological, behavioral, and experiential data do not strongly support the notion that there are clearly identifiable discrete kinds of emotion. The accumulating neuroscience evidence echoes this observation, thus far failing to yield strong evidence of dedicated neural circuits for basic emotion categories (Barrett, 2006b). Although there is good evidence that specific behaviors – such as freezing, the baring of fangs or claws, or hair standing on end – each may depend upon specific brainstem and subcortical nuclei (e.g., Panksepp, 1998), there is little evidence to suggest that a single brain structure is responsible for the production of the complete range of behavioral changes that should be associated with any single emotion category (Barbas, 1995; Cavada, Company, Tejedor, Cruz-Rizzolo, & Reinoso-Suarez, 2000). Similarly, many different cortical and subcortical brain systems are capable of modulating physiological and behavioral correlates of emotion, such as heart rate and respiration or freezing, and no single response system appears to be linked exclusively to a single specific emotion (Cacioppo & Gardner, 1999).

Lesion studies suggest that the normal experience and perception of some emotions, such as disgust and fear, may depend critically upon the integrity of particular brain structures – the insula and amygdala, respectively – but expressive deficits following amygdala or insula lesions typically are not absolute, and imaging studies suggest that both of these structures also appear to participate in the

generation of other emotions, the perception of other types of stimuli, and may even participate in ostensibly cognitive processes, such as orienting attention, as well (Adolphs & Tranel, 1999; Adolphs, Tranel et al., 1999b; Anderson et al., 2003a,b; Anderson & Phelps, 2000, 2001, 2002; Calder, Lawrence, & Young, 2001; Gallagher & Schoenbaum, 1999; Hamann, Ely, Grafton, & Kilts, 1999; Phan, Wager, Taylor, & Liberzon, 2002). Furthermore, alternative explanations for these findings are rarely explored (e.g., Adolphs, Russell, & Tranel, 1999; for a review, see Barrett 2006b). Meta-analyses of studies examining the neural correlates of anger, sadness, fear, disgust, and happiness (Murphy, Nimmo-Smith, & Lawrence, 2003; Phan et al., 2002) also support the conclusion that no single brain structure is exclusively associated with any single kind of emotion, with the exception of a fear-amygdala correspondence that can be accounted for by alternative explanations (Barrett 2006a,b).<sup>1</sup>

In any view of the neural bases of emotion, it is expected that every individual case of anger (or any other emotion) can be associated with some pattern of neural activity. The important question is whether there are strong and consistent correspondences between particular emotions and particular processing systems in the brain, and to date, such correspondences have yet to be identified. It is possible that a number of methodological and theoretical factors at present limit our ability to draw inferences about the neural bases of emotional responses. These include the facts that the way in which emotions are elicited is not constant across studies (allowing for the possibility that variety in method produces variety in brain activation; see Phan et al., 2002, for evidence), that studies may employ stimuli (such as facial expressions) that do not elicit strong emotional responses (Ochsner & Barrett, 2001), that little care may be taken to separate emotion generation from emotion regulation, allowing participants, for example, to spontaneously regulate their responses (Cunningham, Raye, & Johnson, 2004), and that measures of multiple channels of emotional response are rarely collected and correlated with brain activation (Ochsner & Barrett, 2001; Ochsner & Gross, 2004).

Even with those caveats in mind, the existing evidentiary base is not supportive of the modal model's core claim that there is a hardwired set of emotion-specific generators. Behavior, experience, or peripheral and central nervous system responses do not show obvious categorical "footprints" for each kind of emotion. It is possible, of course, that kinds of emotion exist and will reveal themselves once scientists find the right eliciting stimuli or employ better measurement tools. Indeed, Skinner claimed that cognitive processes "didn't exist" because they could not be measured reliably, but then psychological science developed better methods, and now such a claim seems preposterous. It seems worth noting, however, that the self-report, behavioral, physiological, and neural evidence are consistent with one another and seem to point to the same conclusion, namely that it may be time to move beyond the modal model.

### *Is Emotion Generation Automatic?*

This second cornerstone of the modal model – namely that emotions are automatically generated – has great intuitive appeal. Indeed, we "see" evidence (or so

we think) of highly automatic and stereotyped emotional responses in ourselves, in others, and in non-human animals (such as our dogs and cats). But are our emotions generated automatically as the modal model suggests, leading us this way or that depending upon which emotion has been elicited by a particular context? In general, introspection does not give us privileged information regarding the causal mechanisms that give rise to our behavior (Nisbett & Wilson, 1977; Wilson & Dunn, 2004). Therefore, our experience of emotions as arising unbidden, and then taking us over, does not, in and of itself, constitute evidence that emotion generation is intrinsically automatic.

To be fair, the original definition of automaticity was phenomenological in nature. Automatic processing was characterized by the absence of any subjective experience of control during thought, feeling, and behavior. This idea began with James (1890) and was elaborated by Helmholtz (1925) who clearly described the elements of the subjective experience of automaticity in detail: automatic processing is marked by the absence of any subjective experience of awareness (you are unable to self-reflect on your processing attempts), agency (you do not experience yourself as the agent of your own behavior), effort (you do not experience processing as effortful), and control (you are unaware that automatic processes might be occurring and you are unable to counteract them). By contrast, varieties of controlled processing are defined by the presence of a feeling of awareness, agency, effort, or control.

These ideas were further developed by Bargh (1994), who argued that these four features would be better considered as separate, distinct qualities that can be true of any cognitive process, and can combine in a componential fashion to place processes somewhere along an automatic-controlled continuum. As a result, it is possible to speak of varieties of automatic processing or forms of conditional automaticity, which require a goal to be initiated but run to completion outside of awareness (Bargh, 1989). These essentially phenomenological characteristics continue to dominate the distinction between automatic and controlled processing (e.g., Bargh & Ferguson, 2000), such that *feelings* of control and the *operation* of controlled processes are often confused (Barrett, Tugade, & Engle, 2004).

If we define controlled processing the way that control of attention is defined, as processing that proceeds (either consciously or unconsciously) or is shaped by an internally represented goal state (as opposed to processing that is driven strictly by the stimulus properties of the situation), then evidence from the cognitive literature gives controlled processing a role in phenomena that we typically experience as automatic (Barrett et al., 2004). For example, when goal states affect the processing of information or behavior outside of conscious awareness (e.g., Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Chartrand & Bargh, 1996; Winter & Uleman, 1984), we would call this a form of controlled processing (because attention is being driven by a goal state). The goal itself need not be intended, nor in any way represented in consciousness for controlled processing to proceed. When the goal is intended, or is otherwise conscious, this can result in what social psychologists have called "unintended goal dependent automaticity" — the effects of such processing are unintended and are generally not represented in consciousness (for a discussion, see Bargh, 1989). The idea is that controlled

processing can operate even at early perceptual stages, influencing how sensory information is selected, taken in, and processed (for reviews, see Luck & Hillyard, 2000; Posner & DiGirolamo, 2000; Shiffrin, 1988), well before the subjective experience of “seeing,” “hearing”, or “feeling” an input is generated. This sort of goal-based, controlled attention can “tune” more automatic, stimulus-driven forms of attention, including the ability of a stimulus to capture attention (for a review, see Pashler, Johnston, & Ruthruff, 2001). In current views from cognitive psychology, automatic processing (i.e., that which is stimulus-driven) is often intrinsically influenced by top-down, controlled forms of attention, making it somewhat artificial to separate the two in practice.

In principle, therefore, there need be no correspondence between a conscious feeling of control and the operation of controlled processing in the case of emotion. Indeed, they can be thought of as orthogonal, which means that one cannot be used to indicate when the other is occurring. It is possible, therefore, that controlled processing (as we mean it in this chapter, to refer to goal-dependent processing) may be more central to emotion generation than we have supposed, even though we do not experience any sense of agency or control or intention to generate an emotional response as controlled processes contribute to the formulation of an emotion. For example, if one has the goal of cooperating with another person, one may be less likely to take offense (and get angry) when the person tells an off-color joke than if one has either a competitive interpersonal goal or no specific interaction goal at all. In the view we are advancing, the lack of anger does not result from emotion regulation, but rather, anger may not be generated in the first place.

What does the evidence say regarding automatic emotion generation? Although there is ample anecdotal evidence (including our own first-person experience) for the automaticity of anger, sadness, and so on, there is surprisingly little empirical evidence to support the idea that emotions are inescapable, involuntary sets of synchronized changes in response systems, as depicted in Figure 4.1. Furthermore, when researchers have attempted to elicit discrete instances of emotion in the lab, the methodologies used do not allow clear determination of whether the generation of emotion responses was free from deliberative processing.

One type of methodology involves scripted mood inductions that may include the deliberate use of mental imagery and recall of autobiographical experiences to generate emotions (e.g., Posse et al., 2003; Schneider et al., 1997). In such experiments controlled deliberative processes play a clear role in generating the experiences under investigation (for imaging evidence consistent with this view, see Phan et al., 2002). A second methodology presents participants with film clips thought to elicit a discrete type of feeling such as happiness or sadness (e.g. Gross & Levenson, 1995; Lane, Reiman, Ahern, Schwartz, & Davidson, 1997; Levesque et al., 2003; Reiman et al., 1997). Although participants might report feeling happy or sad in response to such films, they are free to consciously and deliberately think about and judge film contents, and may even choose to deliberately regulate their responses (Erber, 1996).

Part of the difficulty in determining whether emotions are generated

automatically or deliberately is that there is no clear dividing line between automatic and controlled processes, and commonly used experimental methods that typically allow behavioral researchers to distinguish between automatic and deliberative processing may be difficult to employ in the context of emotion. Consider, for example, the fact that one of the most commonly employed means of demonstrating automaticity for a primary task – which in this case would be the generation of an emotion in response to evocative stimulus – is to ask participants to engage in a secondary task that drains attentional resources. If performance on the primary task does not change when participants are placed under “cognitive load,” then we can infer that the processes supporting primary task performance are comparatively automatic in as much as they do not require the deliberate application of attention in order to proceed. In the context of emotion, however, even if the underlying processes that generate an emotion are not affected by the secondary task, our ability to measure behavioral correlates of emotion may be obscured: in the presence of a secondary task, self-reports of experience will be difficult to make and may likely change because participants are attending to stimuli relevant to the secondary task; autonomic responses might change because participants are more aroused when doing two things at once; and behavioral responses might change for the same reason.

Another commonly used methodology for evaluating automaticity involves comparing responses to subliminal and supraliminal presentations of stimuli. For investigating questions about emotion generation this methodology may be similarly problematic. At issue are the facts that it is difficult to verify that subliminal presentations resulted in no conscious perception of stimuli, that it is difficult to ask questions about emotional experiences generated by stimuli which participants did not perceive, and that the behavioral effects generated by subliminal presentations often are neither robust nor reliable (Merikle & Reingold, 1998; Pessoa et al., 2006; Phillips et al., 2004).

All told, there is little evidence for completely automatic instances of anger, sadness, fear, and so on. The emotion literature continues to rely on a categorical distinction between automatic and controlled processing that may be problematic, and methodological challenges make it difficult to interpret the existing literature. Some of these problems may be solved by reinterpreting the existing evidence in terms of conditional automaticity, where a given emotional response can be characterized in terms of various automaticity subcomponents, but this would not solve the more general problem that one of the basic postulates of the modal model – signature response patterns for each emotion – is not unambiguously supported by the available empirical evidence. Without such signatures it is difficult, if not impossible, to clearly identify when a given kind of emotion has occurred, and that makes it difficult to develop a cumulative body of scientific knowledge about emotion.

In contrast to the ongoing questions about automatic emotion elicitation, there is ample evidence that a simple form of affective processing – the computation of value and its effects – is largely automatic. Affect can be characterized by hedonic valence (positive–negative, pleasant–unpleasant), and in some models also by degree of activation or felt arousal (Cacioppo & Gardner, 1999; Russell & Barrett,

1999). Various theories suggest that a quick determination of whether a stimulus is good or bad is essential for identifying potential threats and orienting attention to potentially goal-relevant stimuli (LeDoux, 2000; Öhman & Mineka, 2001; Scherer, 2001). Although the representational basis of such valenced evaluations has never been specified precisely, theorists have speculated that valenced information is represented at various levels of the neuroaxis (for discussions, see Barrett, 2006c; Cacioppo & Gardner, 1999; Duckworth et al., 2002; Russell, 2003), including both stimulus–response associations and more abstract, semantic representations. The activation of such representations and their ability to guide or bias perception and action is often referred to as unconscious affect. Unconscious affect is produced by a set of action-oriented affective systems that allows a person to deal with immediate dangers and rewards (Gray, 2004).

Three kinds of behavioral evidence support the idea that people can automatically evaluate stimuli or events for affective value or valence. First, many behavioral studies have found that the subliminal or nonconscious presentation of a valenced stimulus can generate autonomic responses (Esteves, Parra, Dimberg, & Öhman, 1994; Öhman & Mineka, 2001), changes in the activity of facial musculature (Dimberg, Thunberg, & Elmehed, 2000) and behavior (Chen & Bargh, 1999; Mikulincer, Hirschberger, Nachmias, & Gillath, 2001), can bias perception of subsequently presented stimuli in a valence-congruent fashion (Murphy, Monahan, & Zajonc, 1995; Murphy & Zajonc, 1993), and can generate “mystery moods” that are misattributed to other causes (Chartrand et al., in press; Winkielman, Berridge, & Wilbarger, 2005). Second, brief supraliminal presentation of a valenced stimulus facilitates access to valence congruent behavioral responses that seem affective in nature (Chen & Bargh, 1999; Cunningham, Preacher, & Banaji, 2001; Fazio, 2001; Nosek & Banaji, 2001). Third, behavioral and autonomic indices of affective response may implicitly reveal changes in the evaluation of the stimulus not reflected explicitly in conscious experience (Bechara, Damasio, Tranel, & Damasio, 1997; Duckworth et al., 2002; Tranel & Damasio, 1993). Taken together, these findings indicate that stimulus valence may be computed rapidly, and can influence subsequent behavior and experience. It is important to note, however, that these data do not provide a firm foundation for broader claims about whether specific emotions (such as anger, sadness, or fear) are automatically generated.

## THE SHAPE OF THINGS TO COME

In the two preceding sections of this chapter, we first examined and then critiqued the core assumptions that have guided emotion research for much of the last century. In this section, we explore the shape of things to come, as emotion research moves beyond the modal model and develops newer models. With neuroscience as their foundation, we expect these newer models to extend well beyond a small set of “basic” emotions, and to avoid the reification of the automatic versus controlled processing distinction, embracing instead an understanding of the intrinsic interplay between top-down (driven by the state of the

organism) and bottom-up (stimulus-driven) properties of emotion processing. As a result, any given emotional response may be characterized by the extent to which it is experienced as automatic, or meets the various criteria for conditional automaticity. In our view, these newer models will provide a framework that may lead to different questions, and perhaps to better answers, about what emotions are and how they are generated. In the section that follows, we describe one family of models that we believe may be useful as emotion researchers begin to move beyond the modal model.

### *Parallel Distributed Processing Models*

Parallel distributed processing (PDP) models (also called neural networks or connectionist models) are designed to explain psychological functions by appealing to the manner in which processing is achieved in the brain. PDP models are neurally inspired, meaning that they do not actually attempt to model processing at the level of individual neurons in the brain, but that they are generally consistent with how the brain processes information. Scientists used to conceptualize the brain as a hierarchical set of specialized processing networks, but more recent neuroscience evidence suggests that the brain is a set of distributed, interacting networks. In a PDP model, networks of neuron-like units (called nodes) pass activation to one another in parallel. Nodes in a PDP model can represent information at any level of analysis. For example, in “localist” models, each node represents a type of psychological function or process (e.g., an instance of emotion). In “distributed” models, the psychological function or process is represented by a pattern of activation distributed across a group of nodes. The idea is that multiple brain circuits process different types of inputs in parallel, with the processing in each circuit limiting, shaping, and constraining the way in which the system achieves a single coherent “solution.” For the purposes of this chapter, the “solution” is an instance of emotion that suits the particular goals of the individual and constraints of the context.

Early attempts to understand emotion using PDP ideas were inspired by appraisal models, and were therefore firmly grounded in the modal model. Appraisal models made explicit reference to ideas from parallel distributed processing (Scherer, 2001), and nonlinearity and neural network modeling (Frijda & Zeelenberg, 2001). For example, Wehrle & Scherer (2001) used a “black box” computational model that consisted of a set of formulas or algorithms to calculate emotional outputs (e.g., facial or vocal behaviors) on the basis of concrete input parameters (such as a profile of appraisals).

Most of the PDP-inspired models that have been offered to date are “localist” to some extent. For example, Thagard and Nerb (2002) used a “localist” model where particular emotions were nodes within a larger processing system. Other models propose an emotion architecture where behavioral subcomponents of an emotional response (appraisals, physiologic reactions, facial movements, instrumental behaviors, and contextual information) cause an emotion in a way that is similar to what James originally had in mind. One early dynamic system model (Fogel et al., 1992) proposed that components of an emotional response are

computed and proceed in parallel fashion and impose mutual constraints on one another to produce an instance of emotion. The result was a self-organizing system that produced the now familiar emotions of anger, sadness, fear, and so on. A more recent dynamic systems model (Lewis, 2005) replaced appraisals, physiologic reactions, and so on, with other components such as arousal, action tendencies, attention orientation, and feeling, but the general idea remained the same: what people usually take to be subcomponents of an emotional response actually have a more causal role in emotion generation. They organize a putative emotion system into coordinated response patterns, thereby causing an emotional response.

A new class of PDP models of emotion with distributed representations more explicitly depicts emotions as emergent phenomena. Distributed models are more neurally plausible than localist models (because the brain uses distributed representations), and they have several important advantages (for a discussion see Queller, 2002). One significant advantage is that a new node or processing unit does not have to be created every time a new instance of emotion is encountered. Instead, activation patterns across a fixed set of units can be altered to represent a new instance. For example, Wagar & Thagard (2004) present a model of core affective responding where collections of nodes represent areas of the brain that are thought to interact to produce a valenced (positive or negative) response. Their model is based on ideas from Damasio's (1994) somatic marker hypothesis. A collection of nodes for the ventral medial prefrontal cortex (VMPFC) directs appraisals of and responses to the situation. A second set of nodes, these representing the amygdala, direct bodily states associated with the current stimulus situation. Together, these nodes compute an initial affective response. Nodes for the hippocampus (coding information about the current context) gate the initial affective information to the nodes representing the nucleus accumbens (NAcc, which directs motor outputs). If the hippocampus allows, the affective signal is passed to the NAcc, which then sends information to nodes for motor/effector sites that then produce context-consistent behavior. The affective signal information is also routed back to the VMPFC and other prefrontal areas to further direct subsequent processing. Although we may not agree with the assumptions this model makes about the functions associated with specific brain systems (see below), we believe that the use of neuroscience data to inform construction of computational models of emotion is an important direction for future work.

### *A Constraint Satisfaction Approach to Characterizing Emotion*

Although initial forays into modeling emotion using PDP networks have taken advantage of some of the properties of parallel distributed networks – for example, the idea that multiple brain circuits process different types of inputs in parallel – they may not yet have realized their full potential. One of the most intriguing properties of PDP networks is the way in which they reach an overall “solution” by taking into account multiple constraints provided by different types of information represented within the network. This process of finding the best fitting solution is known as *constraint satisfaction*. In this section we illustrate some ways in which

constraint satisfaction approach could play an integral role in future models of emotion that take into account current behavioral, physiological, neuro-psychological, and imaging work concerning the nature and heterogeneity of emotional responding and move beyond assumptions made by the modal model. A summary of our ideas is presented in Figure 4.2.

**A Processing Continuum vs. Categories of Processing** Constraint satisfaction models have been used to account for the same phenomena as traditional dual-process models, including stereotyping (Queller & Smith, 2002), impression formation (Kunda & Thagard, 1996), decision making (Simon & Holyoak, 2002), personality (Shoda, Tiernan, & Mischel, 2002), and the self (Carver & Scheier, 2002). In constraint satisfaction logic, processes are not categorically automatic or controlled. Instead, processes combine in a componential fashion so that a given processing event emerges somewhere along an automatic-controlled continuum, and can be characterized as having some degree of conditional automaticity (for a discussion of conditional automaticity, see Bargh, 1989, 1994). We believe that a constraint satisfaction model of emotional processing that is similar in structure to Wagar & Thagard (2004) holds great promise for the scientific study of emotion.

To illustrate how constraint satisfaction works, we borrow the example of perceiving a tomato from Kosslyn (1995). To see a tomato, your visual system first senses some object out in the world and encodes a set of object properties that constrain what the object might be. Let us say the property red is encoded. At this point, the object could be an apple, a tomato, a strawberry, a flower, a fire engine, a shirt, a book, etc. It cannot, however, be a banana, or any other object that is not red. At the same time, you encode the object as fragrant – this would rule out any object that does not have a fragrant scent (e.g., a fire engine, book, shirt, and so on). You also encode that it is round, ruling out all non-round objects, like flowers and strawberries. When you touch the object, it is soft and pliable, and therefore cannot be an apple. Each object property, or piece of information is a constraint on the way in which the stimulus is perceived, helping narrow down the space of potential objects to the one that most viably satisfies the constraints currently available. Not all sensory features or pieces of information provide constraints on the final processing solution, however. For example, the tomato might have a blemish on its skin, which is a constraint that does not match anything in memory, or that might even resemble best a blemish once seen on an apple.

It is possible that an instance of emotion emerges in a similar fashion, out of processes in the brain that attempt to satisfy and minimize differences between different representational constraints that are simultaneously active in a given instance of processing. That is, the generation of an emotion episode results from a heterogeneous network of bottom-up (stimulus-driven) and top-down (goal or organism-driven) processes that are organized into a coherent interpretation and action plan. All this occurs in parallel, and in real time probably happens in the blink of an eye. The result is an emotional episode that people experience more or less as a gestalt.

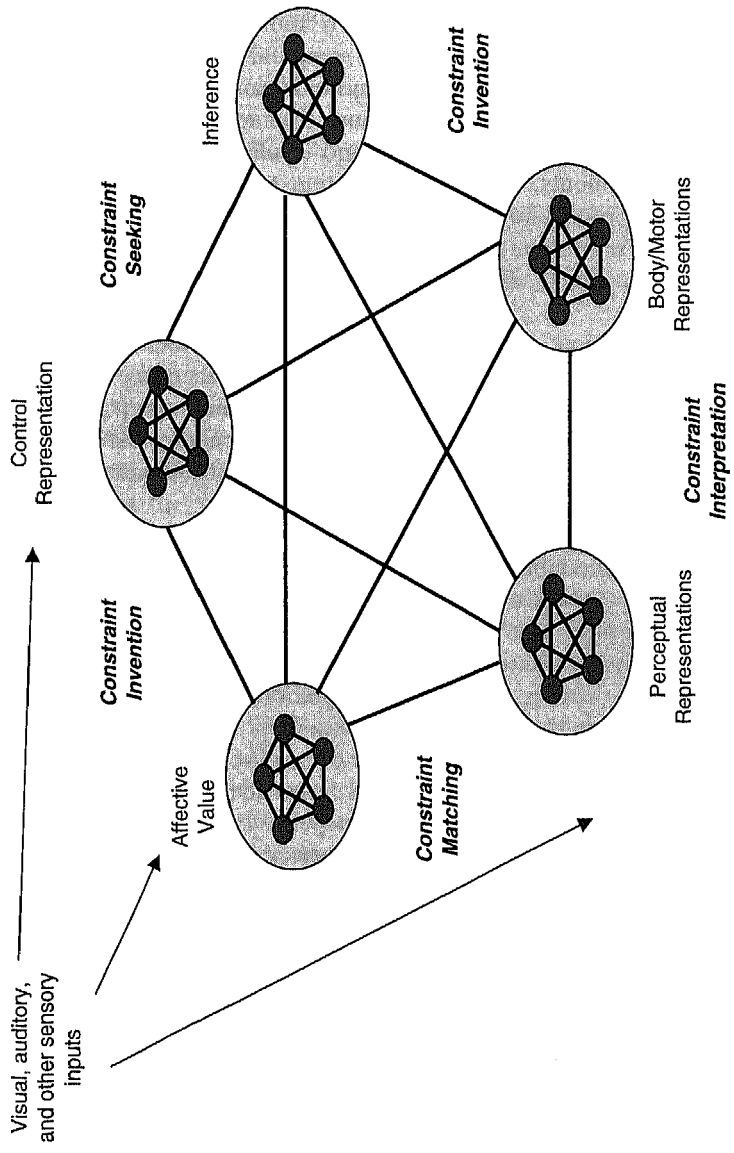


FIGURE 4.2 A constraint satisfaction framework for emotion.

**Bottom-up processing in emotion generation** Bottom-up processing describes how the sensory properties of the stimulus situation drive neural systems during the generation of emotion. These are processes that compute the initial affective value of a stimulus during stimulus recognition and stimulus identification, and are associated with numerous subcortical and posterior cortical areas of the brain including the amygdala, the ventral striatum, temporal parietal junction, and the hippocampus (located in the medial temporal lobe adjacent to the amygdala). Working together, these bottom-up systems can rapidly ensure responses in a variety of output systems, including those that are typically associated with emotional responding (like autonomic and endocrine changes, voluntary behavior, facial movements, etc.). They function to direct attention to stimuli that require further processing (e.g., Holland & Gallagher, 1999). In addition, outputs computed by bottom-up affective evaluators play an important role in establishing a person's core affective state at a given moment in time (Barrett, 2006a).

Constraint satisfaction can be used to describe two ways in which neural systems can generate emotion bottom-up, driven by aspects of the stimulus input. The first method can be termed *constraint matching*, which occurs when a stimulus first is encountered. The second method can be termed *constraint interpretation*, which follows after a stimulus is recognized via the constraint matching process.

*Constraint matching* (Kosslyn, 1995) involves matching available sensory inputs with stored unimodal representations in perceptual memory. In the tomato example that we borrowed from Kosslyn (1995), constraint matching proceeds via a number of bottom-up (or stimulus-driven) processes to extract information about shape, color, texture, and size. These features constrain one another during the matching process, leading to stimulus recognition. The result is that the sensory input implicates a single representation in perceptual memory, and the existence of the stimulus is recognized as familiar or not. At this point, the stimulus is not yet associated with a specific name and is not identified as belonging to a specific conceptual category (that is the next step).

Constraint matching is thought to be supported by posterior regions of the brain that represent different types of modality-specific inputs. Information about the spatial location, size, and movement of objects in the "dorsal stream" for visual processing links the occipital and parietal lobes. Information about the identity, form, and visual features of objects (such as facial expressions or body movements) in the "ventral stream" links the occipital and temporal lobes (Kosslyn & Koenig, 1992).

A person begins the initial computation of affective value during the constraint matching process. In the tomato example, this would involve estimating the probability that the stimulus features of this particular tomato predict subsequent threat (e.g., food poisoning) or reward (e.g., satiation of hunger). Affective value is computed with two neural systems that have been implicated in initially evaluating the valence of stimuli and generating appropriate responses in a predominantly stimulus-driven fashion. The centerpiece of the first system is the *amygdala*, shown in numerous studies to be important for rapidly detecting arousing stimuli

(Cunningham, Johnson et al., 2004; Morris, Öhman, & Dolan, 1999; Whalen et al., 1998), recognizing fearful and ambiguously threatening facial expressions (Adams, Gordon, Baird, Ambady, & Kleck, 2003; Adolphs, Baron-Cohen, & Tranel, 2002; Adolphs, Tranel, Damasio, & Damasio, 1994), fear conditioning (LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; LeDoux, 2000; Phelps et al., 2001), encoding and consolidating affectively charged positive and negative episodic memories (Cahill & McGaugh, 1995; Hamann, Ely, Grafton, & Kilts, 1999), and enhancing perception of arousing stimuli (Anderson & Phelps, 2001). Recent evidence suggests that the amygdala is involved in computing the predictive value of a stimulus (that is, its value in predicting threat or even reward; Kim et al., 2003, 2004).

The second system involved in computing the value of a stimulus involves the *striatum*, especially its ventral portion (known as the ventral striatum), which is important for the anticipation (Knutson, Fong, Adams, Varner, & Hommer, 2001; O'Doherty, Deichmann, Critchley, & Dolan, 2002) and receipt (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; Elliott, Friston, & Dolan, 2000) of rewarding stimuli with primary or secondary reinforcing properties, as well as learning about the rewarding properties of stimuli in general (Berridge & Robinson, 1998; Schultz, Tremblay, & Hollerman, 2000). More generally, the striatum may be involved in responding to events of immediate motivational significance. Once dopamine marks the salience of an event (such as when dopamine neurons respond strongly to an incoming event; Horvitz, 2000, 2002; Nieoullon & Coquerel, 2003), that event is given privileged access to voluntary motor output systems associated with the striatum and other parts of the basal ganglia.

Although the functional relationship between bottom-up affect encoding systems and perceptual memory systems has not yet been fully worked out and currently is an important topic of research (see e.g., Anderson & Phelps, 2001; George, Driver, & Dolan, 2001), it is clear that affective and perceptual encoding systems can interact in two ways during constraint matching. In some cases – for example those that involve complex perceptual inputs such as photographic images and films that require organization by perceptual representation systems – affect encoding systems may compute an affective value only after perceptual systems have recognized the input. In other cases, however, subcortical inputs from sensory organs to the amygdala and ventral striatum may support the rapid and independent computation of affective value, which can, in turn, constrain further processing in perceptual systems (Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003a; Vuilleumier, Armony, Driver, & Dolan, 2003).

Working together, these bottom-up systems can rapidly ensure responses in a variety of output systems, including those that are typically associated with emotional responding (like autonomic and endocrine changes, voluntary behavior, facial movements, etc.). In a sense, these are the systems that produce the unconscious affective response that prepares a person to deal with imminent threat or reward (Gray, 2004). In addition, outputs computed by bottom-up affective evaluators play an important role in establishing a person's core affective state at a given moment in time (Barrett, 2006a). That is, a person's affective state at ar

given moment is available but not necessarily experienced as feelings of pleasure/displeasure and activation/deactivation.

The second way in which emotions may be generated bottom-up may be termed *constraint interpretation* (Kosslyn, 1995). Constraint interpretation follows after a stimulus is recognized via the constraint matching process. Outputs from unimodal perceptual memory systems and assessments of valence and arousal from affective evaluation systems come together with category information in the conceptual system. As a coherent interpretation of their gestalt is formed, additional conceptual representations about a given object may be activated, including knowledge about the mental – in this case emotional – states likely to arise in conjunction with the object (Barsalou, Niedenthal et al., 2003), what types of situations elicit those emotional states, and so on. The nature of the interpretation formed is represented as what may be termed an attractor state in a neural network, and may depend upon the initial affective salience assigned to the stimulus, the type of emotion knowledge retrieved, what other representations recently have been activated in representational networks (which form additional constraints on the interpretive process), as well as other constraints that influence the overall state at the system, such as that resulting from the prior affective computation (which we think of as “mood”) or from processing of the broader situational context. Typically it is the case that the attractor state with the strongest and most coherent representation inhibits all other possible interpretive representations. In that sense, one attractor “wins,” and gains access to action planning. In this way, the core affective state resulting from the constraint matching process may evolve into an emotional episode: as conceptual knowledge about objects and emotions is activated, the situational and intentional causal antecedents of the core affective state may be specified, resulting in an emotion episode (for a discussion, see Barrett, 2006a).

Two types of brain systems have been implicated in the representation of conceptual and contextual information that constrains affective computations of the amygdala and ventral striatum in a bottom-up fashion. The first represents explicit, declarative knowledge about the kinds of emotional responses appropriate to a given situation as well as the kinds of behavioral intentions associated with those responses. This type of knowledge has been associated with the temporal parietal junction and the left inferior frontal lobe (Wagner et al., 2001), which may be activated when individuals access contextual knowledge to interpret emotional stimuli (Ochsner, Knierim, et al., 2004; Ochsner, Ray et al., 2004). The second type of system involves the hippocampus (located in the medial temporal lobe adjacent to the amygdala) and plays an essential role in encoding the relationships among multiple modalities of inputs that together comprise a representation of stimuli in their current episodic context (Cohen, Poldrack, & Eichenbaum, 1997; Eichenbaum, Otto, & Cohen, 1992; McClelland, McNaughton, & O’Reilly, 1995). Importantly, the hippocampus helps consolidate these representations in a format that can be flexibly and explicitly retrieved later on. These multimodal representations may provide a constraint on current behavior, registering a match between current and past situational contexts, thereby indicating affordances for certain kinds of appraisal (e.g., knowing that at work, a grimace from your boss means he is

not unhappy with you, but with a co-worker). In Figure 4.2, we have depicted conceptual knowledge as perceptual and motor representations, in line with recent research that the conceptual system consists of perceptual, rather than abstract, amodal, symbols (Barsalou, Simmons, Barbey, & Wilson, 2003).

**Top-down processing in emotion generation** In some cases, an emotional episode ends when all of the activated constraints are interpreted as a coherent whole and appropriate responses are selected and executed. In many cases, however, a specific emotional response will not emerge. The stimulus context may be ambiguous or impoverished so that information sufficient to coherently interpret an object has not been encoded, or it may happen that multiple competing and closely related interpretations of the inputs are simultaneously activated and it is not possible to easily resolve which interpretation is best suited to the particular circumstance. Goal states that are active at the time of processing provide an internal reason why the network does not easily settle into a single, clear, emotional response. Goal states can be consciously initiated or passively primed by the stimulus environment, but either way, they can bias the final solution by providing additional constraints to the system.

At this point, the system may actively use top-down processes to engage in *constraint seeking* (Kosslyn, 1995). Constraint seeking involves testing hypotheses about the possible meaning of a stimulus input given the activation of ambiguous or competing attractors. In the process of doing so, several brain systems may come into play, each of which is associated primarily with top-down processes involved in the generation of emotion.

The first is the orbital frontal cortex (OFC), which receives multimodal sensory inputs and is robustly interconnected with the amygdala and ventral striatum (Carmichael & Price, 1996; Ongur, Ferry, & Price, 2003). The OFC can be thought to implicitly provide top-down constraints during the generation of emotional responses, implicitly constraining them based on the flexible representations of situation appropriate behavioral goals that have been learned through experience and instrumental action (Ochsner & Gross, 2005). This conclusion is suggested by the fact that OFC neurons rapidly change firing properties (as stimulus-reward mappings change), whereas amygdala and striatal neurons do not (Rolls, 2000; Schoenbaum, Chiba, & Gallagher, 1998), and the fact that OFC lesions render responses to affective stimuli inflexible and situationally inappropriate, as evidenced by impaired social behavior (Beer, Heerey, Keltner, Scabini, & Knight, 2003; Rolls, Hornak, Wade, & McGrath, 1994), impaired extinction of conditioned fear responses (Quirk, Russo, Barron, & Lebron, 2000), an inability to alter stimulus-reward associations (Fellows & Farah, 2003, 2005; O'Doherty, Critchley, Deichmann, & Dolan, 2003; Rolls et al., 1994), and impaired evaluation of the relative costs and benefits of risky decisions (Bechara, 2003).

The second system is the lateral prefrontal cortex (LPFC), generally implicated in the maintenance and manipulation of information (Miller & Cohen, 2001). In the context of emotion, LPFC may play an important role in strategically accessing conceptual memories that aid in the deliberate appraisal of an event's emotional significance (Davidson, Jackson, & Kalin, 2000; Ochsner & Barrett,

2001; Ochsner, Knierim et al., 2004; Ochsner, Ray et al., 2004), such as when one needs to understand the kinds of emotions that might be experienced in a given context or when one tries to understand the current situation in terms of similar ones experienced in the past.

The third system related to top-down emotion generation is the medial prefrontal cortex (MPFC), and especially its dorsal portion (the evolutionary neurodevelopmental precursor of lateral prefrontal cortex, to be described next), which plays an essential role in drawing inferences about one's own or other individuals' current affective – or more generally, mental – states (Gallagher, Jack, Roepstorff, & Frith, 2002; Gusnard, Akbudak, Shulman, & Raichle, 2001; Lane, Fink, Chau, & Dolan, 1997; see Gallagher & Frith, 2003 and Ochsner, Knierim, et al., 2004 for reviews). The MPFC has interconnections with autonomic centers (Ongur & Price, 2000), and may provide an interface between higher cognitive representations that are informationally explicit on one hand, and comparatively visceral representations that are not explicit and provide “gut feelings” on the other. MPFC may come into play when one becomes aware of, reflects upon, monitors, and/or labels emotional states, including when emotional understanding requires drawing inferences about one's own behavioral intentions or the intentions of others (Ochsner, Knierim, et al., 2004).

The fourth system involved in constraint seeking is important for signaling when top-down processing is needed. The anterior cingulate cortex, and especially its dorsal portion (dACC), plays an important role in monitoring conflict between competing response tendencies and signaling the need for lateral prefrontal regions to implement control processes that bring performance into line with goals (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Cohen, Botvinick, & Carter, 2000). In the context of emotion, dACC activation may reflect the failure to achieve a single constraint-satisfied interpretation of an emotional stimulus, thereby triggering constraint seeking. dACC then may work hand-in-hand with LPFC to guide the controlled application of emotion knowledge to seek additional information (Ochsner & Gross, 2004, 2005) that may enable the system to settle into a stable attractor state.

The top-down processes that are used to seek information in memory or in the perceptual input may be used to generate emotion in the absence of any external inputs that have affective value. In such cases, *constraint inventing* (Kosslyn, 1995) may be used to generate an emotional response to a situation that otherwise would have been interpreted as fairly neutral. Constraint inventing involves the deliberate construction and maintenance of representations in working memory constructed from information stored in the conceptual system. Although very little research has directly investigated this mode of top-down emotion generation, we recently conducted a study that examined this issue directly. This study is described in the final section below.

### *Advantages of a Constraint Satisfaction Approach to Emotion*

Thus far, we have sketched how an instance of emotion can be thought of as an emergent phenomenon built out of processes that attempt to satisfy and minimize

