Orbitofrontal cortex, decision-making and drug addiction

Geoffrey Schoenbaum[1,2], Matthew R. Roesch[1] and Thomas A. Stalnaker[1]

[1]Department of Anatomy and Neurobiology, University of Maryland School of Medicine, 20 Penn Street, HSF-2 S251, Baltimore, MD 21201, USA
[2]Department of Psychiatry, University of Maryland School of Medicine, 20 Penn Street, HSF-2 S251, Baltimore, MD 21201, USA

The orbitofrontal cortex, as a part of prefrontal cortex, is implicated in executive function. However, within this broad region, the orbitofrontal cortex is distinguished by its unique pattern of connections with crucial subcortical associative learning nodes, such as basolateral amygdala and nucleus accumbens. By virtue of these connections, the orbitofrontal cortex is uniquely positioned to use associative information to project into the future, and to use the value of perceived or expected outcomes to guide decisions. This review will discuss recent evidence that supports this proposal and will examine evidence that loss of this signal, as the result of drug-induced changes in these brain circuits, might account for the maladaptive decision-making that characterizes drug addiction.

Introduction

Our ability to form expectations about the desirability or value of impending events underlies much of our emotion and behavior. In fact, two broad functions are crucially subserved by the formation of such expectations. On the one hand, expectations guide our immediate behavior, allowing us to pursue goals and avoid potential harm. On the other hand, expectations can be compared with actual outcomes to facilitate learning so that future behavior can become more adaptive. Both of these functions require that information about expected outcomes be maintained in memory so that it can be compared and integrated with information about internal state and current goals. Such an integrative process generates a signal that we will refer to as an outcome expectancy, a term long-used by learning theorists to refer to an internal representation of the consequences likely to follow a specific act [1]. The disruption of such a signal would be expected to create a myriad of difficulties, in the ability both to make adaptive decisions and to learn from negative consequences of decisions. In this review, we first describe recent evidence that the orbitofrontal cortex (OFC) plays a crucial role in the generation and use of outcome expectancies. Subsequently, we will discuss recent evidence that the maladaptive decisions that characterize drug addiction reflect, in part, a disruption of this signal as a result of drug-induced changes in the OFC and related brain areas.

Neural activity in the OFC and OFC-dependent behavior reflect a crucial role of the OFC in the generation of outcome expectancies

The ability to maintain information so that it can be manipulated, integrated with other information and then used to guide behavior has been variously described as working, scratchpad or representational memory, and it depends crucially on the prefrontal cortex [2]. Within the prefrontal cortex, the OFC, by its connections with limbic areas, is uniquely positioned to enable associative information regarding outcomes or consequences to access representational memory (Box 1). Indeed a growing number of studies suggest that a neural correlate of the expected value of outcomes is present and perhaps generated in the OFC. For example, human neuroimaging studies show that blood flow changes in the OFC during anticipation of expected outcomes and also when the value of an expected outcome is modified or not delivered [3–6]. This activation appears to reflect the incentive value of these items and is observed when that information is being used to guide decisions [7]. These results suggest that neurons in the OFC increase activity when such information is processed. Accordingly, neural activity in the OFC that precedes predicted rewards or punishments increases, typically reflecting the incentive values of these outcomes [8–11]. For example, when monkeys are presented with visual cues paired with differently preferred rewards, neurons in the OFC fire selectively according to whether the anticipated outcome is the preferred or non-preferred reward within that trial block [10]. Moreover, Roesch and Olson [11] have recently demonstrated that firing in the OFC tracks several other specific metrics of outcome value. For example, neurons fire differently for a reward depending on its expected size, the anticipated time required to obtain it and the possible aversive consequences associated with inappropriate behavior [11,12].

Such anticipatory activity appears to be a common feature of firing activity in the OFC across many tasks in which events occur in a sequential, and thus predictable, order (Box 2). Importantly, however, these selective responses can be observed in the absence of any signaling cues, and they are acquired as animals learn that particular cues predict a specific outcome. In other words, this selective activity represents the expectation of an animal, based on experience, of likely outcomes.

Box 1. The anatomy of the orbitofrontal circuit in rats and primates

Rose and Woolsey [53] proposed that prefrontal cortex might be defined by the projections of the mediodorsal thalamus (MD) rather than by 'stratiographic analogy' [54]. This definition provides a foundation on which to define prefrontal homologs across species. However, it is the functional and anatomical similarities that truly define homologous areas (Figure I of this box).

In the rat, the MD can be divided into three segments [55,56]. Projections from the medial and central segments of the MD define a region that includes the orbital areas and the ventral and dorsal agranular insular cortices [55–58]. These regions of the MD in rat receive direct afferents from the amygdala, medial temporal lobe, ventral pallidum and ventral tegmental area, and they receive olfactory input from the piriform cortex [55,56,59]. This pattern of connectivity is similar to that of the medially located, magnocellular division of primate MD, which defines the orbital prefrontal subdivision in primates [60–62]. Thus, a defined region in the orbital area of rat prefrontal cortex is likely to receive input from thalamus that is very similar to that reaching primate orbital prefrontal cortex. Based, in part, on this pattern of input, the projection fields of medial and central MD in the orbital and agranular insular areas of rat prefrontal cortex have been proposed as homologous to the primate orbitofrontal region [55,57,63–65]. These areas in rodents include the dorsal and ventral agranular insular cortex, and the lateral and ventrolateral orbital regions. This conception of the rat orbitofrontal cortex (OFC) does not include the medial or ventromedial orbital cortex, which lie along the medial wall of the hemisphere. This region has patterns of connectivity with the MD and other areas that are more similar to other regions on the medial wall.

Other important connections highlight the similarity between the rat OFC and the primate OFC. Perhaps most notable are reciprocal connections with the basolateral complex of the amygdala (ABL), a region thought to be involved in affective or motivational aspects of learning [66–74]. In primate, these connections have been invoked to explain specific similarities in behavioral abnormalities resulting from damage to either the OFC or the ABL [14,17,75–77]. Reciprocal connections between basolateral amygdala and areas within rat OFC, particularly the agranular insular cortex [58,78–80], suggest that interactions between these structures might be similarly important for regulation of behavioral functions in rats. In addition, in both rats and primates, the OFC provides a strong efferent projection to the nucleus accumbens, overlapping with innervation from limbic structures such as the ABL and subiculum [81–84]. The specific circuitry connecting the OFC, limbic structures and nucleus accumbens presents a striking parallel across species that suggests possible similarities in functional interactions among these major components of the forebrain [81,84,85].
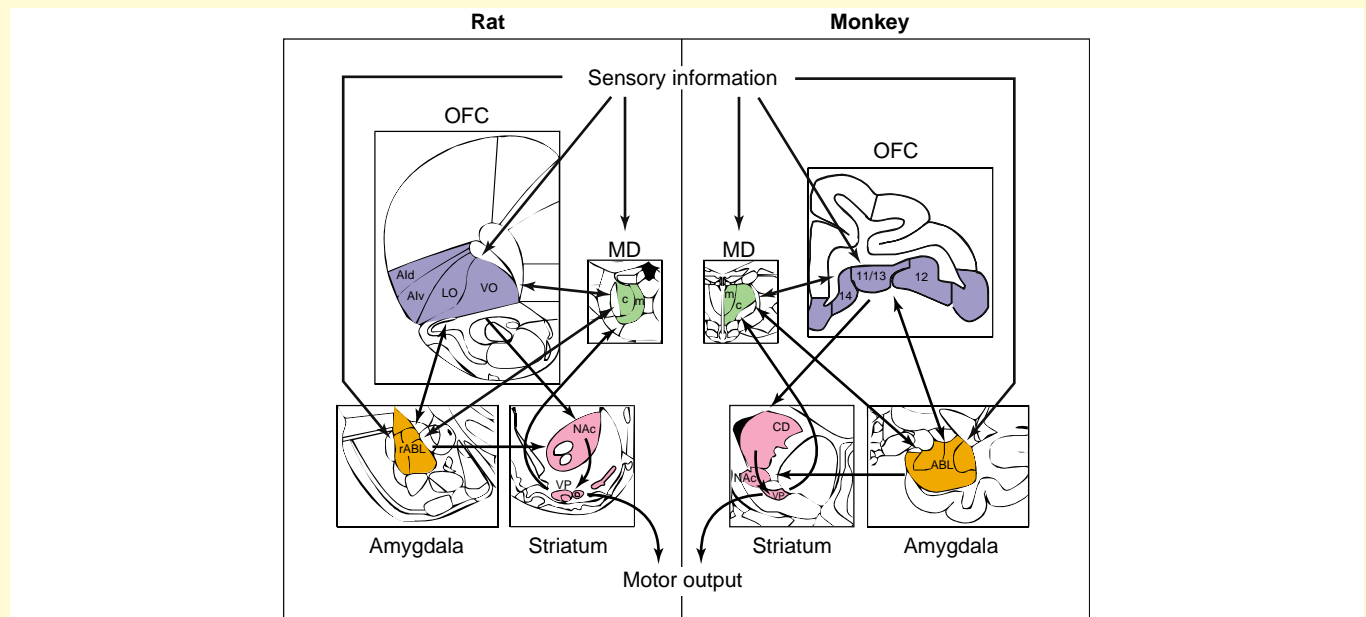

Figure I. Anatomical relationships of the OFC (blue) in rats and monkeys. Based on their pattern of connectivity with the mediodorsal thalamus (MD, green), amygdala (orange) and striatum (pink), the orbital and agranular insular areas in rat prefrontal cortex are homologous to the primate OFC. In both species, the OFC receives robust input from sensory cortices and associative information from the amygdala, and sends outputs to the motor system through the striatum. Each box illustrates a representative coronal section. Additional abbreviations: AId, dorsal agranular insula; AIv, ventral agranular insula; c, central; CD, caudate; LO, lateral orbital; m, medial; NAc, nucleus accumbens core; rABL, rostral basolateral amygdala; VO, ventral orbital, including ventrolateral and ventromedial orbital regions; VP, ventral pallidum.

These features are illustrated in Figure 1, which shows the population response of OFC neurons recorded in rats as they learn and reverse novel odor-discrimination problems [8,9,13]. In this simple task, the rat must learn that one odor predicts reward in a nearby fluid well, whereas the other odor predicts punishment. Early in learning, neurons in the OFC respond to one but not to the other outcome. At the same time, the neurons also begin to respond in anticipation of their preferred outcome. Over a number of studies, 15–20% of the neurons in the OFC developed such activity in this task, firing in anticipation

of either sucrose or quinine presentation [8,9,13]. The activity in this neural population reflects the value of the expected outcomes, maintained in what we have defined here as representational memory.

After learning, these neurons come to be activated by the cues that predict their preferred outcomes, thereby signaling the expected outcome even before a response is made. This is evident in the population response presented in Figure 1, which exhibits higher activity, after learning, in response to the odor cue that predicts the preferred outcome of the neuronal population.