

response forcefully informs the organism about what might follow, which then motivates behavioral and mental coping.

Freud and Spence were about as different as theorists could be, but both came from very different sets of assumptions to the same insight: that affective outcomes influence thought and action, not by pushing from behind, but by pulling from in front. They discovered the power of prospection by which organisms can reorganize themselves to maximize efficiency and minimize energy expenditure through anticipatory change.

Consistent with this emphasis on the prospective power of emotion, Baumeister et al. (2007) proposed that the function of conscious emotional experience is really self-education. When people are embarrassed or humiliated, for example, their emotional reactions make the situation highly memorable. When the person anticipates that the same kind of situation will occur again, a fractional anticipatory affective response to that simulation allows preparation, planning, and choice (Barsalou 2008). As LeDoux² is fond of saying, “Emotion is memory,” by which he means that much emotional processing involves the memory of past emotional moments to motivate preparation for similar situations. Similarly, Baumeister et al. (2006) noted that “behavior chases emotion,” or, as emphasized in this chapter, emotion often motivates behavior through incentives that “pull” behavior for future satisfaction rather than by being “pushed” by current deficits.

SUMMARY

In the present view, emotions are seen, not as tightly organized, biological entities, but as sets of evaluative reactions to psychologically significant situations. Common human emotions are reactions to recurrent life situations. Some are common to all animals, whereas others are occasioned by the hypersocial nature of humans. The position was taken that emotions are most readily categorized in terms of the situations they concern, rather than in terms of responses, which depend on the context in which they occur. Other conclusions were that: (a) the sharp distinction between cognition and emotion has probably outlived its usefulness; (b) bodily and other kinds of feelings typically function as foretastes of things to come, providing anticipatory information that allows preparation and coping; (c) the fact that affective feelings are pleasant or unpleasant gives emotions motivational power; (d) the fact that people experience fractional parts of emotional reactions in anticipation of similar situations

allows learning from past emotions to serve as incentives for future behavior.

ACKNOWLEDGMENTS

This work was partially supported by grants from the National Institute for Mental Health (MH 50074) and the National Science Foundation (BCS-1252079) to Gerald L. Clore.

Thanks to David Reinhard and Philip Chow for comments on an earlier draft.

1.6 WHAT ARE EMOTIONAL STATES, AND WHAT ARE THEIR FUNCTIONS?

Edmund T. Rolls

EMOTIONS AS STATES ELICITED BY INSTRUMENTAL REINFORCERS

Emotions can usefully be defined (operationally) as states elicited by rewards and punishers that have particular functions (Rolls, 1999; Rolls, 2005b, 2014a). The functions are defined herein, and include working to obtain or avoid the rewards and punishers. A reward is anything for which an animal (which includes humans) will work. A punisher is anything that an animal will escape from or avoid. Rewards and punishers are the goals for instrumental actions. An example of an emotion might thus be the happiness produced by being given a particular reward, such as a pleasant touch, praise, or winning a large sum of money. Another example of an emotion might be fear produced by the sound of a rapidly approaching bus, or the sight of an angry expression on someone’s face. We will work to avoid such stimuli, which are punishing. Another example would be frustration, anger, or sadness produced by the omission of an expected reward, or the termination of a reward such as the death of a loved one. Another example would be relief, produced by the omission or termination of a punishing stimulus, such as the removal of a painful stimulus, or sailing out of danger. These examples indicate how emotions can be produced by the delivery, omission, or termination of rewarding or punishing stimuli, and go some way to indicate how different emotions could be produced and classified in terms of the rewards and punishers received, omitted, or terminated. A diagram summarizing some of the emotions associated with the delivery of a reward or punisher or a stimulus associated with them, or with

the omission of a reward or punisher, is shown in Figure Q1.6.1.

Before accepting this approach, we should consider whether there are any exceptions to the proposed rule. Are any emotions caused by stimuli, events, or remembered events that are not rewarding or punishing? Do any rewarding or punishing stimuli not cause emotions? We will consider these questions in more detail later. The point is that if there are no major exceptions, or if any exceptions can be clearly encapsulated, then we may have a good working definition at least of what causes emotions. Moreover, it is worth pointing out that many approaches to or theories of emotion have in common that part of the process involves “appraisal” (Keltner et al., 2013; Moors et al., 2013). In all these theories, the concept of appraisal presumably involves assessing whether something is rewarding or punishing. The description in terms of reward or punishment adopted here seems more tightly and operationally specified.

I consider elsewhere a slightly more formal definition than “rewards” or “punishers,” in which

the concept of “reinforcers” is introduced, and it is shown that emotions can be usefully seen as states produced by instrumental reinforcing stimuli (Rolls, 2014a). Instrumental reinforcers are stimuli that, if their occurrence, termination, or omission is made contingent upon the making of a response, alter the probability of the future emission of that response. Some stimuli are unlearned reinforcers (e.g., the taste of food if the animal is hungry, or pain); while others may become reinforcing by associative learning, because of their association with such primary reinforcers, thereby becoming “secondary reinforcers.”

I define emotion in terms of instrumental reinforcers when the behavior is under the control of a goal, the reward or punishment, partly because this is the situation in which the major evolutionary advantage of emotion is most apparent, as described later. If the behavior is being performed as a habit, using a stimulus–response association built up over many learning trials, then the behavior may be performed almost automatically, without much emotion, so habit-based instrumental responses are not central to my definition

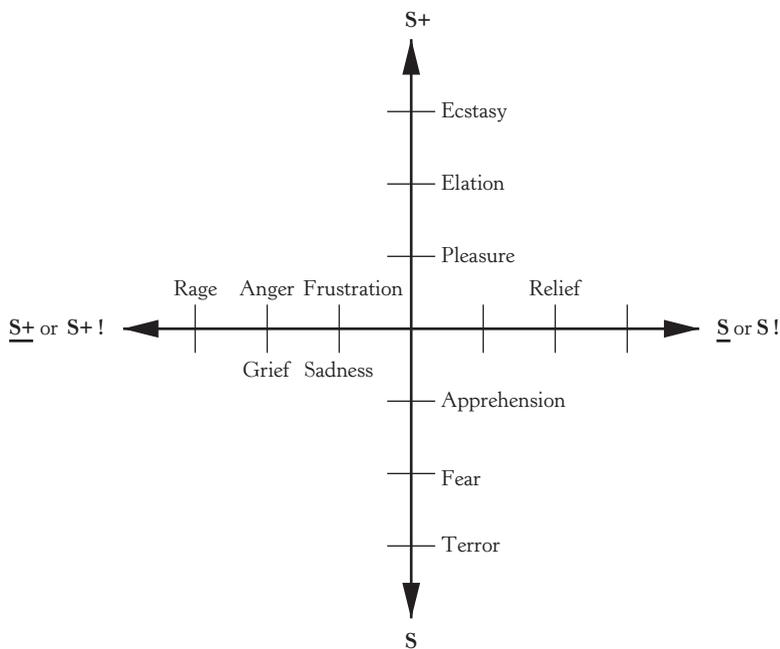


FIGURE Q1.6.1 Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the center of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists of (1) the presentation of a positive reinforcer (S+), (2) the presentation of a negative reinforcer (S-), (3) the omission of a positive reinforcer (S+) or the termination of a positive reinforcer (S+!), and (4) the omission of a negative reinforcer (S-) or the termination of a negative reinforcer (S-!). It should be understood that each different reinforcer will produce different emotional states: this diagram just summarizes the types of emotion that may be elicited by different contingencies, but the actual emotions will be different for each reinforcer.

of emotion. When rewards and punishers are present, there is also typically the opportunity for classical conditioning; for example, increased heart rate in response to a conditioned stimulus associated with the delivery of an aversive stimulus, or salivation upon the sight of food. Such classically conditioned responses may be useful in preparing the body for action, such as performing an action to avoid the aversive stimulus, or for eating, but these classically conditioned responses function mainly in the preparation for instrumental actions. For these reasons, and the evolutionary advantages of instrumental actions performed to obtain gene-specified goals, my theory focusses on stimuli that can lead to instrumental, goal-oriented actions, which for these reasons are central to my definition of emotion and its functions (Rolls, 2014a).

This foundation has been developed (Rolls, 2014a) to show how a very wide range of emotions can be accounted for, as a result of the operation of a number of factors, including the following:

1. The *reinforcement contingency* (e.g., whether reward or punishment is given, or withheld) (see Figure Q1.6.1).
2. The *intensity* of the reinforcer (see Figure Q1.6.1).
3. Any environmental stimulus might have a *number of different reinforcement associations*. (For example, a stimulus might be associated both with the presentation of a reward and of a punisher, allowing states such as conflict and guilt to arise.)
4. Emotions elicited by stimuli associated with *different primary reinforcers* will be different.
5. Emotions elicited by *different secondary reinforcing stimuli* will be different from each other (even if the primary reinforcer is similar).
6. The emotion elicited can depend on whether an *active or passive behavioral response* is possible. (For example, if an active behavioral response to the omission of a positive reinforcer can occur, then anger might be produced, but if only passive behavior is possible; then sadness, depression or grief might occur.)

By combining these six factors, it is possible to account for a very wide range of emotions (Rolls, 2014a). It is also worth noting that emotions can be produced just as much by the recall of reinforcing events as by external reinforcing stimuli; that

cognitive processing (whether conscious or not) is important in many emotions, for very complex cognitive processing may be required to determine whether or not environmental events are reinforcing. Indeed, emotions normally consist of cognitive processing that analyses the stimulus, and then determines its reinforcing valence; and then an elicited mood change if the valence is positive or negative. I note that a mood or affective state may occur in the absence of an external stimulus, as in some types of depression, but that normally the mood or affective state is produced by an external stimulus, with the whole process of stimulus representation, evaluation in terms of reward or punishment, and the resulting mood or affect being referred to as “emotion” (Rolls, 2014a).

THE FUNCTIONS OF EMOTION

The functions of emotion also provide insight into the nature of emotion. These functions, described more fully elsewhere (Rolls, 2014a), can be summarized as follows:

1. The *elicitation of autonomic responses* (e.g., a change in heart rate) and *endocrine responses* (e.g., the release of adrenaline): These prepare the body for action, and are responses (not instrumental actions) produced by stimuli that produce emotions, and can be classically conditioned.
2. *Flexibility of behavioral responses to reinforcing stimuli*: Emotional (and motivational) states allow a simple interface between sensory inputs and action systems. The essence of this idea is that goals for behavior are specified by reward and punishment evaluation. When an environmental stimulus has been decoded as a primary reward or punishment, or (after previous stimulus–reinforcer association learning) a secondary rewarding or punishing stimulus, then it becomes a goal for action. The human can then perform any action to obtain the reward, or to avoid the punisher. (Instrumental learning typically allows any action to be learned, though some actions may be more easily learned than others—Lieberman, 2000; Pearce, 2008). Thus there is flexibility of action, and this is in contrast with stimulus–response, or habit, learning in which a particular response to a particular stimulus

is learned. The emotional route to action is flexible not only because any action can be performed to obtain the reward or avoid the punishment, but also because the human can learn in as little as one trial that a reward or punishment is associated with a particular stimulus, in what is termed “stimulus–reinforcer association learning.”

To summarize and formalize, two processes are involved in emotional behavior. The first is stimulus–reinforcer association learning; emotional states are produced as a result (Rolls, 2014a). This process is implemented in structures such as the orbitofrontal cortex and amygdala (Figures Q1.6.2 and Q1.6.3) (Grabenhorst & Rolls, 2011; Rolls, 2014a; Rolls & Grabenhorst, 2008). The second is instrumental learning of an action made to approach and obtain the reward or to avoid or escape from the punisher. This is action–outcome learning, and involves brain regions such as the cingulate cortex when the actions are being guided by the goals, and the striatum and rest of the basal ganglia when the behavior becomes automatic and habit-based; that is, uses stimulus–response connections (Figures Q1.6.2 and Q1.6.3) (Rolls, 2009, 2014a, 2014b; Rushworth et al., 2011). Emotion is an integral part of this, for it is the state elicited in the first stage, by stimuli that are decoded as rewards or punishers, and this state has the property of being motivating. The motivation is to obtain the reward or avoid the punisher (the goals for the action), and animals must be “built” to obtain certain rewards and avoid certain punishers. Indeed, primary or unlearned rewards and punishers are specified by genes that effectively specify the goals for action. This is the solution that natural selection has found for how genes can influence behavior to promote their fitness (as measured by reproductive success), and for how the brain could interface sensory systems with action systems, and is an important part of Rolls’s theory of emotion (2005b, 2014a).

Choosing among available rewards with their associated costs, and avoiding punishers with their associated costs, is a process that can take place both implicitly (unconsciously), and explicitly by using a language system to enable long-term plans to be made (Rolls, 2008b, 2014a). These many different brain systems, some involving implicit evaluation of rewards, and others explicit, verbal, conscious, evaluation of rewards and planned long-term goals, must all enter into the selector of behavior (see Figure Q1.6.2).

The implication is that operation by animals (including humans) using reward and punishment

systems tuned to dimensions of the environment that increase fitness provides a mode of operation that can work in organisms that evolve by natural selection. It is clearly a natural outcome of Darwinian evolution to operate using reward and punishment systems tuned to fitness-related dimensions of the environment, if arbitrary actions are to be made by the animals, rather than just preprogrammed movements such as tropisms, taxes, reflexes, and fixed action patterns. This view of brain design in terms of reward and punishment systems built by genes that gain their adaptive value by being tuned to a goal for action offers, I believe, a deep insight into how natural selection has shaped many brain systems, and is a fascinating outcome of Darwinian thought (Rolls, 2014a).

3. Emotion is *motivating*, as just described. For example, fear learned by stimulus–reinforcement association provides the motivation for actions performed to avoid noxious stimuli.
4. *Communication*: Monkeys for example, may communicate their emotional state to others, such as by making facial expressions (such as an open-mouthed threat display to indicate the extent to which they are willing to compete for resources), and this may influence the behavior of other animals. This aspect of emotion was emphasized by Darwin (1872) and has been studied more recently by Ekman (1993), and, in terms of the brain mechanisms.
5. *Social bonding*: Examples of this are the emotions associated with the attachment of the parents to their young, and the attachment of the young to their parents.
6. The current mood state can affect the *cognitive evaluation of events or memories* (Blaney, 1986; Keltner et al., 2013). For example, happy memories are more likely to be recalled when one is happy. Another example is that when people are in a depressed mood, they tend to recall memories that were stored when they were depressed. The recall of depressing memories when depressed can have the effect of perpetuating the depression, and this may be a factor with relevance to the etiology and treatment of depression. The interactions between mood and memory systems using neural networks that capture the effects of interconnectivity from

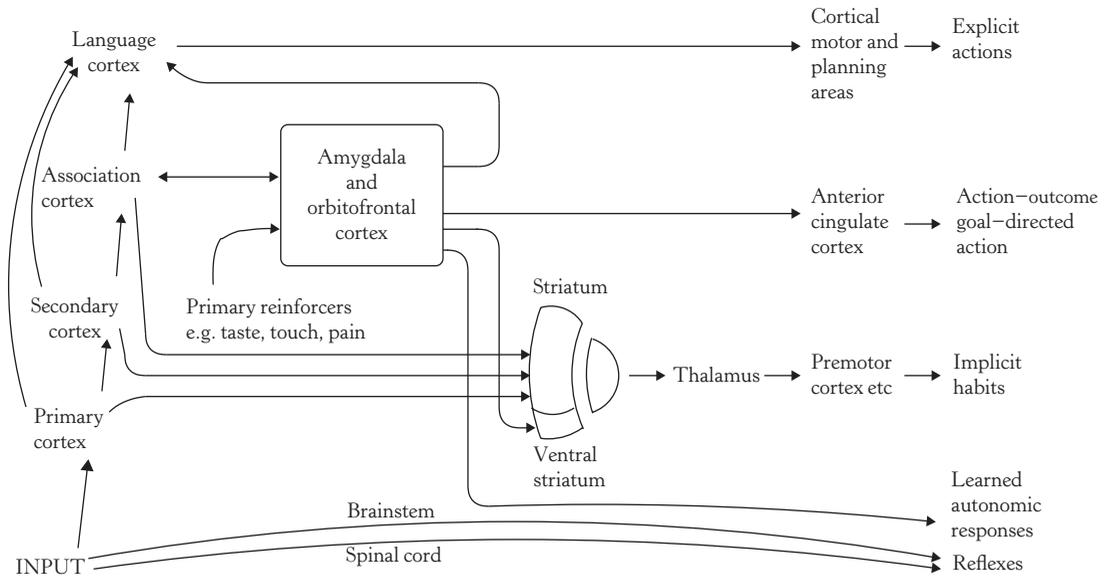


FIGURE Q1.6.2 *Dual routes to the initiation of action in response to rewarding and punishing stimuli.* The inputs from different sensory systems to brain structures such as the orbitofrontal cortex and amygdala allow these brain structures to evaluate the reward- or punishment-related value of incoming stimuli, or of remembered stimuli. The different sensory inputs enable evaluations within the orbitofrontal cortex and amygdala based mainly on the primary (unlearned) reinforcement value for taste, touch and olfactory stimuli, and on the secondary (learned) reinforcement value for visual and auditory stimuli. In the case of vision, the “association cortex” that outputs representations of objects to the amygdala and orbitofrontal cortex is the inferior temporal visual cortex. One route for the outputs from these evaluative brain structures is via projections directly to structures such as the basal ganglia (including the striatum and ventral striatum) to enable implicit, direct behavioral responses based on the reward- or punishment-related evaluation of the stimuli to be made. The second route is via the language systems of the brain, which allow explicit decisions involving multistep syntactic planning to be implemented.

emotional to perceptual and cognitive systems have been analyzed (Rolls, 2008b; Rolls & Stringer, 2001).

7. Emotion may facilitate the *storage of memories*. One way this occurs is that episodic memory (i.e., one’s memory of particular episodes) is facilitated by emotional states (Kesner and Rolls, 2015; Rolls, 2008b). A second way in which emotion may affect the storage of memories is that the current emotional state may be stored with episodic memories, providing a mechanism for the current emotional state to affect which memories are recalled. A third way that emotion may affect the storage of memories is by guiding the cerebral cortex in the representations of the world that are set up, using backprojections (Rolls, 2008b, 2016).
8. Another function of emotion is that by enduring for minutes or longer after a reinforcing stimulus has occurred, it may

help to produce *persistent and continuing motivation and direction of behavior*, to help achieve a goal or goals.

9. Emotion may trigger the *recall of memories* stored in neocortical representations. Amygdala backprojections to the cortex could perform this for emotion in a way analogous to that in which the hippocampus could implement the retrieval in the neocortex of recent (episodic) memories (Rolls, 2008b; Rolls & Stringer, 2001).

DIFFERENT SYSTEMS FOR EMOTIONAL LEARNING AND MEMORY

When stimuli are paired with primary reinforcers, associations that perform many types of functions are formed. Some are as follows, and are described in more detail in *Emotion and Decision-Making Explained* (Rolls, 2014a). The importance of this is that many processes take place during emotion,

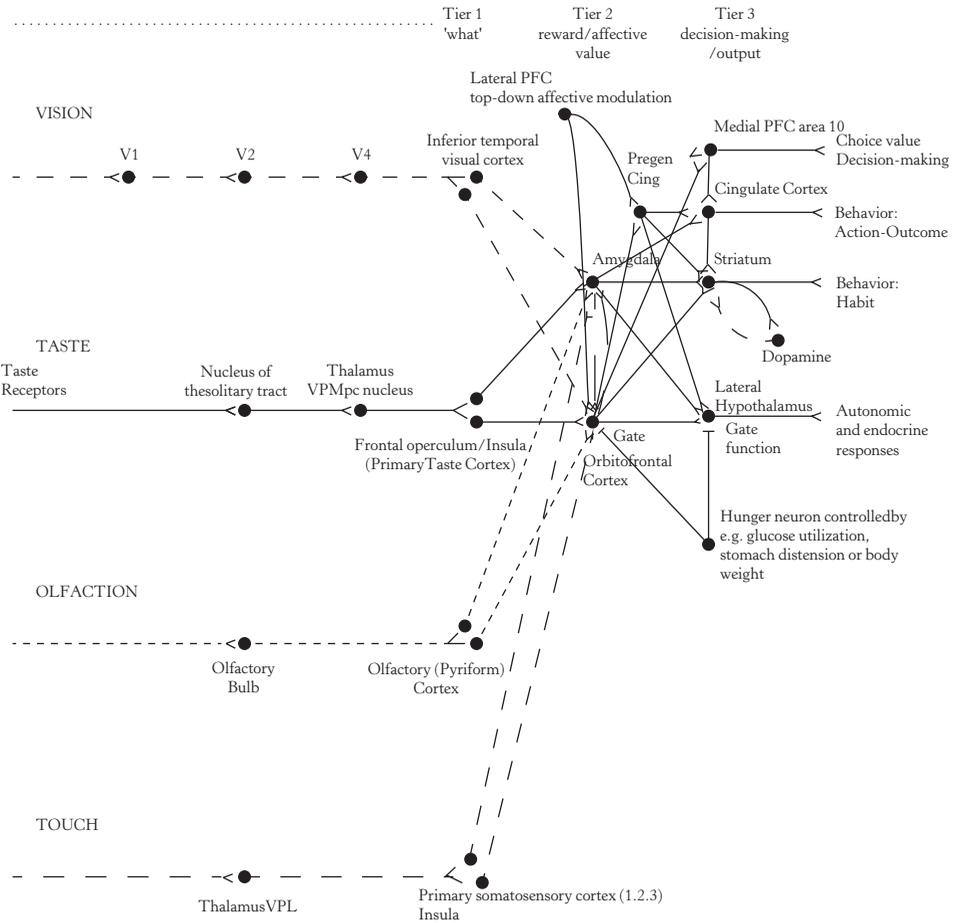


FIGURE Q1.6.3 Schematic diagram showing some of the gustatory, olfactory, visual, and somatosensory pathways to the orbitofrontal cortex, and some of the outputs of the orbitofrontal cortex, in primates. The secondary taste cortex, and the secondary olfactory cortex, are within the orbitofrontal cortex. V1: primary visual cortex. V4: visual cortical area V4. PreGen Cing: pregenual cingulate cortex. “Gate” refers to the finding that inputs such as the taste, smell, and sight of food in some brain regions only produce effects when hunger is present (Rolls, 2005b). The column of brain regions including and below the inferior temporal visual cortex represents brain regions in which what stimulus is present is made explicit in the neuronal representation, but not its reward or affective value, which are represented in the next tier of brain regions, the orbitofrontal cortex and amygdala, and in the anterior cingulate cortex. In areas beyond these such as medial prefrontal cortex area 10, choices or decisions about reward value are taken, with the mechanisms described elsewhere (Rolls, 2008b, 2014a; Rolls & Deco, 2010). Medial PFC area 10: medial prefrontal cortex area 10; VPMpc: ventral posteromedial thalamic nucleus.

and they can all contribute to the richness and sometimes the inconsistency of what happens during emotional behavior. Understanding the diversity of these processes provides a foundation for analyses and descriptions of emotional behavior.

First, as shown in Figure Q1.6.4, *Pavlovian (classical) conditioning* (in which a stimulus is paired with another stimulus or response, and where the actions have no influence on the pairing) has the potential to create multiple associative representations in the brain, as described next (Cardinal et al., 2002; Rolls, 2014a).

1. *Stimulus–Response association*: The conditioned stimulus (CS) may become directly associated with the unconditioned response (UR), a simple stimulus–response association that carries no information about the identity of the unconditioned stimulus (US; pathway 1 in Figure Q1.6.4). Such US-elicited responses include preparatory responses that are not specific to the type of US involved (e.g. orienting to a stimulus, or increased arousal), and “consummatory” responses that are specific to the US, such as

salivation to food, or blinking to an air puff applied to the eye, or approach to a food.

A single US may elicit both preparatory and consummatory responses, and thus the CS may enter into simple S-R associations with several types of response.

2. *A representation of affect, i.e. an emotional state:* The CS can evoke a representation of affect; i.e. an emotional state, such as fear or the expectation of reward (pathway 2 in Figure Q1.6.4). It is demonstrated operationally by the phenomenon of transreinforcer blocking (Cardinal et al., 2002). However, I note that, at least in humans, affective states normally have content; that is, they are about particular reinforcers (such as feeling happy because I am seeing a friend, or feeling happy because I am receiving a gift), and these states are better described by the third type of association, listed next.
3. *Conditioned-stimulus (CS)–unconditioned stimulus (US) associations.* The CS can become associated with the specific sensory properties of the US including its visual appearance, sound, and smell and its “consummatory” (primary reinforcing) properties such as its taste, nutritive value, and feel (pathway 3 in Figure Q1.6.4). This is the process involved in stimulus-reinforcer association learning, and in the brain involves structures such as the orbitofrontal cortex and amygdala (Grabenhorst and Rolls, 2011; Rolls, 2014a; Rolls and Grabenhorst, 2008).

Different pathways in the brain are involved in the Pavlovian learned autonomic and skeletal responses to a CS, and in the affective representation or state (e.g. fear), which may itself enter into associations and influence choice (Rolls, 2014a).

Second, in instrumental learning, there is a contingency between the behavior and the reinforcing outcome. A number of different learning processes may operate during this procedure, which it turns out may have somewhat different brain implementations (Cardinal et al., 2002; Rolls, 2014a). One key process is action-outcome learning. The outcome is represented as reward or affective value, such as that implemented by the firing of orbitofrontal cortex neurons that respond to the taste of food only if hunger is present. Other processes influence instrumental learning, including Pavlovian processes that can facilitate performance (as in Pavlovian-instrumental transfer).

Furthermore, approach to a food may be under Pavlovian rather than instrumental control.

We must be aware of the fact that after overtraining, responses may become inflexibly linked to stimuli, and that the goals, and the reward value of the goals, may no longer be directly influencing behavior in an ongoing way. My theory is that normally we want because we like. Indeed, that is inherent in my theory, for the genes that make a stimulus (such as a sweet taste) rewarding (i.e. wanted, a goal for action) also make the stimulus liked (i.e. accepted, with a subjective correlate of pleasure, pleasantness, and affective liking). If behavior becomes overlearned and a habit or stimulus–response connection is built up by another brain system, then animals may make automatic responses that are not goal-directed. There has been confusion in the literature caused by overlooking this point (Berridge et al., 2009). The fact that behavior can become stimulus–response and no longer be under the control of the goal need not surprise us. Normally, and certainly during learning before habits set in, we want a goal, and when we get the goal, we like it: goal stimuli normally specify what is wanted and what is liked (Rolls, 2014a).

The impact of this analysis in the present context of approach and avoidance is as follows. There are many different brain processes involved in learning emotional responses; different brain regions are involved in the different types of learning, and what is learned in each system may be somewhat independent of what is learned in the other systems (Rolls, 2014a). We should thus not assume that emotion is a single unified process. There may be many different underlying processes that take place, and they are not always consistent with each other. My hypothesis is that it is the states elicited by instrumental reinforcers that are emotional states. Stimuli that produce such states may have other effects, eliciting perhaps autonomic responses, approach, fixed action patterns, and learned habits, and while these responses are adaptive, they do not require the intervening states involved in instrumental actions made to obtain goals, and are therefore not crucial to emotional states. The emotional states produced by such instrumental reinforcers are states of the brain, the firing of neurons, that action systems in the brain seek to obtain or avoid by producing instrumental actions. The emotional states include representations of goals (rewards and punishers) and are independent of the actions made to obtain the goals. These emotional states may or may not be conscious: my approach to consciousness suggests that emotional states may gain access to consciousness especially when we must perform reasoning that involves these states, and

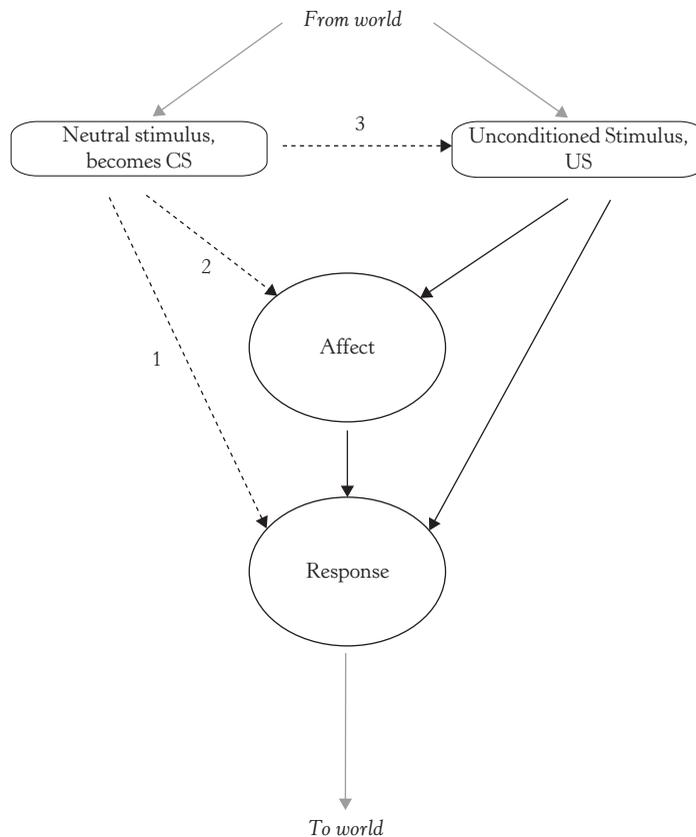


FIGURE Q1.6.4 Pavlovian conditioning has the potential to create associations between a conditioned stimulus (CS) and representations of the unconditioned stimulus (US), central affective or emotional states such as fear, and unconditioned responses. Dashed lines represent associatively learned links. Several different types of response may be involved, including preparatory responses that are not specific to the type of US involved (e.g., orienting to a stimulus, or increased arousal), and “consummatory” responses that are specific to the US such as salivation to food, or blinking to an air puff applied to the eye (after Cardinal, Parkinson, Hall, & Everitt, 2002).

correcting errors in such reasoning (Rolls, 2007a, 2007b, 2008a, 2014a).

A SEPARATE, RATIONAL, REASONING, CONSCIOUS SYSTEM FOR IDENTIFYING EMOTIONAL GOALS

I have put forward a position elsewhere that in addition to the gene-based goal system for emotion described here, there is a separate rational—that is, reasoning—system that can plan ahead and work for what are sometimes different, long-term, goals (Rolls, 1997b, 2003, 2004, 2005a, 2005b, 2007a, 2007b, 2008a, 2011a, 2012, 2014a). This type of processing involves multi-step trains of thought, such as might be required to formulate a plan with many steps. Each step has its own

symbols (e.g., a word to represent a person), and so syntactic linking (binding) is needed between the symbols within each step, and some syntactic (relational) links must be made between symbols in different steps. I have argued that when we correct such multi-step plans or trains of thought, we need to think about these first-order thoughts, and the system that does this is thus a higher-order thought system (in that it is thinking about first-order thoughts).

There is a fundamentally important distinction here: working for a gene-specified reward, as in many emotions, is performed for the interests of the “selfish” genes. Working for rationally planned rewards may be performed in the interest of the particular individual (e.g. the person), and not in the interests of the genotype (Rolls, 2011a, 2014a).

It is suggested that this arbitrary symbol-manipulation using important aspects of language processing and used for planning but not in initiating all types of behavior is close to what “consciousness” is about. In particular, consciousness may *be* the state that arises in a system that can think about (or reflect on) its own (or other people’s) thoughts; that is, in a system capable of second or higher-order thoughts (Carruthers, 1996; Dennett, 1991; Gennaro, 2004; Rolls, 1995, 1997a, 1997b, 1999, 2004, 2007a, 2014a; Rosenthal, 1986, 1990, 1993, 2004, 2005).

It is of great interest to comment on how the evolution of a system for flexible planning might affect emotions. Consider grief, which may occur when a reward is terminated and no immediate action is possible (see Rolls, 1990, 2014a). It may be adaptive by leading to a cessation of the formerly rewarded behavior and thus facilitating the possible identification of other positive reinforcers in the environment. In humans, grief may be particularly and especially potent because it is represented in a system that can plan ahead, and understand the enduring implications of the loss.

The question then arises of how decisions are made in animals such as humans that have both the implicit, direct reward-based, and the explicit, rational, planning systems (see Figure Q1.6.2) (Rolls, 2008b). One particular situation in which the first, implicit, system may be especially important is when rapid reactions to stimuli with reward or punishment value must be made, for then structures such as the orbitofrontal cortex may be especially important (Rolls, 2014a). Another is when there may be too many factors to be taken into account easily by the explicit, rational, planning system, when the implicit system may be used to guide action. In contrast, when the implicit system continually makes errors, it would then be beneficial for the organism to switch from automatic habit, or from action-outcome, goal-directed behavior, to the explicit conscious control system that can evaluate with its long-term planning “algorithms” what action should be performed next. Indeed, it would be adaptive for the explicit system to regularly be assessing performance by the more automatic system, and to “switch itself on” to control behavior quite frequently, as otherwise, the adaptive value of having the explicit system would be less than optimal.

It may be expected that there is often a conflict between these systems, in that the first, implicit, system is able to guide behavior particularly to obtain the greatest immediate reinforcement,

whereas the explicit system can potentially enable immediate rewards to be deferred, and longer-term, multi-step, plans to be formed that may be in the interests of the individual, not the genes. For example, an individual might decide not to have children, but instead to devote himself or herself to being a creative individual, or to enjoying opera, and so forth. This type of conflict will occur in animals with a syntactic planning ability; that is, in humans and any other animals that have the ability to process a series of “if . . . then” stages of planning. This is a property of the human language system, and the extent to which it may be a property of non-human primates is not yet fully clear. In any case, such conflict may be an important aspect of the operation of at least the human mind, because it is so essential for humans to correctly decide, at every moment, whether to invest in a relationship or a group that may offer long-term benefits, or whether to directly pursue immediate benefits (Rolls, 2008b, 2011a, 2014a).

DECISION-MAKING MECHANISMS IN THE BRAIN, AND HOW THEY ARE INFLUENCED BY “NOISE” IN THE BRAIN

Recently, a theoretical foundation for understanding decision-making in the brain has been emerging (Deco et al., 2013; Deco et al., 2009; Rolls, 2008b, 2014a; Rolls & Deco, 2010; Wang, 2002). A fundamental part of the brain’s architecture is a neural network that has positive internal feedback between its neurons, and that can fall into one of a number of states, each one of which corresponds to a decision, and consists of one winning population of neurons that is firing at a high rate, and inhibits the other populations. When the decision process starts, if the inputs are relatively equal, the state that is reached is influenced by the “noisy”—that is, random—spike timings of the firings of the neurons in the different populations. This type of noise in decision-making processes may operate at many different stages of brain processing, and may even influence the way in which decisions are influenced on different occasions between the unconscious emotional system, and rational decision-making processes (Rolls, 2004, 2005a, 2007a, 2007b, 2008a, 2008b, 2014a; Rolls & Deco, 2010). I emphasize that by “rational,” I mean here “reasoned.” In this way, noise in the brain may influence what behavioral responses or actions are made to emotional stimuli, including, for example,

whether actions are based on activity in the emotional or reasoning brain systems.

CONCLUSIONS

In this section, I have outlined a theory of emotion. This provides an account of approach and avoidance in terms of, first, associative learning to define which stimuli are goals (this involves associative learning between stimuli that are primary and secondary reinforcers), and second, of instrumental or action-outcome (i.e., action-reinforcer) association learning. Evidence on the brain mechanisms involved in emotion is described in more detail elsewhere (Rolls, 2014a, 2014b).

1.7 ACTIVE INFERENCE AND EMOTION

Karl J. Friston, Mateus Joffily,
Lisa Feldman Barrett, and Anil K. Seth

OVERVIEW

In what follows, we briefly survey the theoretical (active inference) framework, under which we consider the nature of emotional constructs. We then consider three dominant themes within this framework; namely, (i) emotion as changes in free energy (Joffily & Coricelli, 2013); (ii) interoceptive inference and counterfactual processing (Seth, 2013); and (iii) Conceptual Act Theory (Barrett, 2014). Having established the theoretical fundamentals, we then revisit the basic attributes of emotions and identify their representational and neurophysiological correlates—showing that these themes converge on one conclusion. This conclusion is that emotional processing optimizes the expected precision of (or confidence in) beliefs about the affiliative and interoceptive consequences of behavior and autonomic control. Finally, the evidence for this conclusion is reviewed briefly, in terms of theoretical neurobiology and neuropsychology.

THE FREE ENERGY PRINCIPLE AND ACTIVE INFERENCE

Active inference is a corollary of the free energy principle, which says that changes in brain states (e.g., synaptic activity and efficacy) minimize a quantity called *variational free energy* (Friston, 2010). This quantity reports the surprise about observed outcomes based on the predictions of an internal or generative model (of sensorimotor contingencies). A popular process theory for this minimization is

predictive coding—in which the brain constructs top-down predictions of sensory inputs in a hierarchical fashion and uses the ensuing prediction errors to update its predictions (Mumford, 1992; Rao & Ballard, 1999). There is a considerable amount of anatomical and physiological evidence that is consistent with predictive coding (e.g., de Gardelle, Waszczuk, et al., 2013); including the hierarchical organization of extrinsic (between-area) neuronal connections and asymmetries between forward and backward connections (Bastos, Usrey, et al., 2012).

Predictive coding is one example of *message passing* or *belief propagation* that can be seen as a particular instance of Bayesian inference. In other words, the reciprocal exchange of top-down predictions and bottom-up prediction errors corresponds to inferring the likeliest causes of our sensations. Irrespective of the message-passing scheme used by the brain, the implication is that synaptic activity and connectivity encode probabilistic beliefs about (hidden) causes of the sensorium. These beliefs are used to generate predictions, which are compared against sensory evidence to enable Bayesian belief-updating (e.g., by prediction errors).

So what is the nature of these beliefs? In this setting, “beliefs” are probability distributions over fictive constructs that can explain sensory samples (e.g., my visual impressions are currently caused by a smiling face). In turn, probability distributions are encoded by their sufficient statistics—like their mean (or expectation) and variance (or inverse precision). This means neuronal activity (and connectivity) encodes the sufficient statistics of beliefs in terms of expectations and precisions. The distinction between expectations and precisions is important: expectations encode the contents of a belief (e.g., a smiling face), while precision encodes the confidence in that expectation. Crucially, both have to be encoded—and both have to be minimized with respect to free energy or surprise. This is like estimating the difference in group means and the standard error of that difference in classical statistics, based on the same data.

Neurobiologically, it is thought that expectations about the state of the world are encoded by synaptic activity in hierarchical sensory and motor systems, while precision is encoded by the efficacy of message-passing established by neuromodulatory synaptic gain (Brown & Friston, 2012; Clark, 2013). This immediately implicates modulatory neurotransmission (and synchronous gain) in the encoding of precision or confidence of our expectations. This will become important in our treatment of emotions.

other region of the brain is activated in all of these tasks” (p. 65). Indeed, in a recent analysis of the *functional diversity* of brain regions, the anterior insula emerged as one of the most diverse (Anderson, Kinnison, & Pessoa, 2013).

Thus, the anterior insula is robustly engaged during both cognitive and emotional contexts and provides an interface between the two, as exemplified by the findings of a response-conflict task by my research group (Choi, Padmala, & Pessoa, 2012). Participants were asked to indicate whether a picture was a house or building, while ignoring task-irrelevant words. Interference-related responses were observed in the anterior insula; that is, larger responses when an incongruent stimulus was encountered. At the same time, responses were observed in threat trials (without actual shock delivery), during which presumably participants monitored for the occurrence of shock. In addition to being driven by the cognitive and emotional dimensions of the task, an interaction of the two was also observed in the anterior insula; namely, larger conflict responses occurred when threat was encountered. Furthermore, this pattern of responses increased based on individual differences in trait anxiety—that is, high-anxious individuals exhibited increased interactions. Combined, this three-way statistical interaction (interference x emotion x trait anxiety) nicely illustrates the confluence of different signals in the anterior insula (see also Gu et al., 2012).

In summary, the medial PFC and the anterior insula are located at the intersection between emotion and executive function. The importance of these regions in cognition–emotion interactions is captured by their proposed status as *hub* regions that play a prominent role in the *integration* of information by mixing signals that have distinct compositions (Figure Q8.7.1B). Consider, for instance, the representation of the entire body state that is observed in the anterior insula. This representation involves signals from multiple parts of the body, including visceral signals, in addition to temperature, pain, itch, muscular sensations, sensual touch, and other feelings from the body (Craig, 2002). At the same time, the anterior insula is consistently engaged during executive function. Accordingly, executive and interoceptive signals are contained in the anterior insula. A related situation occurs in the dorsomedial PFC. Of note, the anterior part of the cingulate gyrus probably has a more extensive descending projection system than any other cortical region (Vogt & Vogt, 2009), with robust projections to autonomic regulatory structures. This suggests that affective signals that mobilize the body are mixed with executive signals in the medial PFC. Taken together, the

anterior insula and medial PFC provide the substrate for ample cognitive–emotional integration that, in broad terms, includes both bodily “input” and “output” signals. In addition, these regions do not work in isolation. During cognitive–emotional interactions, they interact with the lateral PFC and the parietal cortex (Figure Q8.7.1B).

A third type of communication depends on the diffuse action of neuromodulatory signals, including the action of dopamine and norepinephrine. Widespread modulatory connections originating from these systems reach large portions of the cortical surface and several subcortical areas; thereby they are able to rapidly influence brain responses. Neuromodulatory effects have been documented in animal models of stress (Arnsten, 2009; Panksepp, 1998). For example, exposure to acute stress leads to release of norepinephrine across a widely distributed brain network, including the PFC (Arnsten & Li, 2005; Hermans et al., 2011; Joels et al., 2006). Acute stress also results in activation of the dopamine system, which has considerable effects on PFC function and associated tasks, particularly working memory (Arnsten, 2009; Deutch & Roth, 1990).

CONCLUSIONS

The *dual competition* model describes how affective significance influences competition at *both* the perceptual and the executive levels. To focus the discussion, the interactions were described mostly *from* emotion *to* perception/cognition. But interactions are bidirectional, of course. A prototypical example of influences *from* cognition *to* emotion is provided by emotion regulation (Ochsner & Gross, 2005). In the end, as described in my answers to Questions 4 and 5 in this volume, emotion and cognition interact so strongly that a demarcation between them turns out to be a fruitless enterprise. In the end, we must speak of an *emotion-cognition amalgam*. To study the constituent elements is productive only insofar as it further reveals the properties of the new compound element in all its complexity and richness.

8.8 EMOTIONAL VS. RATIONAL SYSTEMS, AND DECISIONS BETWEEN THEM

Edmund T. Rolls

Rolls’s theory of emotions is that emotions are states elicited by gene-specified instrumental reinforcers (or stimuli associated with them) that

are the goals for action (Rolls, 1999, 2005b, 2014, 2017b, 2018). Emotions provide an efficient way for genes to influence behavior in their own self-interest, by specifying the goals for action instead of particular responses to particular stimuli.

These emotional states may or may not be conscious: my approach to consciousness suggests that emotional states may gain access to consciousness, especially when we must perform reasoning that involves these states, and correcting errors in such reasoning (Rolls, 2007a, 2007b, 2008a, 2014).

A SEPARATE, RATIONAL, REASONING, CONSCIOUS SYSTEM FOR IDENTIFYING EMOTIONAL GOALS

I have put forward a position elsewhere that in addition to the gene-based goal system for emotion, there is a separate rational—that is, reasoning—system that can plan ahead and work for what are sometimes different, long-term, goals (Rolls, 1997b, 2003, 2004, 2005a, 2005b, 2007a, 2007b, 2008a, 2011, 2012, 2014). This type of processing involves multistep trains of thought, as might be required to formulate a plan with many steps. Each step has its own symbols (e.g., a word to represent a person), so syntactic linking (binding) is needed between the symbols within each step, and some syntactic (relational) links must be made between symbols in different steps. I have argued that when we correct such multistep plans or trains of thought, we need to think about these first-order thoughts, and the system that does this is thus a higher-order thought system (in that it is thinking about first-order thoughts).

There is a fundamentally important distinction here: working for a gene-specified reward, as in many emotions, is performed for the interests of the “selfish” genes. Working for rationally planned rewards may be performed in the interest of the particular individual (e.g., the person, the phenotype), and not in the interests of the genotype (Rolls, 2011, 2014).

It is suggested that this arbitrary symbol manipulation using important aspects of language processing, and used for planning but not in initiating all types of behavior, is close to what consciousness is about. In particular, consciousness may *be* the state that arises in a system that can think about (or reflect on) its own (or other people’s) thoughts; that is, in a system capable of second- or higher-order thoughts (Carruthers, 1996; Dennett, 1991; Gennaro, 2004; Rolls, 1995,

1997a, 1997b, 1999, 2004, 2007a, 2014; Rosenthal, 1986, 1990, 1993, 2004; Rosenthal, 2005).

It is of great interest to comment on how the evolution of a system for flexible planning might affect emotions. Consider grief, which may occur when a reward is terminated and no immediate action is possible (see Rolls, 1990, 2014). It may be adaptive by leading to a cessation of the formerly rewarded behavior and thus facilitating the possible identification of other positive reinforcers in the environment. In humans, grief may be particularly and especially potent because it becomes represented in a system that can plan ahead and understand the enduring implications of the loss (Cheng et al., 2016; Rolls, 2016b, 2017a).

The question then arises of how decisions are made in animals such as humans that have both the implicit, direct reward-based, and the explicit, rational, planning systems (see Figure Q8.8.1) (Rolls, 2008b, 2016a). One particular situation in which the first, implicit, system may be especially important is when rapid reactions to stimuli with reward or punishment value must be made, for then structures such as the orbitofrontal cortex may be especially important (Rolls, 2014). Another is when there may be too many factors to be taken into account easily by the explicit, rational, planning system, when the implicit system may be used to guide action. In contrast, when the implicit system continually makes errors, it would then be beneficial for the organism to switch from automatic habit, or from action–outcome goal-directed behavior, to the explicit conscious control system, which can evaluate with its long-term planning algorithms what action should be performed next. Indeed, it would be adaptive for the explicit system to regularly be assessing performance by the more automatic system, and to “switch itself in” to control behavior quite frequently, as otherwise the adaptive value of having the explicit system would be less than optimal.

It may be expected that there is often a conflict between these systems, in that the first, implicit, system is able to guide behavior particularly to obtain the greatest immediate reinforcement, whereas the explicit system can potentially enable immediate rewards to be deferred, and longer-term, multi-step plans to be formed that may be in the interests of the individual, not the genes. For example, an individual might decide not to have children, but instead to devote himself or herself to being a creative individual, or to enjoying opera, and so forth. This type of conflict will occur in animals with a syntactic planning ability; that is, in humans and any other animals that have the ability to process a series of “if . . . then”

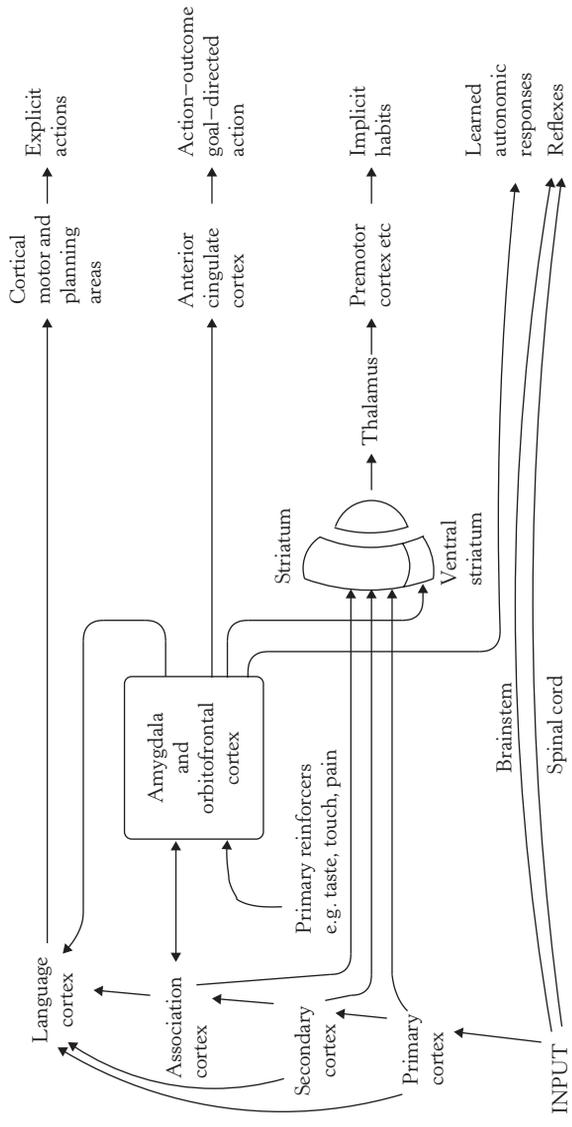


FIGURE Q8.8.1 *Dual routes to the initiation of action in response to rewarding and punishing stimuli.* The inputs from different sensory systems to brain structures such as the orbitofrontal cortex and amygdala allow these brain structures to evaluate the reward- or punishment-related value of incoming stimuli, or of remembered stimuli. The different sensory inputs enable evaluations within the orbitofrontal cortex and amygdala based mainly on the primary (unlearned) reinforcement value for taste, touch, and olfactory stimuli, and on the secondary (learned) reinforcement value for visual and auditory stimuli. In the case of vision, the “association cortex” that outputs representations of objects to the amygdala and orbitofrontal cortex is the inferior temporal visual cortex. One route for the outputs from these evaluative brain structures is via projections directly to structures such as the basal ganglia (including the striatum and ventral striatum) to enable implicit, direct behavioral responses based on the reward- or punishment-related evaluation of the stimuli to be made. The second route is via the language systems of the brain, which allow explicit decisions involving multistep syntactic planning to be implemented.

stages of planning. This is a property of the human language system, and the extent to which it may be a property of non-human primates is not yet fully clear. In any case, such conflict may be an important aspect of the operation of at least the human mind, because it is so essential for humans to correctly decide, at every moment, whether to invest in a relationship or a group that may offer long-term benefits, or whether to directly pursue immediate benefits (Rolls, 2008b, 2011, 2014).

DECISION-MAKING MECHANISMS IN THE BRAIN, AND HOW THEY ARE INFLUENCED BY “NOISE” IN THE BRAIN

Recently, a theoretical foundation for understanding decision-making in the brain has been emerging (Deco et al., 2013; Deco et al., 2009; Rolls, 2008b, 2014, 2016a; Rolls & Deco, 2010; Wang, 2002). A fundamental part of the architecture is a neural network that has positive internal feedback between its neurons, and that can fall into one of a number of states, each one of which corresponds to a decision, and consists of one winning population of neurons that is firing at a high rate and inhibits the other populations. When the decision process starts, if the inputs are relatively equal, the state that is reached is influenced by the “noisy”—that is, random—spike timings of the firings of the neurons in the different populations. This type of noise in decision-making processes may occur at many different stages of brain processing, and may even influence the way in which decisions are influenced on different occasions between the unconscious emotional system and the rational decision-making processes (Rolls, 2004, 2005a, 2007a, 2007b, 2008a, 2008b, 2014, 2016a; Rolls & Deco, 2010). I emphasize that by “rational” I mean here “reasoned.” In this way, noise in the brain may influence what behavioral actions or responses are made to emotional stimuli, including, for example, whether actions are based on activity in the emotional or the reasoning brain systems.

8.9 AFTERWORD

How Do Emotion and Cognition Interact?

Alexander J. Shackman and
Regina C. Lapate

Emotion and cognition seem fundamentally different. Emotion is hot, bright, and

quick; infused with vivid feelings of pleasure or pain and manifesting in readily discerned changes in the body. In contrast, cognition is cold, gray, and slow; devoid of substantial hedonic, motivational, or somatomotor features. These differences in phenomenology and psychophysiology led classical thinkers and philosophers to treat emotion and cognition as distinct, often warring, mental faculties. And yet the two decades since the publication of the first edition of *The Nature of Emotion* have witnessed the emergence and widespread adoption of new tools for objectively assaying both the mind and the brain. What have these new data taught us about the interplay of emotion and cognition?

At the broadest level, Okon-Singer and colleagues remind us that emotional cues, states, moods, traits, and disorders can, and often do, influence key components of cognition, including attention, working memory, and cognitive control. Drawing on biased-competition models of cognition (Desimone & Duncan, 1995; Miller & Cohen, 2001), Pessoa suggests that affect can prejudice the competition for limited cognitive resources at virtually every level of the information-processing hierarchy, from perception to “executive” cognition. Berggren and Derakshan argue that emotionally salient stimuli enjoy privileged access to attention and memory. Clore tells us that mood and affect confer “positive or negative value on whatever is in mind at the time”; that mood and emotion represent a kind of information that can bias judgments, evaluations, and choices in valence-congruent ways (see also Nettle & Bateson, 2012). Cools et al. make a related point, highlighting evidence that emotionally salient stimuli (e.g., emotional faces, Pavlovian cues) can bias instrumental approach and avoidance in a valence-congruent manner (i.e., positive stimuli facilitate approach and inhibit avoidance, whereas negative stimuli exert the opposite effect).

EMOTION CAN INFLUENCE COGNITION

Emotion Hijacks Attention

Several contributors emphasized the consequences of emotion for selective attention. Clore focuses on mood and affect, suggesting that a key function of affect is to capture attention, and that “whatever seizes one’s attention then becomes input for other mental processes.” Okon-Singer et al. make a conceptually similar point: “Once lodged in working memory, threat-related information is poised to