Neural representation of reward information: coding by single cells and populations in rat orbitofrontal cortex

van Duuren, E.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Neural representation of reward information: coding by single cells and populations in rat orbitofrontal cortex

Esther van Duuren
Neural representation of reward information: coding by single cells and populations in rat orbitofrontal cortex

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van Doctor aan
de Universiteit van Amsterdam
op gezag van de Rector Magnificus
Prof. Dr. D.C. van den Boom
ten overstaan van een door het college van promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel

op dinsdag 16 september 2008, te 12:00 uur

door

Esther van Duuren
geboren te ‘s-Gravenhage
The research conducted for this thesis was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)

M.C. Escher’s “The six days of the creation” © 2008 The M.C. Escher Company B.V. - Baarn - The Netherlands. All rights reserved. www.mcescher.com

## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Coding of reward magnitude in the orbitofrontal cortex of the rat</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>of the rat during a five-odor olfactory discrimination task</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Population coding of reward magnitude in the orbitofrontal cortex of the rat</td>
<td>49</td>
</tr>
<tr>
<td>4</td>
<td>Single cell and population coding of expected reward probability in the orbitofrontal cortex of the rat</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>Pharmacological manipulation of neuronal ensemble activity by reverse microdialysis in freely moving rats: a comparative study of the effects of Tetrodotoxin, Lidocaine and Muscimol</td>
<td>109</td>
</tr>
<tr>
<td>6</td>
<td>General Discussion</td>
<td>129</td>
</tr>
</tbody>
</table>

References 145
Dutch Summary 165
List of Publications 169
Famous Last Words 171
Chapter I

General Introduction
Het is nu zes uur 's avonds, het werk zit erop. Ik kan nu een wandeling gaan maken, of ik kan naar de club gaan; ik kan de toren beklimmen om de zon onder te zien gaan, maar ik kan ook naar het theater gaan; ik kan bij deze vriend op bezoek gaan, maar ook bij een andere; ik kan zelfs de stadspoort uitlopen, de wijde wereld in en nooit meer terugkomen. Dit hangt allemaal alleen van mij af; ik heb daartoe de volledige vrijheid. Toch doe ik dat allemaal niet: ik ga even vrijwillig naar huis, naar mijn vrouw.

Dat is precies hetzelfde als dat het water zou zeggen: ik kan hoge golven maken (jazeker, in zee tijdens een storm), ik kan wild omlaag stromen (jazeker, in de bedding van een rivier), ik kan schuimend en bruisend naar beneden storten (jazeker, in een waterval), ik kan in een straal recht de lucht inspuiten (jazeker, als fontein), ik kan ten slotte zelfs verkoken en verdwijnen (jazeker, bij 80° Réamur). Toch doe ik dat nu allemaal niet: ik blijf gewoon vrijwillig als stil en helder water hier in deze spiegelende vijver.

Net zoals het water dit allemaal alleen kan doen door tussenkomst van de oorzaken die het op een of andere manier bewerken, kan die man wat hij meent te kunnen alleen onder diezelfde voorwaarde. Zonder de oorzaken is het hem onmogelijk; zijn de oorzaken echter aanwezig, dan moet hij wel – net zoals het water, zodra het in de juiste omstandigheden wordt gebracht. (...) Dit alles is volmaakt verenigbaar met het 'Ik kan doen wat ik wil' van ons zelfbewustzijn, waarin tot op de dag van vandaag sommige hersenloze filosofasters de vrijheid van de wil menen te bespeuren en deze dierenovereenkomstig als een vaststaand feit van het bewustzijn doen gelden.”

A. Schopenhauer, 1839

At this time, it is still a matter of debate in both philosophy and science whether man truly possesses ‘free will’, or whether his actions are determined. To be able to resolve this controversy, one needs to get insight in the process of how we exercise ‘free will’, that is, how we make our choices. To this end, we need to explore the neural mechanisms underlying the process of decision-making.

One area in the brain considered to be involved in decision-making is the prefrontal cortex, which is located in the rostral pole of the frontal lobes. Initially, insight in the
functional properties of the prefrontal cortex was provided by deficits observed in human patients with prefrontal damage. The most famous patient who sustained frontal damage was Phineas Gage, a railroad foreman whose prefrontal cortex was damaged by an iron bar penetrating through his head during a work-related accident in 1848. Although Gage had no impairments in retrograde and anterograde long-term memory, the alterations in his social and cognitive abilities were so severe that people described him as ‘no longer Gage’. Over the years following this accident, Gage, as noted by his physician, turned into a ‘fitful, irreverent person, indulging at times in the grossest profanity, but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires’ (O’Driscoll and Leach, 1998). Additional research with prefrontal patients revealed that they have normal IQ scores, but suffer from a variety of social and emotional disturbances, have personality changes and display impulsiveness and alterations in goal-directed behavior and decision-making (Bechara 1994, 2004; Brazzelli et al., 1994; Rolls et al., 1994; Verin et al., 1993). These impairments are indicative of the importance of the prefrontal cortex in human functioning, especially in the optimization of complex behavior.

Dysfunction of the prefrontal cortex is also implicated in a number of psychiatric disorders, including schizophrenia and obsessive-compulsive disorder. One of the most consistent abnormalities observed in schizophrenic patients is a decreased volume of the prefrontal cortex (Antonova et al., 2004). In addition, schizophrenia is associated with hypofrontality, meaning that these patients fail to activate the prefrontal cortex during executive functioning (Glahn et al., 2005; Ragland et al., 2007). In contrast, hyperactivity of the orbitofrontal cortex, located in the ventral part of the prefrontal cortex, is demonstrated in patients with obsessive-compulsive disorder (Breiter et al., 1996; Rauch et al., 1994). Patients with obsessive-compulsive disorder have impaired executive functioning and cognitive flexibility, which is consistent with the idea of the orbitofrontal cortex being necessary for the guidance of goal-directed behavior (Cavedini et al., 2006).

This thesis focuses on this particular part of the prefrontal cortex, the orbitofrontal cortex. The orbitofrontal cortex is specifically implicated in the guidance of goal-directed behavior, and hence decision-making, for which it uses predictive information about future outcomes. The following overview will provide, besides general information regarding connectivity and functionality of the prefrontal cortex, in more detail the behavioral and neurophysiological evidence for involvement of the orbitofrontal cortex in the coding of expectations for future rewards and its involvement in the guidance of behavior. Finally, the aims and outline of this thesis will be presented.
The prefrontal cortex: anatomy and connectivity

Demarcation, subdivision and comparative anatomy

The prefrontal cortex is located in the most rostral area of the frontal lobes and is part of the frontal cortex. Historically, the frontal cortex of primates has been divided into two regions: the electrically excitable primary motor and premotor area, which produces motor activity on stimulation, and the prefrontal cortex (Kolb, 1984, 1990). This division furthermore relates to the presence of a granular layer IV, which is absent in the primary motor and premotor cortical areas. However, based on the observation that this granular layer is absent in rodents, other criteria were needed to be able to delineate the prefrontal cortex across different species. Another anatomical characteristic used to define the prefrontal cortex has been originally proposed by Rose and Woolsey (1948), and considers the prefrontal cortex as that part of the frontal cortex receiving projections from the mediodorsal thalamic nucleus. According to this definition, almost all mammalian species probably do have a prefrontal cortex, which in primates and rodents includes the entire cortex of the dorsolateral, orbital and medial frontal lobe rostral to the precentral motor cortex (Öngür and Price, 2000).

The prefrontal cortex in primates can be roughly subdivided into a medial, orbital (or orbitomedial) and dorsolateral region (Fuster, 1997). Whereas the dorsolateral area is implicated in higher cognitive processes, the medial and orbital parts are thought to be more involved in the conduct of emotional behavior. Within these regions, a large variety of cytoarchitectonically distinct subareas can be distinguished, which may contribute differentially to the cognitive and emotional processes mediated by this area (Barbas, 1996; Carmichael and Price, 1994; Fuster, 1997). The rat prefrontal cortex can be subdivided into a number of areas as well, including (i) a medial part, consisting of the medial agranular cortex (area Fr2), the dorsal and ventral anterior cingulate and the prelimbic, infralimbic and medial orbital cortices, (ii) a lateral part that comprises the dorsal and ventral agranular insular cortices and the lateral orbital area, and (iii) a ventral part that encompasses the ventral orbital and ventral lateral orbital cortices (Kolb, 1984; Uylings and Van Eden, 1990) (Fig. 1).

Differences in prefrontal cytoarchitecture and connectional characteristics between primates and rodents make it difficult to establish homologies between the different prefrontal cortical areas in these species. However, based on an overall similarity in anatomy and function, the prelimbic, infralimbic, and orbital areas and the lateral prefrontal cortex in the rat are considered homologous to the primate prefrontal cortex (Brown and Bowman, 2002; Kolb, 1984; Öngür and Price, 2000; Uylings et al., 2003). Whether the prefrontal cortex in the rat contains an equivalent of the primate dorsolateral prefrontal cortex is still a matter of debate. According to
Preuss (1995) there is no such equivalent in the rat, but Uylings et al. (2003) argued that, based on anatomical and functional data, regions of the prefrontal cortex in rats (i.e. Fr2 and the prelimbic region) do display features that resemble characteristics of the primate dorsolateral prefrontal cortex, although it should be kept in mind that in the rat the prefrontal cortex is not as differentiated as in primates.

**Figure 1.** Diagram illustrating the subdivision of the rat prefrontal cortex. (a) Lateral view, (b) coronal section at the location indicated by the arrow in (a). Different shadings represent the major subdivision of the prefrontal cortex in a medial, ventral and lateral part. Abbreviations: ACg, anterior cingulate cortex; AID, dorsal agranular insular cortex; AIV, ventral agranular insular cortex; AOM, medial anterior olfactory nucleus; AOV, ventral anterior olfactory nucleus; cc, corpus callosum; Cg2, cingulate cortex area 2; gcc, genu of corpus callosum; IL, infralimbic cortex; LO, lateral orbital cortex; M1, primary motor area; MO, medial orbital cortex; OB, olfactory bulb; PrL, prelimbic cortex; PrC, precentral cortex; VLO, ventrolateral orbital cortex; VO, ventral orbital cortex. Adapted from Dalley et al., Neurosci Biobehav Rev 28: 771-784, 2004.

**Connectivity**

The main excitatory neurotransmitter in the prefrontal cortex is the amino acid glutamate, which is provided to the prefrontal cortex amongst others by means of its widespread cortico-cortical connections. Both in rats and primates the various subareas of the prefrontal cortex are highly interconnected (Carmichael and Price,
In primates, the distinction between a dorsolateral and an orbitomedial part is also reflected in the pattern of cortico-cortical connections (Fig. 2). The dorsolateral prefrontal cortex receives, besides input from (pre-) motor and orbitomedial areas, input from most sensory modalities, as well as from the parietal, temporal and cingulate cortices. There is a functional distinction between a dorsal and ventral information stream towards the dorsolateral prefrontal cortex, directed to respectively areas located in and around the principal sulcus, and to the inferior convexity in ventrolateral prefrontal cortex (Ungerleider, 1998). The dorsal stream contains fibers from primarily the dorsal parietal lobe, conveying information related to spatial orientation, whereas the ventral streams contains more fibers from the temporal lobe, transmitting information related to the identification and recognition of objects. Hence, the representation of these aspects of the external world in the prefrontal cortex is considered important for the function this area fulfills in cognitive functions such as working memory and the guidance of complex behaviors (Ungerleider, 1998). The cortico-cortical connections of the primate medial and orbital areas characterize the caudal orbital area as visceral- and olfactory-related, and more rostrolateral orbital areas as related more strongly to exteroceptive senses (i.e. vision, audition) (Price et al., 1996).

In rats, the pattern of cortico-cortical connections of the prefrontal cortex is less differentiated as compared to those in primate prefrontal cortex. Area Fr2 and the dorsal anterior cingulate cortex have reciprocal connections with the somatosensory, motor and visual cortices and with temporal association cortices, such as the perirhinal cortex. The prelimbic, infralimbic, and ventral anterior cingulate cortex are reciprocally connected with the perirhinal and entorhinal cortices, as well as with the lateral part of the prefrontal cortex. Furthermore, predominantly the lateral and ventral parts of the prefrontal cortex receive olfactory and gustatory inputs (Carmichael and Price, 1996; Groenwegen and Uylings, 2000; Uylings et al., 2003). In this respect, Fuster (1997, 2004) views the interaction between prefrontal cortex and other cortices as a ‘perception-action cycle’, in which the cycle represents the circular flow of information by which an organism relates to its environment. In this hierarchical concept, the prefrontal cortex is connected with other cortices as long as the behavior contains novelty, uncertainty or ambiguity and has to bridge time intervals using short-term memory. As soon as these constraints disappear and behavior becomes automatic, these functional interactions disappear or weaken; the action is then executed via lower brain structures (Fuster, 2004).
A large number of other brain areas is reciprocally connected with the prefrontal cortex. As already noted, both in rats and primates a topographically organized reciprocal connection exists between the prefrontal cortex and the mediodorsal nucleus of the thalamus, as well as with the intralaminar and midline nuclei, the rostral part of the ventral thalamic complex and the anterior medial thalamic nucleus (Barbas et al., 1991; Berendse and Groenewegen, 1991; Ray and Price, 1992; Reep et al., 1996). In both species, the dorsomedial prefrontal cortex
receives via the mediodorsal thalamic nucleus more exteroceptive sensory input, whereas the ventromedial, orbital and agranular insular areas receive mostly limbic, visceral and olfactory inputs (Groenewegen and Uylings, 2000). In primates, reciprocal connections with the amygdala are organized with a medial to lateral topography across the main amygdaloïd nuclei (the basal, accessory basal and lateral amygdala) and prefrontal areas, with medial areas in both structures projecting to each other and lateral areas projecting to lateral areas (Carmichael and Price, 1995). In rats, projections to the PFC primarily originate from the basolateral amygdala, terminating in the medial and lateral subdivisions of the prefrontal cortex (McDonald, 1991). Hippocampal input originating from area CA1 and the subiculum is also topographically organized and projects in both rats and primates to the medial prefrontal cortex, especially the prelimbic, infralimbic and medial orbital areas, and the lateral prefrontal cortex (Jay and Witter, 1991; Jay et al, 2002). The prefrontal cortex projects to the dorsal and ventral striatum, receiving indirect input from the striatum via the pallidum and thalamus (Groenewegen et al., 1999). There is also a connection with the hypothalamus that is reciprocated, a projection that both in rats and primates primarily applies to the ventral and medial prefrontal cortex (Rempel-Clower and Barbas, 1998; Saper, 1985; Sesack et al., 1989). This connection of specific parts of the prefrontal cortex with the hypothalamus implies the involvement of the prefrontal cortex in autonomic and endocrine functions (Robbins, 2000). In addition, both in rats and primates, the prefrontal areas projecting to the hypothalamus also project directly to various areas in the brainstem that are targeted by the hypothalamus as well, including the periaqueductal grey, the peribrachial nuclei, the nucleus of the solitary tract and parts of the reticular formation (Sesack et al., 1989; Uylings and Van Eden, 1990).

The prefrontal cortex is reciprocally connected with the main sources of the forebrain monoaminergic and cholinergic neurotransmitter systems, including dopamine containing cells in the ventral tegmental area, the noradrenergic cells of the locus coeruleus, serotonergic neurons of the raphe nucleus and the cholinergic magnocellular basal nucleus (Robbins, 2000). These projections enable the prefrontal cortex, by means of its excitatory glutamatergic projection neurons, to exert control over processes within all of these structures, and to regulate input coming from these areas (Robbins, 2000). For example, stimulation of the prefrontal cortex would be expected to excite dopaminergic neurons in the ventral tegmental area, producing an increase in prefrontal dopamine. This was suggested by Jedema and Moghddam (1996), who showed that stimulation of the medial prefrontal cortex with AMPA (which mimics the effect of glutamate), causes an elevation of dopamine within this area. Furthermore, Sara and Herve-Minvielle (1995) demonstrated a strong inhibitory control of prefrontal neurons on the activity of neurons within the locus coeruleus.
Prefrontal cortical functions

Executive functions and working memory

The prefrontal cortex is generally considered to mediate executive functions, which are cognitive control processes that are needed to optimize intended, complex behaviors. These control processes include attentional selection and resistance to interference, monitoring, behavioral inhibition, task switching, planning and decision-making (Dalley et al., 2004). The mediation of cognitive control requires all kinds of information, and the prefrontal cortex, because of its widespread connectivity, is a suitable area to integrate all information needed for the complex sequencing of behavior. This idea is supported by deficits that occur when the prefrontal cortex is damaged: impairments are most apparent in situations in which cognitive control is needed. Prefrontal patients show altered social and goal-directed behavior, impulsiveness or disinhibited behavior and misinterpretation of other people's moods, and they suffer from personality changes (Bechara, 1994, 2005; Brazzelli et al., 1994; Rolls et al., 1994; Verin et al., 1993). These impairments are indicative of the involvement of the prefrontal cortex in impulsive control, behavioral inhibition and decision-making, which are considered executive functions necessary to optimize behavior.

In addition, the prefrontal cortex is thought to play an important role in working memory, which is a temporary memory system used for the active maintenance and elaboration of memory representations for future use, and intimately related to executive functioning. Studies in rodents and primates examining the involvement of the prefrontal cortex in working memory make use of paradigms with delayed response contingencies, including spatial delayed alternation tasks in rats or delayed non-matching to sample tasks in rats and primates. Rats with lesions of prefrontal areas show delay-dependent disruption of task performance (Delatour et al., 1999; Robbins et al., 2000). Furthermore, electrophysiological studies in primates demonstrate the activation of neurons in the dorsolateral prefrontal cortex specifically during the delay (Fuster, 1973; Kojima and Goldman-Rakic, 1982). This activity persists from the time a cue is presented until the response is performed, and is also found to be specific for locations within the visual field (Goldman-Rakic, 1996). These activations are thought to reflect information that is held ‘on-line’, so that it is directly available to be used in decision-making processes (Funahashi, 2006; Goldman-Rakic, 1995). Brain imaging studies in humans have consistently demonstrated prefrontal activations during the performance of variants of the delayed matching-to-sample task as well (see for reviews: D’Esposito et al., 2000; Fletcher and Henson, 2001; Passingham and Sakai, 2004). Furthermore, patients with frontal lobe lesions show impairments in a variety of working memory tasks as well (Müller and Knight, 2006).
It is known that working memory processes are modulated by the dopaminergic system (Dalley et al., 2004; Robbins, 2000). Dopaminergic manipulations in rats and primates that either decrease or increase dopamine levels in the prefrontal cortex produce deficits in working memory tasks, which implies that working memory in the prefrontal cortex may depend on an optimal level of dopamine, presumably according to an inverted ‘U-shaped’ function (Robbins, 2002; Williams and Goldman-Rakic, 1995).

**Attentional processes and attentional set shifting**

The prefrontal cortex has a prominent role in attentional functions (Dalley et al., 2004). A behavioral paradigm that is widely used to assess attentional processes and executive functions in rats is the 5-choice serial reaction time task. In this paradigm, a brief visual stimulus that signals reward can be presented in one of five spatial apertures. Upon a response of the animal in this aperture within a certain period after illumination, a reward is presented. This task captures different aspects of attention and inhibition, including attentional capacity, as indexed by the accuracy of responding, and inhibitory response control, such as compulsive (perseveration) and impulsive (premature) responding (Chudasama et al., 2003b). Lesions of different subareas of the prefrontal cortex produce differential effects on performance in this task: lesions encompassing the cingulate and prelimbic cortices produce deficits in attentional selectivity and a slower latency to respond correctly. Lesions comprising the infralimbic area cause deficits in response inhibition (impulsive premature responding), whereas orbitofrontal lesions increase perseverative responding, but only when the inter-trial interval is long and unpredictable (Chudasama et al., 2003b; Muir et al., 1996).

Other paradigms used to examine attentional processes make use of attentional set shifting. For example, in a primate analogue of the Wisconsin Card Sorting Test, a series of compound visual discriminations is used that requires the monkey either to maintain attention on a certain dimension (e.g. focusing on the color of items presented), and to transfer behavioral control from one pair of exemplars to another within that same relevant perceptual dimension, for example from one pair of blue shapes to another (intra-dimensional shift), or to shift its attentional set from a previously relevant perceptual dimension to another, previously irrelevant dimension, for example from a pair of blue shapes to a pair of white lines (extra-dimensional shift) (Dias et al., 1996). Lesions of the dorsolateral prefrontal cortex, but not the orbitofrontal cortex, impair extra-dimensional attentional set shifting, whereas lesions of the orbitofrontal cortex produce deficits in reversal learning (in which the previously acquired stimulus-reward contingencies are reversed). Similar findings are obtained in an attentional set shifting paradigm in rats: animals with lesions of the medial prefrontal cortex trained to discriminate bowls containing food on the basis
of odor, digging medium or the texture covering the bowls, are impaired in extra-dimensional set shifting, whereas lesions of the orbitofrontal cortex do not impair the acquisition of intra- or extra dimensional set shifting, but do cause deficits in reversal learning (Birrell and Brown, 2000; Brown and Bowman, 2002). These behavioral findings support the idea that the medial prefrontal cortex is necessary to preserve attentional selectivity to relevant stimulus features during behavior, whereas ventral and lateral regions appear to be more critical for inhibitory response control or behavioral adaptation to previously established, but altering reward contingencies.

The ascending monoaminergic and cholinergic systems are found to contribute to different aspects of the 5-choice serial reaction time task (Robbins, 2002). The observation that lesions of the cholinergic neurons of the magnocellular basal nucleus produce deficits in performance of the 5-choice serial reaction time task is in line with the finding in rats that neuronal activity associated with increased attentional demands within the medial prefrontal cortex is modulated by cholinergic afferents (Gill et al., 2000). Depletion of noradrenalin and dopamine from the medial prefrontal cortex also results in attentional impairments in the 5-choice serial reaction time task (Robbins, 2002), suggesting the involvement of multiple neurotransmitter systems during task performance.

**Associative learning: action-outcome associations, habits and extinction learning**

During conditioning, animals learn contingencies between particular stimuli or actions and outcomes, which allow the prediction of events occurring in the environment. When animals perform a goal-directed action, which corresponds to the human-oriented concept of ‘intended’ behavior, they learn the contingency between performing a certain behavioral act and the outcome (also called action-outcome learning). When the outcome is as desired, an action-outcome association is learned. This in contrast to stimulus-response learning, a form of learning that does not integrate knowledge of outcome in the learned association and which can result in habit formation by overtraining, in which actions that were originally goal-directed become automatic and habitual.

The prelimbic cortex is required for the detection of action-outcome relationships, but seems less important for tasks that can be acquired through habit formation (Cardinal et al., 2002). As demonstrated by Ostlund and Balleine (2005) in a devaluation task used to discriminate between stimulus-response and action-outcome learning, the prelimbic medial prefrontal cortex is specifically involved in the acquisition of action-outcome representations, but not in the expression of action-outcome associations during instrumental learning. Furthermore, animals with lesions of prelimbic medial prefrontal cortex or orbitofrontal lesions within this task are impaired in adjusting their behavior upon the changed incentive value of the reinforcer (i.e. they exhibit habit-like task performance) (Balleine and Dickinson,
A possible neural correlate of action-outcome learning has been described in the medial prefrontal cortex by Mulder et al. (2003) during a lever-press task in rats. In contrast to the prelimbic cortex, lesions of infralimbic medial prefrontal cortex caused animals to remain goal-directed (Killcross and Coutureau, 2003) while temporary inactivation of the infralimbic medial prefrontal cortex reinstated goal-directed responding (Coutureau and Killcross, 2003).

The involvement of the prefrontal cortex in classical conditioning has been studied less than its involvement in instrumental conditioning (but see Mulder et al., 2003). In fear conditioning, lesions of the prefrontal cortex generally do not affect acquisition (Morgan et al., 1993). Effects consist of an increase in fear responses to the conditioned stimulus or context (Morgan and LeDoux, 1995). Furthermore, lesions of rat orbitofrontal cortex impair the acquisition of Pavlovian autoshaping, whereas lesions of the infralimbic medial prefrontal cortex do not (Bussey et al., 1997; Chudasama et al., 2003a). However, lesions of the infralimbic medial prefrontal cortex were found to enhance recovery and reinstatement of an appetitive Pavlovian response after extinction (Rhodes and Killcross, 2004, 2007). As regards extinction learning, from lesion studies during extinction of conditioned fear it became apparent that the ventral medial prefrontal cortex (i.e. prelimbic and infralimbic parts of the prefrontal cortex) is involved in the extinction of fear conditioning, whereas the dorsal part of the medial prefrontal cortex does not seem to be essential for this type of learning (Morgan and LeDoux, 1995, 1999). Involvement of the infralimbic cortex in extinction of conditioned fear is furthermore supported by the finding that neuronal firing within this area correlates with extinction of responding to an avertively paired cue (Milad and Quirk, 2002).

An interesting point concerns the functional heterogeneity in the prefrontal cortex in relation to the different components of executive functioning. The behavioral studies in rats and primates as described above are consistent with the notion of functionally dissociable regions within the prefrontal cortex (Bussey et al., 1997; Chudasama et al., 2003, 2004; Dias et al., 1996; Passetti et al., 2002), but it remains largely unknown how these regions are organized, for example in a hierarchical manner or as a heterarchical group (Fuster, 1997b). Since the primary focus of this thesis concerns the rat orbitofrontal cortex, which is a part of the prefrontal cortex that appears to be specifically required for adapting already established behavioral responses in reaction to changes in the environment, the connectivity and functionality of this part of the prefrontal cortex are discussed in more detail below.
The orbitofrontal cortex: connectivity and functions

Connectivity

Thalamic afferents to the various subareas of the orbitofrontal cortex are topographically organized, and originate primarily from the submedial and mediodorsal nuclei (Reep et al., 1996). The ventrolateral area of the orbitofrontal cortex receives projections from the entirety of the submedial nucleus, whereas the other orbital areas receive projections from the periphery of this nucleus. In addition, each orbital area is connected with a particular segment of the mediodorsal nucleus, including the dorsocentral, the central and the lateral segments. In the rat, the medial and central regions of the mediodorsal thalamic nucleus receive direct afferents from a number of structures, including the amygdala, which is a complex of nuclei involved in the processing of affective and motivational aspects of learning, and the piriform cortex sending olfactory input (Ray and Price, 1992). Hence, olfactory and associative information is provided to the orbitofrontal cortex through this thalamic input. These types of information also reach the orbitofrontal cortex directly by means of its reciprocal connections with the basolateral nucleus of the amygdala and piriform cortex (Ray and Price, 1992). The direct and indirect pathways from the piriform cortex converge on the same region in the orbitofrontal cortex (lateral and medial orbitofrontal cortex), but originate from neurons that differ in their morphology (Ray and Price, 1992). Neurons from piriform cortex projecting to the orbitofrontal cortex through the mediodorsal nucleus of the thalamus are large multipolar neurons, whereas cells projecting directly to the orbitofrontal cortex are small and superficially located within the piriform cortex (primarily layers II and III). The existence of these two pathways suggests that olfactory information is carried by two different information streams. A similar pattern is demonstrated by the direct and indirect pathway originating from the basolateral amygdala: neurons projecting via the mediodorsal nucleus of the thalamus appear to be larger and better suited to integrate activity over a wide region than those projecting directly to orbitofrontal cortex (Ray and Price, 1992).

Cortico-cortical connections of the orbitofrontal cortex in rats are more extensive for the ventral and ventrolateral areas than for the medial and lateral orbital areas (Reep et al., 1996). The ventral and ventrolateral parts are connected with the cingulate cortex, FR2, and several areas of the visual and somatosensory cortices (Par1, Par2, PPC, Oc2M and Oc2L). The medial orbital area has connections with the cingulate cortex, FR2 and the posterior parietal cortices, whereas the connections of the lateral part are limited to the granular and agranular insular cortices and the somatosensory cortex. In addition, the orbitofrontal cortex is reciprocally connected with other subareas of the prefrontal cortex (Carmichael and Price, 1996; Öngür and Price, 2000), and provides a strong efferent projection to the nucleus accumbens.
(part of the ventral striatum) that overlaps with innervations from limbic structures, including the amygdala (Berendse et al., 1992; Groenewegen et al., 1990; McDonald, 1991). The nucleus accumbens is thought to act as an interface between the limbic and motor system, integrating information from sensory modalities and cortical areas to prepare for and invigorate behavioral actions (Swanson and Morgenson, 1981; Pennartz et al., 1994).

**Orbitofrontal cortical functions: behavior**

Based on its position in the circuitry outlined above, the orbitofrontal cortex is able to integrate sensory, motivational and associative input with information regarding planned actions, and applies this information to the guidance of goal-directed behavior. This hypothesis is consistent with the effects on behavior when the orbitofrontal cortex is damaged. Humans with lesions of the orbitofrontal cortex show alterations in emotional state, a lack of concern for the future and socially inappropriate or disinhibited behavior (Bechara, 1994, 2004; Berlin et al., 2004; Damasio, 1994; Hornak et al., 2003; Rolls et al., 1994). In formal testing situations, patients with orbitofrontal damage are unable to rapidly learn reversals of previously acquired stimulus-reinforcer associations (Fellows and Farah, 2003; Hornak et al., 2004; Rolls et al., 1994). Interestingly, patients can report verbally that task contingencies have changed, but are unable to alter their behavior accordingly (Rolls et al., 1994). This type of dysfunction, the inability to alter established behavior in response to a change in reinforcer associations, may in part underlie the social deficits found in those patients, since social and emotional interaction often requires the inhibition of inappropriate responses. In addition, it is suggested that emotion-based biasing signals that arise from the body are integrated within the orbitofrontal cortex to regulate decision-making in complex and uncertain behavioral situations (Damasio, 1994; Bechara et al., 1996). These presumed ‘somatic marker’ signals (skin conductance responses or SCRs) were found to be absent in patients with orbitofrontal lesions when they anticipated future outcomes, linking the presence of SCRs to successful performance on a decision-making paradigm in healthy subjects. However, questionability of both the behavioral task used and the interpretation of the findings in these experiments, together with a shortage of evidence for causally linking peripheral feedback to the performance in the task used, make the empirical evidence for this hypothesis rather weak (for a critical review of the somatic marker hypothesis, see Dunn et al., 2006). Support for the idea that the orbitofrontal cortex is involved in the mediation of emotions that influence decision-making comes from recent imaging studies that implicate the orbitofrontal cortex in the experience of regret, which is a negative emotion affecting choice behavior (Coricelli et al., 2005). For example, in a gambling task, normal subjects learned from regret caused by their suboptimal choices and tried to minimize future regret, whereas subjects with
orbitofrontal lesions reported to not experience regret and were unable to anticipate the negative outcomes of their future choices (Camille et al., 2004).

Deficits in reversal learning arising upon orbitofrontal damage as displayed by humans occur in other species as well. Both rats and monkeys are able to normally acquire stimulus-reward associations, but are impaired when the contingencies are altered (Bohn et al., 2003; Chudasama and Robbins, 2003a; Dias et al., 1996; Fellows and Farah, 2003; Ferry et al., 2000; Izquierdo et al., 2004; Kim and Ragozinno, 2005; McAlonan and Brown, 2003; Meunier et al., 1997; Schoenbaum et al., 2002, 2003b). Since the initial acquisition of the stimulus-reward associations remains unaffected and responses at this stage can be inhibited when appropriate, this impairment in reversal learning cannot be explained by a simple and general deficit in response inhibition due to orbitofrontal damage. Hence, the orbitofrontal cortex is thought to be involved not only in the representation of information regarding the incentive value of conditioned and unconditioned stimuli (Pears et al., 2003), but also in the process by which conditioned stimuli access a specific representation of the incentive value of the response outcome to guide behavior (Gallagher et al., 1999; Roberts et al., 2006). According to this hypothesis, damage to the orbitofrontal cortex disables the representation of the conditioned stimulus to access the representation of the associated reinforcer, which is needed to use the motivational guidance provided by the response outcome when environmental changes demand an adaptation in behavioral responses.

Besides the deficits reported in reversal learning, more evidence for this idea is provided by reinforcer devaluation paradigms. In these tasks, rats are trained to associate e.g. a light with a reinforcer. After conditioned responses to the light are established, the incentive value of the food is reduced (‘devaluated’) by pairing it with a toxin in the home cage in the absence of the light, to form a conditioned taste aversion. Upon re-exposure to the light, normal animals spontaneously decrease their conditioned response to the light cue, in accordance with the low value of the food reward. Rats with orbitofrontal lesions show normal acquisition of conditioned responding to the light and a normal reduction in food consumption after food-toxin pairing, but fail to show a decrease in conditioned responding to the light after reinforcer devaluation (Gallagher et al., 1999), which is an effect that is also observed when lesions are made later in the task, namely after associative learning (Pickens et al., 2003). This indicates that the orbitofrontal cortex is not only involved in the acquisition of the cue-outcome associations, but also in the control of conditioned responding according to the new representation of the devaluated response. Similar results are obtained in a devaluation task in primates (Izquierdo et al., 2004b), in which monkeys are trained to discriminate different visual objects to obtain different rewards. After devaluation of one of the rewards by food satiation, monkeys are offered a choice between the visual objects associated with the
devaluated versus the non-devaluated reward. Normal monkeys bias their choice towards the objects that are not associated with the devaluated reward, whereas in monkeys with lesions the effect of reward devaluation is attenuated (Izquierdo et al., 2004b).

It has been demonstrated that normal performance in devaluation paradigms requires an interaction between basolateral amygdala and orbitofrontal cortex (Baxter et al., 2000; Izquierdo et al., 2004a). Disconnection lesions of the basolateral amygdala and orbitofrontal cortex (orbitofrontal cortex damaged in one hemisphere and the amygdala in the other) or combined lesions of these two structures cause a failure in reducing conditioned responding after devaluation (Baxter et al., 2000; Izquierdo et al., 2004a). Examination of the specific contribution of these two structures in the devaluation paradigm reveals that both (bilateral) lesions of the orbitofrontal cortex or basolateral amygdala made prior to learning causes impairments in the devaluation paradigm (Hatfield et al., 1996; Gallagher et al., 1999), but when basolateral amygdala lesions are made after the conditioned response to the light is established no effect on behavior is observed, whereas lesions of the orbitofrontal cortex made at this stage continue to cause behavioral effects (Pickens et al., 2003). Lesions of the orbitofrontal cortex continue to have an effect on behavior even when they are made after the stage when the reinforcer is devaluated (Pickens et al., 2005). This suggests that the basolateral amygdala is involved in coding of the original associations, while the orbitofrontal cortex is involved in monitoring and updating the current value of the reinforcer and to use that information to guide behavior. This dissociable role for the amygdala and orbitofrontal cortex is also evident from an odor-guided discrimination and reversal task, in which lesions of either basolateral amygdala or orbitofrontal cortex exerted different effects on performance (Schoenbaum et al., 2003a). Neither lesions of the amygdala nor orbitofrontal cortex caused deficits in the acquisition of the initial stimulus-reward association, but after reversal orbitofrontal or amygdala lesions caused distinct impairments. Rats with lesions of the basolateral amygdala were impaired on the first reversal, showing no deficit in acquiring the reversal back to the original contingencies. Rats with lesions of the orbitofrontal cortex, however, were impaired on both types of reversals. This is indicative of a somewhat independent role of the orbitofrontal cortex during reversal learning (except for the first reversal), but both structures seem to have a role in the use of associative information during discrimination learning.

In addition to reversal learning and devaluation paradigms, evidence that supports a role of the orbitofrontal cortex in the integration of reinforcer value in decision-making processes comes from studies applying delayed, uncertain or probabilistic reward paradigms. For example, humans with orbitofrontal lesions are impaired in performing the Iowa Gambling Task (Bechara et al., 1997, 1999; Ptak
et al., 2000). In this task, subjects must choose a card from a deck associated with rewards and penalties of different sizes. ‘Bad’, high-risk decks are associated with a large reward but occasionally also with large penalties, and ultimately lead to an overall loss, whereas ‘good’, low-risk decks yield low rewards and no penalties, leading to an overall gain. Normal subjects initially choose high-risk decks, but bias their choice behavior towards the low-risk decks later on. Orbitofrontal patients fail to change their choice behavior towards the low risk decks, despite the fact that they have an accurate account of the task and of the correct strategy. Apparently, the more delayed and probabilistic penalties are not used to guide choice behavior by these patients (Bechara et al., 1997, 1999; Ptak et al., 2000). In addition, damage to the orbitofrontal cortex impairs performance on a task in which subjects are required to choose between two possible outcomes and to bet on their choice; patients with orbitofrontal damage decide slowly and fail to choose the optimal, most likely outcome (Rogers et al., 1999a). Recent imaging studies provide further support for a role of the orbitofrontal cortex in decision-making under uncertainty (Ernst et al., 2004; Hsu et al., 2005). For example, using PET imaging, the choice between small, likely rewards and large, unlikely rewards was shown to increase blood flow in orbital areas (Rogers et al., 1999b). In addition, visual stimuli associated with higher uncertainty elicit increasing activations in orbital areas, which covaries with risk-seeking and risk aversion activity within this area (Tobler et al., 2007).

Behavioral lesion studies in rats indicate involvement of the orbitofrontal cortex in the evaluation of reward under probabilistic or uncertain conditions. Using behavioral tasks in which delay discounting or probability discounting paradigms are applied (respectively the free-choice between immediate small rewards and delayed but larger rewards, or between a certain small reward and an uncertain larger reward with fixed probability of delivery), it was demonstrated that the orbitofrontal cortex modulates the threshold at which animals switch to preferring immediate or reliable but smaller rewards over delayed or unreliable but larger rewards (Kheramin et al., 2003; Mobini et al., 2002). According to these studies, orbitofrontal lesioned rats tend to be more risk-aversive than normal rats when making choices, which is in contrast with the finding in humans that orbitofrontal damage leads to more ‘risky’ behavior. However, Winstanley et al. (2004) reported that rats with orbitofrontal lesions increasingly prefer the large, delayed reward in a delay discounting paradigm. A possible explanation for the discrepancy between these rat studies might be related to whether the orbitofrontal lesion is made before or after training in the behavioral task, since in the study by Winstanley et al. animals were trained before the orbitofrontal cortex was damaged and retested afterwards, whereas Mobini et al. trained and tested postoperatively. A more recent study, which uses an alternative probability discounting task that mimics the fundamental aspects of the gambling tasks as used in humans better than the previously used paradigms for rats,
demonstrated that rats with orbitofrontal lesions made after training indeed preferred the larger but uncertain reward, with stronger risk-taking behavior as demonstrated by human experimental data (Pais-Vieira et al., 2007).

The main findings of these behavioral lesioning studies involving delay- or probability discounting paradigms, an increased choice of risky options in a gambling task and an increased choice of a large uncertain or delayed reward after orbitofrontal damage, probably result from an impaired integration of the representation of the incentive value of the reinforcer with the representation of consequences of responding for that particular reinforcer (Winstanley et al., 2004). Hence the ‘punishment’ (loss of money in the gambling task or highly uncertain or delayed reward) is failing to induce a devaluation of the response outcome and to alter choice behavior.

To summarize, behavioral studies in rats and primates indicate that the orbitofrontal cortex is involved in the flexible guidance of goal-directed behavior, a function that requires the integration of representations of the incentive value of stimuli with schemes for appropriate behavioral responses.

**Orbitofrontal cortical functions: neurophysiology of reinforcement expectancy**

Associations that link initially neutral cues with the incentive value of reinforcers provide a framework for constructing expectancy in goal-directed behavior. Expectations of likely outcomes reflect judgments about chances of receiving a future reinforcer using associative information that is acquired during learning. This predictive information is updated continuously, providing an internalized representation of possible future outcomes that can be used to respond in an appropriate manner when changes in the environment ask for adaptations in behavior. For example, Montague and Berns (2002) developed a computational model (the predictor-valuation model, or PVM) of decision-making, in which the orbitofrontal cortex integrates information regarding rewards and punishments and their predictors, and supports the conversion of disparate types of future rewards into a common neural currency, i.e. a common scale used to compare the valuation of future behavioral acts or stimuli, on which decisions can be based. In another computational model, proposed by Pennartz (1997), expectations for future outcomes are compared with actual outcomes, a process that involves the glutamatergic projections neurons of the orbitofrontal cortex and basolateral amygdala. Subsequently, these projection neurons instruct sensorimotor regions, such as the striatum and pre- and supplementary motor cortical areas, to optimize stimulus-response mapping (Pennartz et al., 2000).

That animals use expectations about future outcomes was already demonstrated in a behavioral study by Tinklepaugh published in 1928. In this study, monkeys perform a delayed-response task for different kinds of food reward. In
particular trials, the monkey is shown her favorite food (banana) placed in one of the food cups in front of her. During the delay, in which the animal must maintain the information about the location of the banana, the experimenter replaces the banana for a less favorite reward, a piece of lettuce. Discovering the lettuce instead of the banana, the monkey, to use the author's terminology, 'becomes angry': she does not touch the lettuce but instead looks around for the banana, and sometimes shrieks to the observers in apparent anger. However, when lettuce is used as cue, so that the monkey expects the lettuce after the delay, she does not become angry but responds normally and eats the lettuce. Apparently, the monkey only becomes angry when the reward is different than expected.

Human brain imaging studies have demonstrated the involvement of the orbitofrontal cortex in expectancy-related signaling (Breiter et al., 2001; Gottfried et al., 2003; O’Doherty et al., 2002). Furthermore, Nobre et al. (1999) showed using PET imaging that when learned and expected stimulus-reward associations were violated, the orbitofrontal cortex is activated. In addition, firing characteristics of orbitofrontal neurons also indicate that predictive information concerning outcomes is present within the orbitofrontal cortex. Neurons in the orbitofrontal cortex do not only respond to sensory features of predictive stimuli, but their firing patterns during sampling of these stimuli also reflect the expected outcomes (Rosenkilde et al., 1981; Thorpe et al., 1983). For example, manipulating the aversiveness of a syringe alters the response of orbitofrontal neurons: when the syringe provides glucose to the animal, no response to the sight of the syringe is observed, but when the same syringe is used to deliver aversive hypertonic saline, neurons do become responsive to the sight of the syringe. Other findings support the occurrence of cue-selective firing within the orbitofrontal cortex that reflects predicted outcomes. Schoenbaum and Eichenbaum (1995a) reported that in an 8-odor olfactory discrimination task, in which four odors are associated with reward and four with no reward, the large majority of neurons discriminates during cue sampling between rewarded and non-rewarded odors, whereas only a minority discriminates between the different, reward-associated odors. Similar findings were obtained in primates performing an odor discrimination task, in which they learn to discriminate eight rewarded odors from two odors associated with an aversive saline solution. Also in this task firing activity of orbitofrontal neurons reflected, besides odor identity, the reinforcement association of the odors (Critchley and Rolls, 1996). In addition, several studies in rats and primates report switching of cue-selective firing of orbitofrontal neurons during reversal learning (Rolls et al., 1996; Schoenbaum et al., 1999; Thorpe et al., 1983).

Expectations for upcoming reinforcers have not only been found to be coded during cue sampling. Studies in rats and primates that incorporated a delay before a particular reinforcer is delivered demonstrated that neural firing in the orbitofrontal
cortex during this delay reflects expectations for different response outcomes, showing differential firing activity during the anticipation of various types or amounts of reinforcers, or reflects relative reward preference (Hikosaka and Watanabe, 2000; Ichihara-Takeda and Funahashi, 2006; Padoa-Schioppa and Assad, 2006; Roesch and Olson, 2004; Roesch et al., 2006; Schoenbaum et al., 1998; Tremblay and Schultz, 1999; Wallis and Miller, 2003). This differential activity appears to represent the expectation of the animal for the likely outcome. However, such activity was also demonstrated in other areas, including the basolateral amygdala (Schoenbaum et al., 1998) and other prefrontal regions (Wallis and Miller, 2003). Interestingly, expectancy-related activity in these areas seems to develop only after this activity has appeared in the orbitofrontal cortex (Roesch and Schoenbaum, in press): outcome-expectant activity was found to appear first in the orbitofrontal cortex and later in the basolateral amygdala. Furthermore, outcome-expectant firing within the basolateral amygdala was also demonstrated to depend on input from the orbitofrontal cortex. In contrast, during cue-sampling, selective activity seems to develop first in the basolateral amygdala, and development of cue-selective firing in OFC seems to be dependent on input from the basolateral amygdala (Saddoris et al., 2005; Schoenbaum et al., 1999, 2003a). As proposed by Roesch and Schoenbaum, the basolateral amygdala takes the lead to activate the network when external cues are present, as is the case during the cue period, whereas in the absence of external cues, for example during the delay-period, the orbitofrontal cortex is dominant in activating the network. Hence, the asymmetrical flow between orbitofrontal cortex and basolateral amygdala is distinguished primarily by the representational memory demand (Roesch and Schoenbaum, in press).

Besides coding of reward predictive information it was demonstrated that the orbitofrontal cortex is also involved in odor-place associations (Lipton et al., 1999). More recent, a role of the orbitofrontal cortex in the representations of spatial goals was demonstrated as well (Feierstein et al., 2006; Roesch et al., 2006). In these studies, the response site was explicitly coupled to differently valued rewards, and orbitofrontal neurons were found to encode the direction of choice or the location of the goal. Moreover, Feierstein et al. (2006) reported that a subset of neurons encoded both the direction and choice outcome. Hence, according to these authors, orbitofrontal neurons, in contrast to a pure valuation signal (Montague and Berns, 2002; Roberts, 2006), encode specific spatial-motor variables for the representation of goal locations, together with the responses needed to obtain that goal.

Aims and outline of this thesis

The general aim of this thesis is to examine the coding of predictive information regarding several reward-related parameters (i.e. magnitude and
probability of reward) by single units and ensembles in the orbitofrontal cortex of the rat. Electrophysiological recordings are made while animals perform an olfactory discrimination learning task, in which they are required to learn response outcomes based on the identity of a predictive stimulus. Since outcomes in this task can be appetitive or aversive, animals need to decide whether they execute a behavioral response or not. Examination of neural activity during this task will provide more information about how reward parameters, which can be considered main determinants of behavior, are represented in the brain, and how these neural representations may regulate decision-making processes.

In Chapter 2, we examine the coding of predictive information about reward magnitude by single units in the orbitofrontal cortex. To this end, we perform ensemble recordings during an olfactory discrimination ‘go/no-go’ task, in which five different odor stimuli are predictive for different amounts of a rewarding sucrose solution, no reward or an aversive reinforcer, i.e. quinine. Since the orbitofrontal cortex is known to code predictive information regarding upcoming reinforcers of different quality, we hypothesized that during several task phases, predictive information about reward quantity would be encoded in the orbitofrontal cortex as well.

In Chapter 3, we elaborate on the encoding of reward magnitude in the orbitofrontal cortex by examining how information about this reward parameter is represented by orbitofrontal ensembles. We use different reconstruction methods to reconstruct whether activity of orbitofrontal ensembles represents predictive information about reward magnitude, and to examine in more detail how this activity develops as the task progresses, during which task phases and on which time scales.

In Chapter 4, we extend our experiments to another reward parameter, the probability of reward. Using a similar paradigm as employed in the previous two chapters, we examine the neural coding of reward probability by both single units and ensembles in the orbitofrontal cortex. As the orbitofrontal cortex is indicated to be involved in decision-making under uncertainty, we hypothesized that neural activity in the orbitofrontal cortex represents information about reward probability. Of special interest in this chapter is the question whether orbitofrontal ensembles, active during an odor-reward probability task, encode the uncertainty or probability of a rewarding outcome.

To add to new developments in the field of ensemble recordings, we have developed the combidrive, as described in Chapter 5. The combidrive combines ensemble recordings with the technique of microdialysis, and might bring closer understanding of the interaction between neurotransmitters and cell populations during ongoing behavior. The general aim of this study was to develop an instrument for studying effects of local drug infusions on ensemble firing patterns during
ongoing behavior and cognitive processing. Specifically, we validated the instrument by comparing how three inactivating compounds (i.e. lidocaine, tetrodotoxin and muscimol) affected the capacity of frontal-cortical ensembles to sustain firing.

In the General Discussion contained in Chapter 6, the results of this thesis are discussed in the context of current ideas on the functional role of the orbitofrontal cortex, especially in relation to reversal learning.
Chapter 2

Neural coding of reward magnitude in the orbitofrontal cortex of the rat during a five-odor olfactory discrimination task
Abstract

The orbitofrontal cortex (OFC) has been suggested to code the motivational value of environmental stimuli and to use this information for the flexible guidance of goal-directed behavior. To examine whether information regarding reward prediction is quantitatively represented in the rat OFC, neural activity was recorded during an olfactory discrimination ‘go/no-go’ task in which five different odor stimuli were predictive for various amounts of reward or an aversive reinforcer. Neural correlates related to both actual and expected reward magnitude were observed. Responses related to reward expectation occurred during the execution of the behavioral response towards the reward site and within a waiting period prior to reinforcement delivery. About one-half of these neurons demonstrated differential firing towards the different reward sizes. These data provide new and strong evidence that reward expectancy, regardless of reward magnitude, is coded by neurons of the rat OFC, and are indicative for representation of quantitative information concerning expected reward. Moreover, neural correlates of reward expectancy appear to be distributed across both motor and non-motor phases of the task.
Introduction

It has been noted for a long time that the magnitude of a primary reinforcer exerts a profound effect on the selection and speed of behavioral responses (Black, 1968; Campbell and Seiden, 1974; Brown and Bowman, 1995; Boysen et al., 2001; Bohn et al., 2003). Likewise, in computational neuroscience different algorithms for Reinforcement Learning (RL) consider reward magnitude an important parameter to be gauged and predicted during sensorimotor processing (Sutton and Barto, 1981; Schultz et al., 1997). In one of these models, in which glutamate serves as a reinforcing signal guiding synaptic modifications necessary for adapting operant behavior, reward-related information is primarily processed by glutamatergic projection neurons of the orbitofrontal cortex (OFC), basolateral amygdala and related limbic areas (Pennartz, 1997). The OFC is known to be involved in the representation of the motivational significance of stimuli and in applying this information to the guidance of goal-directed behavior (Thorpe et al., 1983; Gallagher et al., 1999; Lipton et al., 1999; Schoenbaum et al., 1999, 2003; Yonemori et al., 2000; O’Doherty et al., 2003). Accordingly, rats with OFC lesions exhibit impairments when stimulus-reward contingencies are altered during olfactory discrimination learning, whereas the initial acquisition of those associations remains unaffected (Ferry et al., 2000; Schoenbaum et al., 2002). Monkeys and humans with damage to the OFC demonstrate similar deficits: both show impairments in reversal learning and decision-making (Dias et al., 1996; Meunier et al., 1997; Baxter et al., 2000; Manes et al., 2002; Hornak et al., 2004; Izquierdo et al., 2004).

Another consideration for a central role of the OFC in RL is the neurophysiological evidence for the coding of predictive information regarding upcoming reinforcers. Neurons in primate OFC process information concerning expected outcomes of behavioral responses, showing differential firing activity during the anticipation of various types or amounts of reinforcers (Tremblay and Schultz, 1999; Hikosaka and Watanabe, 2000; Wallis and Miller, 2003; Roesch and Olson, 2004; Ichihara-Takeda and Funahashi, 2006; Padoa-Schioppa and Assad, 2006). In addition, findings in rat OFC suggest predictive neural coding of both appetitive and aversive outcomes (Schoenbaum et al., 1998). However, a difficulty in interpreting the latter findings is that ‘go’ responses for the rewarding outcome were compared with ‘go’ responses for an aversive outcome, the latter being a response that was erroneously made. Thus, the neural activity seen in the waiting period during these ‘false alarm’ responses may not be related to the expectancy of a particular (aversive) outcome, but, considering OFC functioning, for example to a signal reflecting a conflict, error in responding, or an internal requirement for cognitive flexibility that will result in adjustment of the behavioral response. A second motivation for the current study on coding of reward magnitude was that many previous studies
employed only two reward sizes. A limitation of such a design is that it is not possible to characterize the tuning relationship between reward size and neural activity. In the current five-odor olfactory discrimination 'go/no-go' task, three odors were associated with a parametrically varied amount of reward, allowing the comparison of neural activity between trial types with highly similar behavior (namely all correct 'go' responses). Thirdly, while previous studies in rats focused on reward representation in an immobile trial phase preceding reward delivery by a short delay, the current study also paid attention to the possible occurrence of reward-expectancy correlates during the motor phase of the trial, i.e. when the animal actually performed a 'go' response towards the reward site. When effects of the expected reward size could be uncovered during voluntary, operant action, this may help to understand the formation and representation of action-outcome associations (Balleine and Dickinson, 1998; Dayan and Balleine, 2002), an issue that has remained underexposed in orbitofrontal cortex studies performed until now, including those in primates.

**Materials and Methods**

**Subjects**

Data were collected from 7 male Wistar rats (Harlan CPB; Horst, The Netherlands), weighing 325-425 g at the time of surgery. Animals were socially housed in standard macrolon cages, weighed and handled daily, and kept under a reversed 12 hr light/dark cycle (lights off at 7:00 AM) with food available *ad libitum* (standard rat chow; Hope Farms, The Netherlands). Water deprivation started overnight prior to behavioral testing to motivate the animals to perform the task. During performance, animals received a maximum of 7.5 ml fluid, and after finishing the session there was, with a variable delay, free access to water for a 2 hour period. After surgery the animals were housed individually under the same conditions. All experiments were carried out in accordance with the National Guidelines for Animal Experimentation.

**Behavior**

*Apparatus*

Behavioral testing was performed in a plexiglas operant chamber that was placed in a sound attenuated and electrically shielded box, with behavioral events and data collection controlled by a computer. The recording chamber (40 x 37 x 41.5 cm) had a black interior with straight walls and a front panel that contained on the right side an odor sampling port beneath a light indicating trial onset, and on the left side a fluid delivery well. The lower part of the front panel was placed at an angle of
101° with respect to the grid floor, with an additional angle higher up of 161° with respect to the lower part of the front panel to ensure that during recordings, when animals were attached to the recording equipment, ample space was left for the rat to put its snout into the odor sampling port. To detect the odor- and fluid poke responses made by the animals, both the odor sampling port and fluid delivery well were equipped with an infrared beam transmitter and detector. During six recording sessions, licking behavior in the fluid well was monitored with the use of an electronic circuit with fluid contact causing a current of 150 nA and a concomitant change in voltage.

The delivery of odors was regulated by a system of solenoid valves and flow meters (cf. Schoenbaum, 2002) with separate delivery lines for each odor to prevent mixture of odors in the system. To be able to deliver different types of fluid reinforcement (i.e. quinine and sucrose solutions), separate fluid delivery lines were present. Fluid delivery was gravity-driven, with a tap and valves controlling the flow and amount of fluid delivered. The odorants (Tokos BV; Noordscheschut, The Netherlands) were separated into different families, i.e. herbal, floral, woody, citrus and fruity. In order to have distinct odors in a set of five odors used in a single discrimination session, each set contained one odor from each family. Care was taken to ensure that no single family of odors was preferentially associated with a particular trial outcome (positive versus negative and in terms of reward size).

**Shaping and behavioral paradigm**

After the animals were habituated to the recording chamber they were gradually trained on the behavioral procedure of the five-odor olfactory discrimination ‘go-no/go’ task. A set of five different odors was used for any discrimination session: three odors associated with a particular amount of a positive reinforcement (10% sucrose in water, i.e. 0.05, 0.15 and 0.30 ml), one odor with no reward (non-reinforced condition), and one odor with a negative reinforcement (0.15 ml of a 0.015M quinine solution in water). A pilot test in a T-maze was performed to examine the animal’s ability to discriminate between the volumes of sucrose solution used. This test demonstrated that animals selected the larger amount of reward when offered a choice between two alternatives (e.g. 0.05 versus 0.15 ml or 0.15 versus 0.30 ml), indicating they were able to discriminate between the different reward sizes. The odors coupled to the non-reinforced condition and quinine were used to determine whether the animal was actually capable of associating an odor with a particular outcome, as should be visible by withholding responses towards or at the fluid well after sampling the odors that were predictive of these two outcomes.

During shaping, animals were initially trained to make a nose poke in the odor sampling port (“odor poke”), which was sufficient to immediately obtain
reinforcement by visiting the fluid well. At this stage only two different odors were used, associated with 0.05 and 0.15 ml of sucrose solution. At the next stage a third odor was introduced, associated with 0.30 ml of sucrose solution. In this phase animals were trained to make an odor poke with a minimal duration of 2 s: the animals had to wait 1 s in the odor sampling port before the odor was presented in order to have their body stationary during cue sampling, and odor sampling itself was required to last at least 1 s. In the final stage of shaping the odors coupled to the non-reinforced condition and quinine were introduced, together with a waiting period of 1500 ms in the fluid well following the nose poke in the well (“fluid poke”), before reinforcement was delivered. The behavioral sequence comprising the departure from the sampling port to the fluid well, including nose entry and waiting period in the well, will be referred to as the ‘go’ response.

The criterion for behavioral performance was set at 15 trials per positively rewarded trial type, since it was difficult to achieve a reliable acquisition of many more trials than 15 due to the larger reward volumes used in this task. Once they reached criterion, animals were implanted with a headstage containing an array with individually movable tetrodes (“hyperdrive”) and recordings started. In the first recording session the set of odors that was used during shaping was presented again, in order to retrain the animals with familiar odors. Provided that the performance was back at presurgical level, a new set of five odors was given the next session. Each time the discrimination had been learned, as visible by withholding responses towards or at the fluid well during quinine or non-reinforced trial types (usually within 1 session), a new set of five odors was presented.

During the task, odors were presented in a pseudo-random order. After the trial light switched on the animal had 15 s to make an odor poke. If no odor poke was made, the light turned off and an inter-trial interval (with a variable duration of 10-25 s) started. Whenever an odor poke was made, the trial light switched off after 500 ms, followed 500 ms later by the presentation of an odor. After retraction of the animal’s nose out of the odor sampling port or whenever a maximal duration for odor sampling (10 s) was exceeded, odor presentation was terminated. Premature retraction from the odor sampling port (odor poke shorter than the minimal duration of 2 s) resulted in the inter-trial interval. Whenever the animal received reinforcement in the fluid well it had to consume the fluid within 10 s, after which the fluid well was drained by a vacuum line and the inter-trial interval started. An incorrect (‘go’) response after sampling an odor predictive of quinine or the non-rewarded contingency had no further programmed consequences.

**Surgery and electrophysiology**

Animals were anaesthetized with 0.08 ml/100 g Hypnorm i.m. (0.2 mg/ml fentanyl, 10 mg/ml fluanison) and 0.04 ml/100 g Dormicum s.c. (midazolam 1
mg/kg) and mounted in a Kopf stereotaxic frame. Additional local anaesthesia (Xylocaine spray; 10%, Astra) was also given. Body temperature was maintained at 37.5 °C using a heating pad. After exposure of the cranium, five small holes were drilled into the cranium to accommodate surgical screws, one of which serving as ground. Another larger hole was drilled over the OFC in the left hemisphere (3.2 mm anterior, 3.2 mm lateral to bregma according to Paxinos and Watson, 1996). After opening of the dura the base of a hyperdrive (an array of 12 individually drivable tetrodes (13 µm nichrome wire; Kanthal, Palm Coast, Florida) and 2 reference electrodes, spaced apart by at least 310 µm) was lowered onto the exposed cortex (Wilson and McNaughton, 1993; Gray et al., 1995; Pennartz et al., 2004). The hyperdrive was anchored to the screws with dental cement. Immediately after surgery all tetrodes were advanced 1 mm into the brain; in the course of the next three days the tetrodes were gradually lowered until the OFC was reached. Animals were allowed to recover at least 7 days before re-exposure to the behavioral paradigm and the start of the recordings. Recording sessions were initiated when all tetrodes were estimated to have reached the OFC and the animal had obtained its presurgical performance level. In order to record different units during each recording session, all tetrodes were lowered at the start of a recording day with increments of 40 µm. Depending on the amount of neural activity, individual tetrodes could be advanced further. Once the tetrodes were lowered the animal was left to rest in his home cage for at least 2 hours in view of unit recording stability, after which the experimental session started.

Electrophysiological recordings were performed using a Cheetah recording system (Neuralynx, Tucson, Arizona). Signals from the individual leads of the tetrodes were passed through a low noise unity-gain field-effect transistor preamplifier, insulated multi-wire cables and a 72 channel commutator (Dragonfly, Ridgeley, West Virginia) to digitally programmable amplifiers (gain 5000 times; band-pass filtering 0.6-6.0 KHz). Amplifier output was digitized at 32 KHz and stored on a Windows NT station. A 1 ms data sample was taken whenever the signal crossed a preset voltage boundary, so that the width of a window containing the recorded spike was captured in 32 data points. The occurrence of task events in the behavioral chamber was recorded simultaneously, and the behavior of the animals was recorded on videotape.

Data analysis

Single units were isolated by off-line cluster cutting procedures (BBClust/MClust 3.0). Before a cluster of spikes was accepted as a single unit, several parameters and graphs were checked visually, namely the averaged waveform across the four leads, the cluster plots showing spike parameter distributions such as peak amplitudes across the four dimensions, the autocorrelogram and the spike
interval histogram. Since the absence of spike activity during the refractory period (2-3 ms) is indicative for good isolation, units of which the autocorrelogram and the spike interval histogram revealed activity during this period were removed from the analysis.

Correlations between events in the task and changes in firing rate were examined by constructing peri-event time histograms (PETHs) and statistically assessed with the non-parametric Wilcoxon matched-pairs signed-rank (WMPSR) test (P < 0.01) with bin resolutions of 100 and 1500 ms. The 1500 ms bin resolution was used to examine the significance of neural correlates during broader periods such as the entire waiting period of 1500 ms, whereas the 100 ms bin resolution was used to statistically examine the more exact time course of the response. Neural responses were considered significant if firing rates, quantified per bin, were significantly different from a control (baseline) period during the inter-trial interval. This control period consisted of five bins, and any of the bins tested for a significant change in firing during the trial was required to differ significantly from each of these five control bins. In addition, responses had to be significant for both bin resolutions to be considered as such. This procedure enhanced conservativeness in the identification of neural correlates of task events, and, by using a non-parametric test, avoids a number of difficulties and assumptions associated with some other tests that are related to the nature of spike timing distributions and their deviations from baseline activity. Once the WMPSR test indicated a significant deviation in firing rate with respect to baseline, the non-parametric Kruskall-Wallis test (p < 0.05), followed by a post-hoc Mann-Whitney U test (P < 0.05) were used to assess response differences between PETHs pertaining to different odor-reward magnitude pairs.

Behavioral data was analyzed using SPSS for Windows (version 12.0.1). Unless otherwise stated, results are expressed as mean ± SEM values. Movement Time was defined as the interval between nose retraction from the odor port and nose entry into the fluid well, whereas the Overall Response Time was defined as the duration of the entire sequence of odor sampling and moving to the fluid well. The mean response times per reward magnitude were obtained from all different trial types within all the sessions that were used for analysis. These measures were compared across different trial types with the non-parametric Kruskall-Wallis test (P < 0.05), followed by a post-hoc Mann-Whitney U test (P < 0.05).

**Histology**

The final position of the tetrodes was marked by passing a 10 s, 25 µA current through one of the leads of each tetrode in order to induce a lesion and initiate gliosis. The next day, approximately 24 hours after lesioning, animals were perfused transcardially using a 0.9% saline solution followed by 10% formalin.
After removal from the skull the brain was stored in a 10% formalin solution for several days before sectioning. Brain sections (40 µm) were cut using a vibratome and Nissl-stained to reconstruct the tetrode tracks and their final position.

Results

Behavior

For the analysis, data was used from 24 recording sessions, obtained from 7 rats. Animals needed on average about 17 trials for each of the different positively reinforced trial types to reach the criterion of 15 successful trials per reward size; there was no difference in performance level between the different amounts of reward (number of trials needed for 0.05 ml: 16.6 ± 0.4; 0.15 ml: 16.9 ± 0.5; 0.30 ml: 17.4 ± 0.6). Rarely the number of 15 trials per reward size was not reached during a session. For the non-reinforced and quinine condition animals made on average 5.8 ± 0.7 and 4.5 ± 0.7 ‘go’ responses, respectively (sucrose versus quinine or versus non-reinforced: \( P = 0.000 \); paired sampled t-test). In the course of a session these erroneous ‘go’ responses (false alarms) were followed by trials in which animals, after sampling odors predictive for these outcomes, began to withhold responses towards or at the fluid well. Withholding was typically followed by odor pokes with durations shorter than the required 2 s, thus yielding non-valid trial types. This indicates that the animals were actually capable of associating an odor with either a positive or negative outcome. Since the minimal number of trials accepted for the analysis of the electrophysiological data was 6, neural responses to the various stimuli applied during the non-rewarded trial type were not examined any further. However, sessions in which 6 or more trials for quinine were present (n = 9) were used to determine whether neural responses showing reward magnitude differences were compatible with response patterns for quinine (see below). This was done by comparing neural responses during the quinine trials with responses during the trial type with the same amount of sucrose (0.15 ml).

Examination of the Movement Time showed no significant differences between the positively reinforced trial types (0.05 ml: 1.47 ± 0.04 s, 0.15 ml: 1.36 ± 0.03 s and 0.30 ml: 1.37 ± 0.04 s). A comparison between these trial types and the non-reinforced and quinine trials revealed a significant difference: Movement Time during positively reinforced trials taken together (1.40 ± 0.02 s; sample sizes per reward magnitude: 0.05 ml: n = 336; 0.15 ml: n = 343; 0.30 ml: n = 356) was significantly shorter compared to non-rewarded and quinine trial types (non-rewarded ‘go’ responses: 1.50 ± 0.06 s; n = 138; quinine ‘go’ responses: 1.74 ± 0.13 s; n = 107). However, when the entire sequence of odor sampling and moving to the fluid well was taken into consideration (Overall Response Time), rats performed
significantly faster for the 0.15 and 0.30 ml trial types (respectively 3.54 ± 0.06 s and 3.55 ± 0.05 s) as compared to the 0.05 ml sucrose trials (3.74 ± 0.06 s).

**Histology**

Histological verification of the tetrode positions (Fig. 1) showed that the recording sites ranged between 2.7 mm and 4.2 mm anterior to bregma and were limited to the ventral and lateral orbital regions of the OFC. Recording depth ranged from approximately 3 mm to 5.5 mm below cortical surface (Paxinos and Watson, 1996).

![Figure 1](image)

Figure 1. Representative histological section showing the localization of tetrode recording sites. Indicated by the black arrow heads in the upper part of the section are sites at which tetrodes have entered the brain. Several tracks are partially visible, including one track with an endpoint marked by a lesion, indicated by the asterisk. Recordings in all rats were localized in the ventral and lateral regions of the OFC (areas VO and LO), between 2.7 and 4.2 mm anterior from bregma (Paxinos and Watson, 1996).

**Electrophysiology**

**General overview**

During the 24 recording sessions, mainly performed on consecutive days, a total number of 894 single units was recorded in the OFC. The number of single units recorded per session ranged from 24 to 61 (mean ± SEM: 38.9 ± 2.2), with mean firing rates ranging from 0.06 to 48.4 spikes/s. Of these 894 neurons, 141 (16%) showed 176 statistically significant changes in firing rate correlated to one or more main events in the task (Table 1). These neural correlates consisted of responses during the following four events or phases: odor sampling, behavioral activity preceding the nose entry into the fluid well, the waiting period and the
delivery of reinforcement (Fig. 2). The remaining neurons failed to demonstrate any significant task-related modulation as revealed by the statistically assessed histograms and were not examined further. For the analysis of neural correlates only positively rewarded trial types were taken into consideration.

Figure 2. Overview of behavioral correlates observed during task performance. Peri-event time histograms and raster plots showing examples of all main behavioral correlates observed during the task. Examples from four different units recorded in four rats demonstrating correlates related to (A) odor sampling, synchronized on onset of odor presentation, (B) movement activity preceding nose entry into the fluid well, synchronized on fluid well entry, (C) the waiting period of 1500 ms in the fluid well, synchronized on onset of waiting and (D) delivery of reinforcement, synchronized on reward delivery onset. These histograms, as well as the following histograms (figures 3, 5 and 6), are presented with a bin size of 100 ms. In all raster plots, individual consecutive trials are represented as horizontal lines, with the first trial at the top row. Horizontal scales show time (s), vertical scales firing rate (Hz).

During odor sampling, a total of 44 neural responses were found (25%), all consisting of an increase in firing (Fig. 2A). Many of the cells showing this correlate (45.5%, n = 20) started firing when the nose was already in the sampling port but before odor was presented, activity that might reflect the anticipation of odor delivery or the onset of behaviors related to odor sampling. All responses during odor sampling typically peaked within 1500 ms after odor onset. Because the task was not designed to determine whether differential responses during odor sampling were due to different sensory inputs (odor identity) or to expectancy for different
reward magnitudes, the results are inconclusive regarding the coding of expected reward magnitude during this phase. Therefore, the responses observed during odor sampling were not analyzed further.

In ‘go’ trials, the presentation of the odor stimulus was followed by the animal’s movement from the sampling port towards the fluid well and the subsequent waiting period in the fluid well before reinforcement delivery. In both phases, neurons demonstrated significant changes in firing. In the task phase where the animal moved from the sampling port to the fluid well, 33 correlates (19%), consisting of an enhancement in firing rate were found (Fig. 2B). During the waiting period, 47 responses (27%) were found, one of which showing a decrease in firing activity, whereas the other 46 consisted of an enhancement in firing (Fig. 2C).

During the reinforcement delivery phase, all animals displayed similar behavior at the fluid well: for every amount of sucrose solution their snout remained in the fluid well until the end of trial. In response to positive reinforcement, 52 correlates were found (29%), consisting of five different response types. The first type of response consisted of a transient increase in firing rate, peaking within 500 ms after reinforcement delivery (27%, \( n = 14 \)) (Fig. 2D), the second type consisted of an increase in firing rate that peaked between 500 and 1500 ms after delivery (31%, \( n = 16 \)). The third type of response consisted of an increase in firing rate starting almost immediately after the delivery of the reward, remaining during the entire period the animal visited the fluid well (15%, \( n = 8 \)). The fourth type was a rarely encountered transient decrement in firing rate (8%, \( n = 4 \)), whereas the fifth type comprised all other responses showing enhancements in firing rate starting at least 3 s after the delivery of the reward and with variable duration (19%, \( n = 10 \)).

**Neural activity during the movement and waiting period**

After determining the various neural responses to task events, the question arose whether expectancy for the upcoming reinforcement was represented within these different types of neural correlates. Given the absence of changes in sensory input and the animal’s overall immobility, neurons that showed responses within the waiting period in the fluid well may code the predicted outcome. However, to exclude the possibility that these responses were due to anticipatory licking of the animal instead of expected outcome per se, neural responses (\( n = 22 \)) recorded during sessions in which lick detection was available were examined in relation to the extensions and retractions of the tongue in the fluid well. This showed that these responses did not covary with licking behavior, since in these cases anticipatory licking was either completely absent, started after the neural response, or, when overlapping in time with the neural response, continued during reward consumption, when the neural response had already ended.
TABLE 1: Numbers of behavioral correlates.

<table>
<thead>
<tr>
<th></th>
<th>Discriminating Between 2 Sizes</th>
<th>Discriminating Between 3 Sizes</th>
<th>Total Discriminating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odor sampling alone</td>
<td>44</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>&amp; movement period</td>
<td>35</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>&amp; waiting period</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>&amp; reward delivery</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Movement Period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alone</td>
<td>33</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>&amp; odor sampling</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>&amp; waiting period</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp; reward delivery</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Waiting Period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alone</td>
<td>21</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>&amp; odor sampling</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>&amp; movement period</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Reward delivery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alone</td>
<td>26 (19)</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>&amp; odor sampling</td>
<td>2 (1)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>&amp; movement period</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp; waiting period</td>
<td>23 (21)</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>176 (122)</td>
<td>44</td>
<td>23</td>
</tr>
</tbody>
</table>

Overview of the amount of significant behavioral correlates observed during the task in a population of 141 units. Only correlates for the positively rewarded trial types were taken into consideration for calculating the overall number of responses. Responses indicated by 'alone' occurred only to the specific event indicated on the line above, the other responses correlated to two events ('dual correlates'). Units with dual correlates account for the fact that the total number of correlates (176) is higher than the total number of units (141). Responses during odor sampling were not examined in relation to reward magnitude (indicated by '-'; see text). Numbers of correlates after reward delivery between brackets indicate the number of cells that were used to examine differential firing to the various reward sizes (the fifth type of reward responses was left out of the analysis, see text).

Furthermore, a comparison between neural responses in this period between (incorrect) 'go' responses during quinine trials versus 'go' responses during sucrose trials revealed that 10 out of 11 neurons recorded showed differential firing for the sign of the response outcome during this period. Therefore, the present data are
consistent with neural coding of reward expectancy in a subset of OFC neurons during the waiting period.

Expectancy for reward, however, might also occur earlier in the task sequence, for example before or during the movement period of the ‘go’ response. To examine whether neural responses observed during the movement phase were related to the ‘go’ movements of the animal per se or whether they reflected a truly goal-directed action in the context of the task, neural responses obtained during the correct ‘go’ trials were compared with those during performance of the same behavior within the inter-trial interval. This showed that these correlates were task-dependent for 30 out of the 33 neurons exhibiting significant responses in this period: they were not observed when the behavioral sequence (‘leave odor port and go to reward site’) was performed during the inter-trial interval (Fig. 3). In 3 units a significant response occurred during the inter-trial interval at a similar time point as during correct ‘go’ trials, but these activations were not as strong as the responses observed during trials. In addition, a comparison between neural responses during (incorrect) ‘go’ responses during quinine trials versus ‘go’ responses during sucrose trials revealed that during the movement period 12 out of 14 neurons recorded demonstrated differential firing depending on the response outcome.

**Neural activity related to different magnitudes of reward**

Whether neurons showed significant differences in response to the three different reward magnitudes was examined for three phases, namely for the responses occurring during the movement period preceding the fluid poke, the waiting period and the period starting with reward delivery. A total number of 67 of the 122 correlates (55%) found during these three task phases demonstrated statistically significant differences between different amounts of reward, which was either between two (66%) or all three (34%) different reward sizes. The proportion of neurons showing differential activity varied for the three task periods and ranged from 45 to 69% (Table 1). Within the group of 33 neurons demonstrating task-specific neural correlates during the movement period preceding the fluid poke, 17 neurons (52%) demonstrated statistically significant differential firing across different reward sizes, with 7 and 10 neurons discriminating between 2 and 3 different sizes, respectively (Fig. 3). Some neurons displayed firing activity that increased with incremental reward size (18%, n = 3), whereas other neurons displayed the inverse relationship (29%, n 5) or demonstrated the largest or smallest response to the middle reward size (53%, n = 9; Fig. 4A).
Figure 3. Differential firing in relation to expected reward magnitude during the movement period. A. Example of one unit demonstrating an increase in firing rate after leaving the odor port and before nose entry into the fluid well. Activity is synchronized on nose entry into the fluid well, grey marks indicate offset of the odor poke. This single unit discriminated between the two smallest and the largest amount of reward; there was no significant difference in firing between the two lowest reward sizes, but both responses were significantly higher compared to the largest amount of reward. B. Activity of the same unit during the performance of the same behavioral sequence as in A, but now in the inter-trial interval (ITI).

During the waiting period, in which 47 neurons showed significant responses, 21 units (45%) discriminated significantly between different magnitudes of reward (Fig. 5). A total number of 18 of these 21 neurons showed differential activity between two different reward sizes, 3 neurons between three sizes. A similar heterogeneity within the tuning of the neurons was found as during the movement period: units showed a rise in firing activity with increasing reward size (29%, n = 6), whereas other neurons displayed the inverse relationship (24%, n = 5) or showed the largest or smallest activation to the middle reward size (48%, n = 10; Fig. 4B).

Besides expressing neural correlates related to reward magnitude in the periods preceding reinforcement delivery, OFC neurons may code aspects related to the actual reward magnitude. After examining whether neural responses (n = 10) in the period of reward delivery were related to the tongue movements of the animal during licking, none of the responses could be ascribed to licking behavior. Furthermore, due to the different nature of the responses observed in this period, determining whether neurons demonstrated discriminatory responses to the different
Figure 4. Overview of the response profiles to reward size during different task phases, i.e. responses in the movement period (A), the waiting period prior to reinforcement delivery (B) and reward delivery (C). Only profiles of units with single behavioral correlates ('alone' in table 1) and with significant differences between at least two reward sizes are shown. In each graph, different units are represented by different symbols. Some neurons changed their firing pattern monotonically with increasing or decreasing reward size, whereas others showed the largest or smallest response to the middle reward size. On the horizontal scale reward size (ml) is plotted; the vertical scale displays the peak firing rate of individual units in association with different reward sizes, normalized to the response to the 0.05 ml reward.

The magnitudes of reward was done using different criteria for the various response types. For the transient response types, only the time periods in which the responses occurred were compared; for the sustained responses, the entire period with enhanced firing was tested. Responses of the fifth type were left out of the analysis due to the variability of the responses within this group of neurons. Of the 42 neurons showing responses of the remaining four types, 29 neurons (69%) showed significant differential firing across reward sizes, with respectively 19 and 10 neurons discriminating between two and three different reward sizes (Fig. 6). Again all types of response profiles were demonstrated: increasing (21%, n = 6) or decreasing (14%, n = 4) firing activity with increasing reward size, or the largest or smallest response to the middle reward size (66%, n = 19; Fig. 4C).
Figure 5. Differential firing in relation to expected reward magnitude during the waiting period. Example of one single unit. Activity, synchronized on nose entry into the fluid well, did not differ between the two lowest amounts of reward but both differed significantly from the largest amount of reward. Highest activity was demonstrated for the smaller reward sizes. Reinforcement delivery at 1500 ms is indicated by the vertically aligned grey marks.

An additional Kolmogorov-Smirnov test (p < 0.05) was performed for the responses that peaked between 500 and 1500 ms (fig 6A; second response type), since this subgroup often showed a response time course varying with reward magnitude. Taking firing rate values for successive bins as variable, this test revealed that for all neurons of this second type (n = 16) the time course of the response differed either between two (53%) or three (47%) reward sizes as well.

Discussion

The present experiment was performed to examine whether and how information concerning reward magnitude is coded in rat OFC. A total number of 141 out of 894 neurons recorded (16%) demonstrated a behavioral correlate of firing activity during several events in the task, i.e. odor sampling, the movement
period from odor port to the fluid delivery well, the waiting period in the fluid well and the reward delivery period. During the movement and waiting period, respectively 52% and 45% of the neurons showing a significant behavioral correlate demonstrated differential firing across varying reward sizes. After reward delivery, 69% of the neurons showed differential responses towards the various reward sizes. Overall, the percentage of neurons showing significant correlations during this olfactory discrimination task (16%) was lower than in a previous study with an

Figure 6. Differential firing to reward magnitude after reward delivery. Examples of two different units showing differential firing towards the different reward sizes after reward delivery. A. On the left a unit is shown demonstrating firing that differed significantly between the smallest and the two largest rewards, with the largest response for the larger reward sizes. B. On the right a different unit is shown demonstrating differential activity between the two smallest and the largest amount of reward, with the largest response for the lower reward sizes.
eighth-odor olfactory discrimination task (Schoenbaum and Eichenbaum, 1995). Apart from differences in the behavioral paradigms, this might be due to the conservative statistical assessment in the present study, to methodological differences in unit recording and isolation, or to the fact that recording sites in the latter study were located more laterally in the OFC and also comprised the ventral part of the agranular insular cortex, whereas in the present study locations of recordings were limited to the ventral and lateral orbital regions.

Testing the ability of animals in a T-maze task to discriminate between the different reward sizes used in this behavioral task demonstrated that animals were able to discriminate the various reward sizes. In the 5-odor discrimination task, the scores on Overall Response Time revealed that animals responded significantly slower for the smallest amount of reward, indicating that animals learned to discriminate at least between the smallest and the two larger amounts of reward. The finding that this difference was not present within the Movement Time might be explained by the limited number of trials. The differences could have turned out significant when the number of trials per session had been larger. However, other possible explanations include a higher degree of stereotypy or habit-based performance of ‘go’ responses, once initiated, than applies to the odor sampling phase. That the Movement Time was highly similar across different reward sizes makes it likely that the discriminatory neural responses during the movement period were not due to differences in motor-related aspects, including the speed of responding.

**Neural responses during the reward delivery phase: magnitude effects**

As concerns actual reward magnitude, the results indicate that neurons in rat OFC may indeed code information related to this reinforcement parameter, showing differential firing across various reward sizes during reward consumption. Although this discriminative firing appears not to be due to licking behavior, there is the possibility of modulation of these responses by sensory-motor processes (e.g. tasting or ingestion), in so far this is distinguishable from reward processing per se. The observation that the OFC is involved in the processing of information related to the magnitude of the actual reinforcement is in line with the fMRI result in humans that distinct regions within the OFC represent the magnitude of monetary reward (O’Doherty et al., 2001). Furthermore, it fits in with earlier findings concerning neural coding of other aspects of reinforcers in the OFC, such as qualitative properties or motivational value (Thorpe et al., 1983; Tremblay and Schultz, 1999; Hikosaka and Watanabe, 2000; Wallis and Miller, 2003; Roesch and Olson, 2004; Ichihara-Takeda and Funahashi, 2006; Padoa-Schioppa and Assad, 2006). Together with the current result, this implies that at least some major aspects of reinforcers
that might be influential for the selection of the most appropriate response strategy are represented within the OFC.

Neural coding of reward expectancy

In a previous study by Schoenbaum et al. (1998), neurons in the OFC were suggested to code expected outcomes, since they demonstrated discriminatory firing during the delay period in which animals awaited either positive or negative reinforcement, which is a result confirmed in this study. However, a confounding factor in the study by Schoenbaum et al. was the comparison between ‘go’ responses for both positive and negative reinforcers, the latter being erroneously made. The possibility of such a confounding factor is underscored by a study of Ramus and Eichenbaum (2000) on olfactory recognition memory, who reported units in rat orbitofrontal cortex showing firing rate changes during odor sampling that occurred specifically prior to erroneous ‘go’ responses. The advantage of the parametric design in the current study for the examination of reward expectancy is that very similar ‘go’ responses can be compared, since the impending outcomes are all appetitive. Indeed, we found no significant difference in Movement Time across the three positively rewarded trial types, while the waiting period was fixed in duration (1.5 s). Thus, the current design may provide more certainty about the nature of the observed neural correlate, since no change or switch in response strategy is required during these trials. Nonetheless, it remains possible that other cognitive processes besides reward expectancy, like working memory or attention, although not obviously required for correct task performance, are reflected within the waiting or movement period. An additional remark is that, although learning of the discrimination problem by the animals was obvious from both ‘go/no-go’ decisions and the Movement and Overall Response Time, it would be difficult to actually show learning-related changes in single-unit firing. Animals learned to discriminate positively versus negatively reinforced odors within a few trials (average for quinine trials: 4.5 ± 0.7), and the number of spikes during the waiting period per trial appears insufficient to permit a robust analysis of the evolution of the neural representation of stimulus-reward contingencies across these few trials. Resolving this issue will necessitate a new study in which learning is temporally spread out across more trials. Despite this, it does not seem likely that the reward-size differences in neural activity would be attributable to carry-over effects from previous sessions. New odors were used in every new session, and care was taken to avoid constant reinforcement contingencies of the odor families used here. Secondly, interference by previously established contingencies would be expected to prevent the emergence of a consistently slower response to the odor predicting the lowest amount of reward as compared to the high-reward contingencies.
**Reward expectancy: timing and magnitude effects**

Discriminatory activations towards different magnitudes of reward were found to occur during the waiting period, as well as during the movement period, the latter likely being an expectancy-related correlate during the ‘go’ response that to our knowledge has not been described earlier in rat or primate OFC. In addition to the rat OFC, expectancy-related activity was previously found in rat basolateral amygdala (BLA), with neurons demonstrating selective firing during a delay period prior to reinforcement or during predictive odor cues after learning (Schoenbaum et al., 1998, 1999; see Holland and Gallagher, 2004 for review). In addition to these two types of expectancy-related activity, the results of the present study indicate that, at least in the OFC, reward expectancy may also be coded during the movement phase of the ‘go’ response. Although this neural correlate occurred during motor behavior, it was absent when the animal executed the same movement outside the context of the trials, thus excluding a pure motor confound. The presence of this expectancy-related signal matches the results from lesion studies indicating the involvement of the OFC in the flexible guidance of goal-directed behavior (Dias et al., 1997; Meunier et al., 1997; Hornak et al., 2004; Izquierdo et al., 2004), and is also consistent with a role for OFC in encoding motor set as well as action-outcome relationships (Baxter et al., 2000). Especially when reward contingencies are changing, information representing these relationships should be available during task performance, particularly before and during the execution of a behavioral response. The idea that a reward expectancy signal is available during ongoing behavior is furthermore supported by the finding that neurons in the primate OFC may code both long- and short-range reward expectancy (Hikosaka and Watanabe, 2004). It is suggested by these authors that during behavior these different OFC signals serve to adjust motivational levels across different temporal ranges, namely towards an immediate or a more distant outcome. Altogether, the available results support the notion that reinforcement parameters such as magnitude and quality are coded across different temporal phases along task trials, corroborating the hypothesis that OFC neurons collectively code a matrix of reward parameters as a function of the delay towards the moment of outcome. Future recordings will be needed to study the function of neurons with expectancy-related firing during movement periods in more detail, e.g. during different motor behaviors or reversal learning.

A previous study by Roesch et al. (2006) already demonstrated neurons within rat OFC discriminating between two different reward sizes. However, apart from the fact that reward size was not parametrically varied (only two different reward sizes were used), their results are not directly comparable with the results from the current study. Additional differences concern the method of reward application and location of reward delivery. Roesch et al. applies a single bolus of 0.05 ml as a small reward and an additional drop 500 ms after the first bolus as the
large reward, with different reward sizes administered at different locations. In contrast, in the present study all reward sizes were delivered in a single dose and at a common location.

The finding of the current study that information concerning the magnitude of reward is represented in rat OFC is in line with one of the predictions from the HSAT (Hebbian Synapses with Adaptive Thresholds) model of RL (Pennartz, 1997), in which the OFC, together with other related limbic areas, functions as a brain centre for processing and predicting reinforcement, as well as directing Hebbian changes in synaptic strength in brain areas targeted by the OFC. The presence of information comparing the actual versus the expected magnitude of reinforcers within the OFC was modeled to direct behavior of a neural network performing parallel distributed processing towards the most profitable outcome in an efficient and flexible manner. In particular, the observation that about one-half of the OFC neurons (23 out of 52) demonstrating responses during reward delivery also showed expectancy-related activity during the waiting period, is in line with the model, since in the course of learning neurons within the reinforcement processing module of the HSAT network that are initially activated by the reward will become activated during the expectancy period preceding reward as well. Besides the OFC, other brain areas emitting glutamatergic fibers may serve as candidate regions for the reinforcement processing posited by this model, such as medial prefrontal cortex (Pratt and Mizumori, 2001), anterior cingulate cortex and dorsolateral prefrontal cortex (Watanabe, 1996; Leon and Shadlen, 1999; Tremblay and Schulz, 1999; Shidara and Richmond, 2002). That the OFC neurons in the current study exhibited a variety of different tuning curves to reward magnitude – either during the expectancy phase or actual reward delivery – constitutes a result that was not directly predicted by the computational model. This finding is in line with results of Wallis and Miller (2003), demonstrating that neurons within monkey OFC code expected reward magnitude both parametrically and non-parametrically. Such a diversity of tuning is as compatible with coding in a parallel-distributed network as is a monotonic relationship between firing activity and reward magnitude. Adopting a wider perspective on RL than prescribed by any particular model, it should be mentioned that reward variables, including magnitude, are also represented in the striatum (Hollerman et al., 1998; Hassani et al., 2001; Cromwell and Schulz, 2003). Although recent results by Pasupathy and Miller (2005) showed that learning-related changes in firing activity in a cue-saccade association task occurred earlier in striatum than prefrontal cortex, the actual anatomical loci mediating stimulus-response and action-outcome learning remain to be firmly established. The distribution of the information across several cortical and subcortical areas suggests a broader computational system by which RL can be mediated than a basic actor-critic
architecture trained by Temporal Difference Learning that is composed of the striatum and mesencephalic dopamine system (see for review Montague et al., 2004).

**Acknowledgements**

This work was supported by NWO Grant 903-47-084, NWO grant 918.46.609 and BSIK (SenterNovem) grant 03053. We would like to thank Geoffrey Schoenbaum for providing information about the behavioral set-up, Bruce McNaughton for his help with the use of tetrode arrays and David Redish and Peter Lipa for providing the cluster cutting software. Furthermore, we thank Eunjeong Lee and Ton Put for their contribution to respectively the data analysis and graphical illustrations, and our colleagues at the electronic and mechanical workshop of the Netherlands Institute for Neurosciences for their excellent technical assistance.
Chapter 3

Population coding of reward magnitude in the orbitofrontal cortex of the rat
Abstract

Although single cell coding of reward-related information in the orbitofrontal cortex (OFC) has been characterized to some extent, much less is known about the coding properties of orbitofrontal ensembles. We examined population coding of reward magnitude by performing ensemble recordings in rat OFC while animals learned an olfactory discrimination task in which various reinforcers were associated with predictive odor stimuli. Ensemble activity was found to represent information about reward magnitude during several trial phases, namely when animals moved to the reward site, anticipated reward during an immobile period and when they received it. During the anticipation phase, Bayesian and template matching reconstruction algorithms decoded reward size correctly from the population activity significantly above chance level (highest value of 43 and 48%, respectively; chance level 33.3%), whereas decoding performance for the reward delivery phase was 76 and 79%, respectively. In the anticipation phase, the decoding score was only weakly dependent on the size of the neuronal group participating in reconstruction, consistent with a redundant, distributed representation of reward information. In contrast, decoding was specific for temporal segments within the structure of a trial. Decoding performance steeply increased across the first few trials for every rewarded odor, an effect that could not be explained by a nonspecific drift in response strength across trials. Finally, when population responses to a negative reinforcer (quinine) were compared to sucrose reinforcement, coding in the delivery phase appeared to be related to reward quality, and thus was not based on ingested liquid volume.
Introduction

The orbitofrontal cortex (OFC) is thought to contribute to the guidance of goal-directed behavior through the formation of neural representations of predicted outcomes. Studies examining single unit activity in OFC during operant behavior demonstrated that firing rates are modulated by the motivational value of stimuli and represent reward-predictive information during anticipation of various types or amounts of reward (Hikosaka and Watanabe, 2000; Ichihara-Takeda and Funahashi, 2006; Lipton et al., 1999; Padoa-Schioppa and Assad, 2006; Ramus et al., 2007; Roesch and Olson, 2004; Roesch et al., 2006; Schoenbaum et al., 1999, 2003; Simmons and Richmond, 2008; Thorpe et al., 1983; Tremblay and Schultz, 1999; Wallis and Miller, 2003; Yonemori et al., 2000). These single cell studies have not shown how predicted or actual rewards are dynamically represented by populations in OFC. This question is especially relevant for understanding how other, connected brain areas may read out population activity from this structure (Wu and Amari, 2005; Pouget et al., 2000). Until now, population coding analyses have been mainly used for reconstruction of arm movement direction during a reaching task (e.g. Georgopoulos et al., 1986), reconstruction of an animal’s environmental location (e.g. Wilson and McNaughton, 1993), and to predict behavioral responses (e.g. Baeg et al., 2003; Laubach et al., 2000). Only a few studies have thus far described ensemble activity within OFC. Gutierrez et al. (2005) found that ensemble activity in rat OFC discriminated between sucrose versus water reward, but this distinction was made in a free-licking situation, thus without predictive cues and a contingent operant response. In a sensory discrimination task in rats, Schoenbaum and Eichenbaum (1995b) showed that, during stimulus sampling, OFC population activity represented expectation of a reward presented in the following trial. However, it remains unknown how actual and predicted rewards are represented in the population within a specific trial phase during an operant task, and whether such a population code would be specific for different trial phases. Here it is of special interest to compare representations when the animal anticipates reinforcement or when he receives it. For the reinforcer consumption phase, we also asked whether the observed ensemble coding is related to reward quality or can be explained by coding of the ingested volume of liquid reinforcer. In addition, we examined whether reward magnitude is represented in OFC in a sparse or redundant manner, i.e. by a few highly specifically tuned cells or in a broadly distributed way (cf. Narayanan et al., 2005). Finally, we addressed the dynamics by which consistent coding develops as the learning task progresses. It has been hardly feasible to examine learning-related changes on a trial-by-trial basis in single unit recordings from OFC, because single-unit firing patterns show great variability from trial to trial, making it difficult to track systematic changes during learning. To address these various aspects of
population coding, ensemble recordings were made from rat OFC during a five-odor discrimination ‘go/no-go’ task, in which odors were predictive for various amounts of an appetitive sucrose solution, for no reward or an aversive quinine solution.

Materials and Methods

Subjects

All experiments were approved by the Animal Experimentation Committee of the Royal Netherlands Academy of Arts and Sciences and were carried out in accordance with the National Guidelines for Animal Experimentation. Data were collected in 17 sessions recorded from 6 male Wistar rats (Harlan CPB; Horst, The Netherlands; sessions partially overlapped with those presented in a single unit study, Van Duuren et al. 2007). Animals were weighed and handled daily and socially housed in standard type 4 macrolon cages under a reversed 12 hr light/dark cycle (dimmed red light at 7:00 AM). Weight at the time of surgery was 325-425 g. After surgery, animals were housed individually in a larger cage (1 x 1 x 1 m) under similar conditions. Food was available ad libitum, but animals were water restricted. They had access to water with a variable delay after the end of the recording session in the home cage for a 2 hour period.

Behavior

Apparatus

Recordings were made in a black plexiglas operant chamber (40 x 37 x 41.5 cm), placed in a sound attenuated and electrically shielded box. The front panel contained on the right side a light signaling trial onset and an odor sampling port, and on the left side a delivery well for fluid reinforcement. Responses made by the animal in the sampling port and fluid well were registered by an infrared beam transmitter and detector. Behavioral events during task performance and data collection were controlled by a computer. Odor delivery was controlled by a system of solenoid valves and flow meters. To prevent mixing up of odors within the system, separate delivery lines for each odor were present. The different types of fluid reinforcement were delivered with separate fluid lines. Fluid delivery was gravity-driven, with a tap and valves controlling the flow and amount of fluid delivered. The odorants (Tokos BV; Noordscheschut, The Netherlands) were separated into different families, i.e. woody, fruity, herbal, citrus and floral. Each set of odors used in a discrimination session contained one odor from each family. In addition, no single family of odors was preferentially associated with a particular trial outcome.
Behavioral paradigm

When animals were habituated to the recording chamber, the training on the behavioral procedure of the five-odor olfactory discrimination 'go-no/go' task started. In this task, odors signaled whether a 'go' response resulted in a particular amount of a positive reinforcement, or in a negative one. We used five different odors in each session, each of which was uniquely predictive throughout the trial of either an amount of an appetitive sucrose solution (10% sucrose in water, i.e. 0.05, 0.15 and 0.30 ml), no reward (non reinforced-condition), or an aversive quinine solution (0.15 ml of a 0.015 M quinine in water). A particular odor was never used in more than one session. Criterion for behavioral performance was set at 15 trials per positively rewarded trial type, because it was difficult to reliably obtain more trials due to the larger reward volumes. When animals reached this criterion during training, they were implanted with a multi-tetrode array ("hyperdrive") and recordings started. For each recording session a new set of five odors was used. During the task, in which odors were presented pseudo-randomly, the onset of the trial light indicated that animals could initiate a trial by making an odor poke. If no odor poke was made within 15 s after onset of the light, the light switched off and the inter-trial interval started (with a variable duration of 10-25 s). After initiation of an odor poke the trial light switched off after 500 ms, followed 500 ms later by the odor presentation. This interval was included to prevent the animal from moving during cue sampling. Odor sampling itself was required to last at least 1 s. After retraction from the odor sampling port or whenever a maximal duration for odor sampling (10 s) was exceeded, odor presentation was terminated. Premature retraction from the odor sampling port (odor poke shorter than the minimal duration of 2 s) resulted in the onset of the inter-trial interval. The nose poke in the fluid well marked the onset of an immobile waiting period of 1.5 s after which reinforcement was delivered. Whenever reinforcement was delivered, animals had 10 s to consume the reward, after which the inter-trial interval started. Incorrect ('go') responses after sampling an odor predictive of quinine or the non-rewarded contingency had no further programmed consequences. Furthermore, no other, separate reinforcer was applied during these trials. The behavioral sequence comprising the departure from the sampling port to the fluid well, including nose entry and waiting period in the well, will be referred to as the 'go' response.

Surgery, electrophysiology and histology

Animals were anaesthetized with 0.08 ml/100 g Hypnorm i.m. and 0.04 ml/100 g Dormicum s.c. (Roche Netherlands, Woerden) and mounted in a Kopf stereotaxic frame. After exposure of the cranium 5 small holes were drilled into the cranium to accommodate surgical screws, one of which served as ground. Another larger hole was drilled over the OFC in the left hemisphere (centre of the hole 3.2
mm anterior, 3.2 mm lateral to bregma according to Paxinos and Watson, 1996). The dura was opened and the exit bundle of the hyperdrive was lowered onto the exposed cortex. The hole was subsequently filled with a silicone elastomer (Kwik-Sil, World Precision Instruments, Sarasota, Florida), after which the hyperdrive was anchored to the screws with dental cement. The hyperdrive consisted of an array of 12 individually drivable tetrodes and 2 reference electrodes (13 µm nichrome wire; Kanthal, Palm Coast, Florida), spaced apart by at least 310 µm (Gothard, 1996; Gray et al., 1995). Immediately after surgery all tetrodes and reference electrodes were advanced 1 mm into the brain; in the course of the next three days the tetrodes were gradually lowered until the OFC was reached. Animals were allowed to recover at least 7 days before the start of the recordings. Before each session started, tetrodes were lowered with increments of 40 µm to search for novel neurons to be recorded, after which the animal was brought back to his home cage and left to rest for at least two hours for the brain tissue to stabilize. Electrophysiological recordings were made using a Cheetah recording system (Neuralynx, Bozeman, Montana). Signals from the individual leads of the tetrodes were passed through a low noise unity-gain field-effect transistor preamplifier, insulated multi-wire cables and a 72 channel commutator (Dragonfly, Ridgeley, West Virginia) to digitally programmable amplifiers (gain 5000 times; band-pass filtering 0.6-6.0 KHz). Amplifier output was digitized at 32 KHz to record spike waveforms and stored on a Windows NT station. The occurrence of task events in the behavioral chamber was recorded simultaneously.

When all experiments with a given rat were finished, tetrode positions were marked by passing a 10 s, 25 µA current through one of the leads of each tetrode. After approximately 24 hours animals were perfused transcardially with a 0.9% saline solution followed by 10% formalin. Upon removal from the skull, brains were stored in a 10% formalin solution for several days before sectioning. Brain sections of 40 µm were cut using a vibratome and were Nissl-stained to reconstruct the tracks and final positions of the tetrodes. This showed that recording sites ranged from 2.7 mm to 4.7 mm anterior to bregma, and were limited to the ventral and lateral orbital regions of the OFC. Recording depth ranged from approximately 3.0 mm to 5.5 mm (Paxinos and Watson, 1996; Fig.1).

**Data analysis**

**Behavior**

Behavioral data was analyzed using SPSS for Windows (version 11.0). Unless otherwise stated, results are expressed as mean ± SEM values. We distinguished two measures to analyze reaction times during the task: Movement Time was defined as the duration between nose retraction from the odor port and nose entry into the
Figure 1. Localization of tetrode recording sites. As indicated by rectangles, recordings in all rats were localized in the ventral and lateral regions of the OFC, between 2.7 and 4.7 mm anterior from bregma. Recording depth ranged from approximately 3 mm to 5.5 mm (Paxinos and Watson, 1996). Indicated by black arrows in the histological section are several partially visible tetrode tracks. Black asterisks mark the lesion sites that show the final position of three tetrodes.

Electrophysiology
Isolation of single units

Single units were isolated by off-line cluster cutting procedures (BBClust/MClust 3.0). Before a cluster of spikes was accepted as a single unit, several parameters and graphs were checked visually: the averaged waveform across the four leads, the cluster plots showing spike parameter distributions such as peak amplitudes across the four dimensions, the autocorrelogram and the spike interval histogram. Since the absence of spike activity during the refractory period (2 ms) is indicative for good isolation, units of which the autocorrelogram and the spike interval histogram revealed activity during this period were removed from the analysis. For the ensemble analysis, all remaining units were taken into consideration.
Variability of the representation of reward magnitude

To examine variability in firing rates to reward magnitude, we calculated two variability measures (these have also been used for indicating sparseness of neural coding; cf. Perez-Orive et al., 2002; Rolls and Tovee, 1995). Briefly, the population variability ($S_{\text{pop}}$) is indicative of the variability in the mean firing rate of single cells across the population, irrespective of reward size, and is computed by:

$$S = \frac{N}{N-1} \frac{\bar{r}^2 - (\bar{r})^2}{\bar{r}^2}$$

(Eq.1)

where $N$ indicates the number of units and $r_j$ the mean firing rate of neuron $j$ during a particular trial phase, averaged across all three reward conditions. Thus, $\bar{r}$ represents the mean firing rate in the population and $\bar{r}^2$ the mean squared firing rate. The parameter variability ($S_{\text{par}}$), which is indicative of a single cell's response variability attributable to differences in reward size, was calculated in a similar fashion, but index $j$ now indicates each reward size and $N$ the total number of reward sizes ($N = 3$ in the current study). Both measures were calculated for the waiting and reward delivery phase within a time frame of 1.5 s; for the movement period the frame was 1.0 s. Values ranged between 0 and 1, with 1 representing the maximum variability attainable.

Ensemble analysis: general principles of population coding.

Population activity was examined using two different reconstruction methods, the Bayesian method and template matching. Before explaining both methods in more detail, we will first point out the general concepts behind them. A central postulate in systems neurophysiology is that macroscopic parameters (usually in the external world, such as the speed of a moving object) are encoded in the patterns of action potentials generated by neurons. Given these spike patterns, one may also ask how one could make sense out of them, that is how one could understand what type of information about macroscopic objects or properties they represent. To operationalize this general task, we may ask what a given spike train can tell the experimenter about the stimulus or cognitive process that gave rise to the spike train in the first place (Rieke et al., 1997). For example, given a situation in
which a particular spike train may have been evoked by any of three sensory stimuli, the task for a neutral observer, not aware of the actual stimulus, is to decide which of these stimuli was in fact applied to elicit the spike train. This process of \textit{reconstructing} the original stimulus (or other macroscopic variables) from the spike train is termed \textit{decoding}. The current objective is to decode the variable \textit{reward magnitude} from multi-neuron spike trains. This can be attempted for spike trains generated in anticipation of the reward (expected magnitude, as a cognitive variable), or in relation to the actual reward, as it is delivered to the animal.

Decoding can be accomplished successfully on the basis of single-cell firing activity (e.g. Bialek et al., 1991). If one would choose to use the average firing rate of a single neuron in order to decode a macroscopic variable, the procedure would rely on one scalar value. In ensemble recordings studies, however, the firing activity of many neurons offers a potentially much richer source of information; firing-rate values of all simultaneously recorded neurons can be used for decoding, and the series of firing-rate values from all neurons is conceptualized as a \textit{population vector}. If one records $N$ neurons in a given session, a population vector $\mathbf{v}$ can be represented as a series of firing rate values: $\mathbf{v} = (s_1, s_2, \ldots, s_N)$, where $s_i$ is the firing-rate value (scalar) of the first neuron, etc. Geometrically, a vector is an entity in Euclidean space that has both magnitude and direction; for instance, a vector composed of the firing rates of two cells $[v = (s_1, s_2)]$ can be rendered as an arrow in a two-dimensional plane.

If a neutral observer is only provided with a sample of firing-rate values of a given set of recorded neurons, but has no knowledge of how macroscopic parameters ‘map onto’ their firing responses, he would not be able to decode those parameters successfully. Thus, knowledge is needed as to how neurons are ‘tuned’ to the parameter under study; in other words, we need to know what the ‘standard’ parameter mapping or encoding of these neurons is. Thus, besides needing a population vector for actually decoding the sample set of firing-rate responses back to the value of the macroscopic parameter (i.e. the population vector for decoding), we also should have a ‘template’, or separate set of firing-rate values that tells us what the ‘standard’ mapping from the parameter onto firing-rate responses is (i.e. the population vector for encoding). In the current study, a total of 15 trials for each of the five odor-outcome pairs was available, and our standard approach was to extract the ‘template’ (or encoding vector) from the last 6 trials of each pair (i.e. trials 10-15), whereas the sample set for decoding was obtained from trials 1-9 of each pair (see below).

Given the population vectors for encoding and decoding, some kind of mathematically explicit comparison between the vectors must next be carried out to complete the task of decoding. The sample (decoding) set of firing-rates must be compared to the template (encoding vector) to reconstruct which macroscopic parameter value was the most likely one giving rise to the observed sample responses.
While both the template-matching and Bayesian reconstruction methods are described below in more detail, the template-matching method works essentially as follows. If we define a template (encoding vector) for two cells as \([t = (t_1, t_2)]\) with \(t_i\) being the firing-rate response of cell 1 to a given parameter, and a to-be-decoded sample of the same cells as \([d = (d_1, d_2)]\), then it is straightforward to visualize these two vectors as two arrows in two-dimensional space, having the same origin. The similarity (or degree of matching, hence the term ‘template matching’) between the vectors can be expressed as the cosine of the angle between the two vectors, as further explained below.

Although population-coding methods as described have the potency to yield important insights bridging gaps between the single-neuron and behavioral-cognitive levels, some limitations can be delineated. The objective of decoding is naturally limited to the parameter that was varied (in this case, reward magnitude); thus the OFC is likely to encode other macroscopic variables (cf. O’Doherty et al., 2001; Roesch et al., 2006; Wallis and Miller, 2003), which however should be addressed separately. Secondly, because we only measure and consider neural correlates of reward in OFC, no conclusions can be drawn as to how OFC populations come to express the capacity for reward-size coding; afferent pathways and intracortical mechanisms for achieving this capacity must be studied separately. Thirdly, while in the current study mean firing rate (per trial phase) was used as a measure of neural response, other aspects of firing patterns, such as related to spike timing, may make additional contributions to population coding (cf. Narayanan and Laubach, 2005). Besides these aspects there is no ‘code’ in the neural patterns that could be deciphered, or at least the current methods do not provide a way to do this, if possible at all.

**Population coding of reward magnitude: template matching**

For the main analysis, only the three positively rewarded trial types were taken into consideration, since the quinine and non-reinforced conditions did not yield enough trials for a meaningful analysis except for a control procedure (see Results section). As pointed out above, the sessions, all containing 15 correctly performed trials per reward size, were divided into two blocks, of which the first block (trials 1-9) was used for decoding and the second (trials 10-15) for encoding. The first nine trials of the session were chosen for decoding to be able to examine how the representation of reward information builds up during the initial learning phase in the task. We also examined whether a random selection of trials for decoding would provide a similar result; this analysis showed that with randomly selected trials decoding scores were obtained that were generally higher than with the trials 1-9.
Two vectors were constructed for each reward size, denominated as 
\( \mathbf{x} = (x_1, x_2, ..., x_N) \) and 
\( \mathbf{y} = (y_1, y_2, ..., y_N) \), containing the spike counts within a 
specified time window for the encoding (\( \mathbf{x} \)) and the decoding block (\( \mathbf{y} \)), with \( x_i \) and 
\( y_i \) indicating the spike count of cell \( i \) averaged across trials. Thus, the population 
vector \( \mathbf{x} \) is used for the encoding part of the procedure (i.e. for determining the 
“template” or ‘tuning curves’ of the cells towards reward magnitude, using the last 
part of the session, trials 10-15; see above). The population vector \( \mathbf{y} \) is used for the 
decoding part of the procedure, where the spike counts, specific for reward sizes, are 
taken from the same cells, but now from the first part of the session, trials 1-9). The 
two vectors were then compared to calculate the decoding score, which is the 
percentage of correctly identified reward amounts in the decoding phase, based on 
the activity patterns found in the encoding phase (Fig.2). Hence the ensemble code 
for reward magnitude is made up of the different firing rates of all recorded cells 
combined in the en- and decoding phase in relation to reward magnitude. Note that, 
besides mean firing rate per trial phase, other aspects of firing patterns, such as 
related to spike timing, may make additional contributions to ensemble coding (cf. 
Narayanan and Laubach, 2005).

For each trial phase in which we examined population coding of reward size we used 
a standard time window, corresponding with the duration of that particular phase 
within the trial. Reconstruction of reward size for the period during which the 
animal moved from the sampling port to the fluid well was done with a time frame 
of 1 s, whereas for the period when the animal awaited reinforcement with its nose 
in the fluid well, a time-frame of 1.5 s was used. During the reward delivery phase, 
the time frame in which reward size was reconstructed was 10 s, unless otherwise 
mentioned.

With template matching, the similarity between the vectors containing the spike 
count in the defined time window for the encoding and decoding block was 
calculated by computing the cosine of the angle between them (Lehky and Sejnowski, 
1990; Zhang et al., 1998; Louie and Wilson, 2001). A value of 1 represents an exact 
similarity between the two vectors and -1 the exact opposite, whereas 0 (i.e. 
orthogonal) indicates no similarity between the two vectors. We first calculated the 
inner product of \( \mathbf{x} \) and \( \mathbf{y} \):

\[
\sum_{i=1}^{N} x_i y_i
\]  
(Eq.2)

where \( x_i \) and \( y_i \) indicate the average firing rate of neuron \( i \) from a total of \( N \) cells 
within the specified time window for the encoding and decoding block, respectively 
(Fig. 2). Then the cosine was calculated by:
60

\[ \cos \theta = \frac{\sum_{i=1}^{N} x_i y_i}{\|x\| \cdot \|y\|} \]  

(Eq.3) 

with the denominator representing the product of the absolute vector lengths. If the decoding spike vector belonging to a particular reward amount produced the highest cosine value with respect to the encoding vector, then that particular reward size was selected as reconstructed amount of reward (Fig.2).

In several graphs, the decoding score (i.e. the percentage of trials in which the amount of reward was correctly reconstructed) was expressed as a function of time or as a function of the size of the ‘reconstruction ensemble’, i.e. the group of neurons which was subsampled from the entire population and used for the calculations. The maximum size of the reconstruction ensemble was 37: this value represents the median value of the number of cells recorded in all sessions, which ranged between
26 and 60. Calculations were made for each recording session, after which data was averaged. For the assessment of decoding as a function of size of the reconstruction ensemble, the decoding score was calculated a hundred times for each group size, each time with neurons randomly picked from the population recorded in that particular session. Decoding as a function of time was calculated with a reconstruction ensemble of 37 neurons as well: the decoding score was calculated a hundred times per time window, each time with randomly picked neurons. The decoding curves were further analyzed by applying linear regression analysis (P < 0.05) and a one-way ANOVA test with Bonferroni correction (P < 0.05).

Population coding of reward magnitude: Bayesian reconstruction

Bayesian or probabilistic reconstruction was used as previously described by Földiá k (1993), Sanger (1996), Zhang et al. (1998) and Thiel et al. (2007). Briefly, population vectors were calculated as pointed out above and decoding was based on the equation of conditional probability:

\[ P(s \mid y)P(y) = P(y \mid s)P(s) \]  \hspace{1cm} (Eq.4)

where \( P(s \mid y) \) indicates the probability of a reward size \( s \) given the multi-neuron spike pattern vector \( y \), \( P(s) \) indicates the prior probability of reward size, which does not need to be calculated since it has a fixed value of 1/3 due to the three different amounts of rewards that were applied with equal probability across trials. The probability \( P(y) \) for the spike-containing decoding vector \( y \) to occur does not need to be calculated as well: because the reconstructed amount of reward was the most probable reward size of the probability distribution (see below), this can be considered as an unnecessary scaling factor.

Under the assumptions of a Poisson distribution of spike timing and cells firing independently from each other, \( P(y \mid s) \) for every reward size \( s \) was computed by:

\[ P(y \mid s) = P(y \mid x(s)) = \prod_{i=1}^{N} P(y_i \mid x(s)) = \prod_{i=1}^{N} \frac{(x_i(s))^y_i}{y_i!} \exp(-x_i(s)) \]  \hspace{1cm} (Eq.5)

where \( x_i \) is the product of the length of the time window \( \tau \) and the average firing rate \( f_i(s) \) of cell \( i \) for a given reward size, and \( x(s) \) is the multi-neuron spike pattern vector used for encoding (i.e., determining the 'tuning' of neurons to reward size). We calculated the logarithms of the probabilities \( P(y \mid x(s)) \) to avoid working with extremely small values. The most probable reward size of the probability distribution was taken as the reconstructed amount of reward, which is the maximum of
The latter quantity is equivalent to $P(s | y)$ (see Eq. 4). Further data analysis was done as described for template matching.

Analyzing the data with these different decoding methods revealed that in all cases decoding scores for reward amount were slightly higher with template matching than with the Bayesian method. To elaborate on the lower decoding scores for Bayesian method as observed in our results, we recalculated decoding with various variants of the Bayesian method. When the spike timing distributions are approximated by a Poisson distribution, the firing rates within the decoding vector are rounded to integers (see Eq. 5). To examine whether the use of integer values contributes to a lower performance, we tried a gamma distribution instead, but did not observe an improvement. When a Gaussian was used as distribution of spike timing, decoding scores were comparable with those obtained under a Poisson distribution as well. This may be due to the high proportion of cells with low firing rates, which causes the use of a Gaussian distribution to be inappropriate. It should be noted, however, that cells with low firing rates were found to be important for reconstruction, since their removal from the reconstruction ensemble led to lower decoding scores than when these cells were taken into account.

Results

Behavior

Data from 17 recording sessions were used for the analysis, obtained from 6 rats. For all three positively rewarded trial types, animals needed on average about 17 trials to reach the criterion of 15 successful trials per reward size (0.05 ml: 16.7 ± 0.4, proportion correct 88%; 0.15 ml: 16.8 ± 0.6, proportion correct 88%; 0.30 ml: 17.9 ± 0.7, proportion correct 83%). For the non-reinforced and quinine trial types animals made on average 5.4 ± 0.7 and 5.1 ± 0.8 ‘go’ responses, respectively (sucrose versus quinine or versus non-reinforced: P = 0.000; paired sampled t-test). The reason why rats initially perform responses during quinine trials might be that they have to taste the reinforcing fluid in order to learn which outcome to avoid or to approach based on the initially neutral olfactory stimuli.

Examination of the Movement Time showed no significant differences between the positively reinforced trial types (0.05 ml: 1.42 ± 0.04 s, 0.15ml: 1.40 ± 0.04 s and 0.30 ml: 1.41 ± 0.04 s; number of trials per reward size: 255). When the positively reinforced trial types were combined (1.41 ± 0.01 s), Movement Time for sucrose trials was significantly shorter compared to quinine trials, but not compared to the non-rewarded trial types (quinine ‘go’ responses: 1.82 ± 0.16 s; n = 87; non-rewarded ‘go’ responses: 1.51 ± 0.07 s; n = 88). The Overall Response Time revealed that animals responded significantly faster to obtain the highest amount of
reward compared to the lowest amount; comparison of these reward amounts with the middle-sized reward showed no significant difference (0.05 ml: 3.76 ± 0.07 s, 0.15ml: 3.60 ± 0.06 s and 0.30 ml: 3.48 ± 0.06 s). Thus, learning within this task was evident from the faster Overall Response Time for the largest reward and from the lower number of quinine responses, which were performed with a slower Movement Time as well compared to responses during sucrose trials.

Electrophysiology
Variability in the representation of reward magnitude

Over the course of 17 sessions recorded in 6 animals, a total number of 683 single units was obtained. The number of single units recorded per session ranged from 26 to 60 (mean ± SEM: 40.2 ± 2.4), with a mean firing rate of 1.85 ± 0.15 spikes/s. Neurons could be specifically activated during several task phases, including odor sampling, the behavioral phase in which animals moved from the odor sampling port to the fluid well (‘movement phase’), the period of waiting for reinforcement in the fluid well and the reward period (i.e. the period after reward delivery lasting for 10 s) (Fig.3). Examination of ensemble activity focused mainly on two task periods, namely the waiting and reward delivery period, because we expected reward predictive information to be coded especially during the waiting period, and information about the size of the actual reward after reward delivery. When of particular interest, population coding was also examined for the movement phase. Since the task was not designed to determine whether differential responses during cue sampling were due to different sensory inputs (odor identity) or expectancy of varying reward magnitude, results regarding the coding of expected reward magnitude during this phase would be inconclusive. Hence, ensemble activity in this task period was not examined.

Prior to studying population coding principles themselves, we first aimed to assess (i) to what extent the firing-rate variability of single cells is attributable to effects of reward size, and (ii) whether whole OFC populations exhibit a high or low degree of overall firing-rate variability, irrespective of reward size. This was done by calculating two variability measures, namely parameter variability (S_{par}) and population variability (S_{pop}). The mean S_{par}, indicating the response variability of individual neurons associated with variations in reward size (their ‘tuning curves’), was 0.23 for the waiting period and 0.28 for the reward delivery phase (Fig. 4A and B). The mean S_{pop}, which represents the variability in firing rate across the population regardless of reward size effects, was 0.70 for the waiting period and 0.73 for the reward period (Fig. 4C and D). Similar findings were obtained for the movement period, namely a mean S_{par} value of 0.25 and a mean S_{pop} value of 0.76. These results show, first, that there is a high variability of firing rates throughout the
Figure 3. Summary of task-related firing activity of 32 simultaneously recorded OFC neurons during a single session. The horizontal axis denotes time (s), the vertical axis cell number. Color-coded firing rates were calculated by averaging activity over all rewarded trial types during the session, with bin sizes of 100 ms. Neural activity is synchronized on odor presentation (A), the end of the odor poke (B) and reward delivery (C). Onset of the waiting period is 1.5 s prior to reward delivery. Grey colors indicate firing rates that are normalized relative to the maximum firing rate of each neuron, according to a scale shown on the right-hand side. The reward-size variations in firing activity of these cells are used as the basis for studying the ensemble code of reward magnitude.

population, as illustrated by the high $S_{\text{pop}}$ values throughout the 3 trial phases. Second, this phenomenon is not matched by a similarly high reward-related variability when individual cells are considered: OFC neurons displayed less variability in their tuning curves to reward size, since for all three task periods the large majority of the cells had values below 0.5 (movement period: 76%, $n = 518$; waiting period: 79%, $n = 541$; reward period: 76%, $n = 520$). Apparently, other factors must be present besides the reward-modulated response profiles of the individual cells to explain the high variability in the overall population, e.g. differences in baseline firing rate or differences in responsivity during various task phases, although a small subgroup was present with a very high parameter variability (range 0.9-1.0) in all three task phases.

In conclusion, whereas firing rates throughout OFC populations are highly variable, modulation of single-cell firing rate that is attributable to reward size is
relatively subtle, in line with a previous analysis of single-unit activity in OFC (Van Duuren et al., 2007).

Figure 4. Distribution of population variability and parameter variability across orbitofrontal cell populations. A and C: waiting period; B and D: reward period. Values of population variability across sessions were generally in a high range between 0.4 and 1.0, with means of 0.71 and 0.73 for the waiting and reward period, respectively. Parameter variability values varied more strongly, spanning the whole range from 0.0 to 1.0 for both periods, with means of 0.23 and 0.28 for the waiting and reward period, respectively. Similar findings were obtained for the movement period, namely a mean $S_{par}$ value of 0.25 and a mean $S_{pop}$ value of 0.76 (not shown). Note the subgroup of cells with very high parameter variability (0.9-1.0).

Population coding of expected and actual reward magnitude

We will describe first an application of the Bayesian method and template matching in which the magnitude of reward was decoded from ensemble activity in the positively rewarded trial types across the first 9 trials of the session, whereas trials 10 to 15 were used for encoding. During the movement, waiting as well as the reward period, the magnitude of reward was decoded from the population activity with a percentage of correct performance significantly above the 1/3 chance level for both methods (one-way ANOVA; $P = 0.000$ in all three cases). When the decoding score was plotted as a function of the number of cells participating in the reconstruction ensemble, regression analysis showed that with template matching the slope of the decoding curve was significantly positive for all three task phases ($P = 0.000$ in all cases), meaning that the decoding performance improved with an
increasing amount of cells (Fig. 5). The maximum decoding scores were 57% (at n = 30 cells) for the movement period, and 48% (n = 36 cells) and 76% (n = 36 cells) for the waiting and reward period, respectively. We also calculated the decoding score for the subgroup of the sessions that did not show a significant difference in response latency between the lowest and highest amount of reward. Reward magnitude could be reconstructed above chance level in this group as well. This finding indicates that significant population coding is present in OFC even in the absence of overt behavioral differences, which corroborates the notion that OFC activity is not necessarily tied to motor behavior and involves a cognitive process.

The Bayesian method produced similar although generally somewhat lower decoding scores, with maximum scores of 44% for the movement period and waiting period (at respectively n = 26 and n = 24 cells), and 79% for the reward period (n = 36 cells). Also for this method, regression analysis indicated a significantly rising score with increasing ensemble size for the reward period (P = 0.000). For the waiting and movement period, however, no significant effect was found.

Across all analyses presented here, Bayesian reconstruction and template matching produced similar results for the reward period, whereas for the waiting-anticipation phase the decoding scores were similar or lower for the Bayesian as compared to template-matching method. Such differences may be due to several factors (see Materials and Methods). Below, we will concentrate on results obtained with template matching.

**Figure 5.** Decoding of reward magnitude with template matching: dependence on ensemble size. The horizontal axis indicates the size of the reconstruction ensemble, the vertical axis the percentage of trials in which reward size was correctly decoded. In the graphs presented here and below, the horizontal dashed line indicates chance level (33.3%) and dotted lines flanking the curves represent the 95% confidence interval (two times the standard error of proportion).

*Temporal resolution of ensemble coding*

Because the results indicated reward magnitude to be represented in population activity during task performance in a manner that was consistent across
de- and encoding blocks, the question arose what the temporal resolution of population coding was. This question is relevant for understanding OFC function, because when significant coding is present across short time intervals, this may facilitate rapid decision-making and fast behavioral responding. We first asked how the decoding score depends on the width of the time window used for en- and decoding applied to either the waiting or reward period. To this end we selected smaller time windows from the overall time windows of the waiting and reward period (1.5 and 10 s, respectively), and calculated the decoding score for time windows of increasing duration. This provides a cumulative measure of the percentage of decoded reward size over time within each trial phase. The waiting period was divided into 6 time windows that started at 250 ms till 1.5 s, with increments of 250 ms. Time windows used for the reward period were, in addition to the six windows of 250 ms, 2, 3, 4, 5, 6, 8 and 10 s. Regression analysis indicated that with template matching the percentage of reward amount correctly decoded during the waiting and reward period rose with increasing length of the decoding window, with the slope of the decoding curve being significantly positive (waiting period: \( P = 0.039 \); reward period: \( P = 0.000 \); Fig. 6A and B). The maximum decoding scores obtained were 47% and 80% for the waiting and reward period, respectively. A one-way ANOVA indicated that during the waiting period only the decoding score obtained with a decoding window of 1.5 s was significantly above chance level \( (P = 0.001) \), and furthermore revealed that there was no significant variation between the various window widths. In contrast, during the reward period, all time windows with exception of the first one differed significantly from chance level \( (P = 0.000) \). Furthermore, the first two time windows (i.e. 250 and 500 ms) had a significantly lower decoding score than the final 6 time windows (i.e. 3, 4, 5, 6, 8 and 10 s). Thus, for a significant success rate in decoding actual reward size, only a short time segment \( (\sim 0.5 \text{ s}) \) of the population response is needed, whereas the time segment needed for a significant success rate in decoding expected reward is somewhat longer, namely 1.5 s.

We next asked how the decoding capacity of the population varies over time across consecutive segments of the waiting and reward periods. This question probes the time resolution at which significant reward-predictive information is represented across consecutive time segments of a behavioral task phase. Therefore the percentage of correctly decoded reward amount was calculated for successive time segments across the entire length of the overall time windows of the waiting and reward period. Thus, a sequential method was adopted instead of the cumulative method used above. For the waiting period we used three consecutive time segments of 500 ms, for the reward period decoding was calculated during the first 2 s in time segments of 500 ms, followed by four time segments of 1 s, and the final two segments of 2 s. During both the waiting and reward period, regression analysis and
one-way ANOVA did not show a significant variation in decoding success, indicating that the amount of decodable information did not significantly differ across the various time segments within this task period (Fig. 6C and D).

**Figure 6.** Decoding of reward magnitude within specific trial phases. The size of the reconstruction ensemble was 37 neurons. Decoding score using time windows of increasing duration for (A) the waiting and (B) the reward delivery period. The horizontal axis shows the size of the time window (s) from which spikes were taken for reconstruction, the vertical axis the percentage of correctly decoded trials. Panels C and D depict the decoding success for consecutive temporal segments in the waiting and reward period, respectively. The horizontal axis now shows the time segment of trial phase used for reconstruction. In C, the decoding segments are the intervals (0-0.5), (0.5-1.0) and (1.0-1.5) s relative to fluid poke onset. In D, the segments are (0-0.5), (0.5-1.0), (1.0-1.5), (1.5-2.0), (2.0-3.0), (3.0-4.0), (4.0-5.0), (5.0-6.0), (6.0-8.0), (8.0-10.0) s relative to reward delivery.
Development of population coding during task progression

Since the data on Response Times and the number of trials per reinforcer indicated that in the course of a session animals generated predictions about response outcome and learned to discriminate between different amounts of reward, we addressed the question how the representation of reward magnitude by population activity evolved in the course of learning during a session. To this end, the decoding block was divided into four consecutive 2-trial blocks (block 1: trial 1 and 2; block 2: trial 3 and 4, block 3: trial 5 and 6; block 4: trial 7 and 8), and decoding was compared between these four blocks. Trial numbers 10-15 were used for encoding (as above). In all four trial blocks and for both the waiting and reward period, reward magnitude was decoded from the population activity above chance level (one-way ANOVA; P = 0.000). Regression analysis indicated for the waiting period a significant slope of the decoding score for blocks 2, 3 and 4. Decoding with the first trial block differed significantly from all three other blocks, with an improvement in decoding score as the task progressed, except for the final (fourth) trial block. The second block did not differ from the third block, whereas the third block differed significantly from the fourth (Fig. 7A). During the reward period, all blocks showed a significant slope of the decoding curves. Furthermore, the decoding curves differed significantly between the first and the second block and the third versus the fourth, with an improvement of decoding success; the second block did not differ from the third block (Fig. 7B). An additional two-way ANOVA indicated no interaction between cell group size and trial block.

The trial-block sequence observed for the waiting period, in which block 1 is flat whereas the blocks 2, 3 and 4 show increasing decoding with an augmenting number of cells as indicated by the regression analysis, suggests that OFC ensembles have only a weak predictive ability at the beginning of the session (the first block), which is already consistently present in small subsets of neurons. At this time in the session, the decoding score does not rise with an increasing amount of cells; adding more cells to the ensemble adds both ‘confirmatory’ and ‘conflicting’ evidence about the predicted reward size. However, the ability to predict reward size increases with experience over trial blocks (blocks 2, 3 and 4), and the decoding score now comes to depend on the number of cells that participate in the ensemble, as indicated by the regression analysis. This means that at this point, the code has developed in a more consistent and redundant manner within OFC, with less conflicting neural evidence.

Non-specific drift across trials as a possible confounding factor

The increased decoding success found for later trial blocks relative to earlier ones may be either explained by a learning effect or by a non-specific ‘drift’ in ensemble responses over trials since the final block of 5 trials (i.e. trials 10-15) was used for the encoding.
Figure 7. Changes in reconstruction success across consecutive blocks of learning trials. Panels A and B show the percentage of correctly decoded reward amounts per trial block for the waiting and reward period, respectively. In view of the reliability of reconstruction due to the low number of trials per trial block, a reconstruction ensemble of 26 neurons was used, which was the minimal number of neurons present in all sessions. Horizontal axis indicates the size of the reconstruction ensemble. In A and B, encoding vectors were obtained from trials 10 to 15 for each reward size. To maintain clarity, standard errors of proportion are not shown, but these values were generally comparable to those in Fig.4. Panels C and D show mean decoding success for the encoding by late trials (i.e., 10-15) and its temporal mirror image (encoding by early trials, i.e. 1-6) for the waiting (C) and reward period (D). The abscissa denotes the proximity of decoding trials to the encoding block; the larger the decoding block number, the closer that block is to the encoding block. Error bars indicate SEM values.

To check whether the latter, confounding possibility would apply, we compared the 2-trials decoding blocks when the encoding block was situated at the session end (trials 10-15) with 2-trials blocks when the encoding block was situated at the start of the session (encoding block trials 1-6; decoding block 1: trial 7 and 8; block 2: trial 9 and 10, block 3: trial 11 and 12; block 4: trial 13 and 14). Because outcome-prediction learning is expected to be unevenly distributed across trials (with a rapid decrease in go-responses for the quinine and non-rewarded trials early in the session), it is predicted that these decoding procedures should yield separate or at most
partially overlapping curves for the waiting period if the progressive change in decoding is indeed due to learning. Thus, the curve with encoding by late trials should rise more steeply and then stabilize as compared to its temporal mirror image. As illustrated in Fig. 7C, showing the average of the decoding curves per trial block for the waiting period (as in Fig. 7A), the curves indeed confirmed this prediction. A one-way ANOVA (P < 0.05) indicated that for the waiting period all trial blocks of the two different encoding conditions differed significantly from each other, except for the fourth block. For the reward period – in which additional learning may or may not take place - the first and the fourth trial block differed significantly between the two encoding conditions, with a steeper rise across trial blocks when the encoding block was at the session end (Fig. 7D).

Altogether, these results indicate that progressive learning of odor-reward associations coupled to motor responses is accompanied by a quick rise in population coding of expected reward magnitude during the waiting period, followed by relative stabilization – a time course not attributable to non-specific drift. An additional learning-related increase in population coding appears to take place in the reward period, although this effect is less strong.

**Temporal specificity of reward-coding within trials**

In principle, it is possible that population activity does not code reward magnitude only during the single trial phases for which it was examined. For example, information about reward amount may be carried over from one trial period to the next, similar to ‘delay cells’ in e.g. monkey dorsolateral prefrontal cortex (Funahashi et al., 1989). In contrast, one may hypothesize that this type of information is coded by specific neuronal groups, active during a particular trial phase, without any working memory-like activity or carry-over to the next trial phase. To examine this, the time window in which decoding was calculated for the movement period (1 s) was shifted in time with a step size of 0.25 s relative to the onset of this period, whereas the time window for encoding remained unchanged. For each step the amount of decoding success was recalculated. If coding of expected reward magnitude is confined to periods around such a specific event, decoding is expected to approximate chance level shortly before and after the event period. As expected, Fig. 8A shows that during the movement period the decoding score for the expected magnitude of reward was highest when no time-shift was applied (Δt = 0). For both negative and positive shifts in time, the decoding scores decreased rapidly to chance level: the scores differed significantly from the decoding score calculated at time t = 0 (i.e. lower than two times the confidence interval at time t = 0) in the period after + 0.5 and before - 0.5 s, suggesting a large segregation with population coding during earlier and later trial phases.
The same procedure was applied for the waiting period, having a time-frame of 1.5 s (Fig. 8B). Relative to the maximal decoding score obtained when no time-shift was applied, the score dropped slightly within + 1.5 s after event onset, but only slowly returned to chance level after the + 1.5 s period. The decoding score differed significantly from the decoding score at time $t = 0$ in the period after + 2.0 s. The slow decay is likely related to a certain consistency in ensemble firing patterns across the phases of expected versus actual reward. In contrast, negative shifts in time produced a steep decrease in decoding score; the decoding scores were found to be significantly lower compared to the score at time $t = 0$ in the period before – 0.2 s.

**Figure 8.** Temporal specificity of coding reward size assessed with reference to (A) the movement, (B) the waiting and (C) the reward delivery period. A reconstruction ensemble of 26 cells was used, and the decoding time windows were 1 s for the movement period and 1.5 s for both the waiting and reward period. The abscissa plots the time (s) by which the decoding vector was shifted in 0.25 s steps relative to the encoding vector, with $t = 0$ at the offset of odor sampling (A), the onset of fluid poking (B) and reward delivery (C). The decoding score is significantly lower (i.e. lower than two times confidence interval) compared to time $t = 0$ before – 0.5 s and after 0.5 s for the movement period. For the waiting period, the decoding score is significantly lower before – 0.25 s and after + 2.0 s, for the reward delivery period before – 0.25 s and after + 0.75 s. Parallel vertical lines in indicate the average onset of odor poking (OP), fluid poke onset (FP), reward delivery (R) and onset of the inter-trial interval (IT). For reasons of clarity, standard errors of proportion are not shown. Horizontal lines represent two times the standard deviation above or below chance level.
For the reward period, a time window of 1.5 s was applied as well (Fig. 8C). Also in this case, the percentage of correct decoding decreased relatively slowly within +3.0 after event onset. Shifting leftwards, the decoding scores were found to be significantly lower as compared to the score at time t = 0 in the period before −0.25 s, again suggesting a segregation with earlier trial phases, and in the period after +0.75 s. We also examined a time window of 10 s for the reward period (data not shown), which provided a comparable result, although the decay slopes for positive and negative shifts in time were less steep because an overlap between the encoding and decoding window remained until boundaries of +10 or -10 s were reached by time shifting.

We performed an additional analysis to examine whether the template for encoding the magnitude during the reward phase can be used for decoding the expected magnitude during the anticipatory period, and how such matching develops as learning progresses. If significant matching exists, this would confirm the consistency of the ensemble pattern across the two trial phases. To this end, we used the reward period for encoding and the firing activity during the anticipatory period for decoding. We divided the decoding block into four consecutive 2-trial blocks (block 1: trial 1 and 2; block 2: trial 3 and 4, block 3: trial 5 and 6; block 4: trial 7 and 8), and compared decoding between these four blocks. Trial numbers 10-15 were used for encoding. This analysis showed that in all four trial blocks there was a modest but significant decoding of reward magnitude (P < 0.05), but the fourth block showed a significantly lower decoding score with respect to the previous three trial blocks (mean ± SEM: block 1: 38.5 ± 0.2%; block 2: 38.6 ± 0.2%; block 3: 38.3 ± 0.3%; block 4: 35.3 ± 0.4%; Fig.9). This indicates that initially the representation of reward magnitude is similar between these two task phases, meaning that in this phase of the learning task there is a large overlap in reward-related ensemble patterns of the anticipatory and delivery phases. During later stages of the task, however, the neural representations of the reward and anticipatory phases become more differentiated. Apparently, during task progression the activity of the cells involved in magnitude representation shows a marked change upon the transition from the anticipatory to the delivery phase, which can be viewed as a partial decorrelation of the code.

In conclusion, whereas the ensemble firing patterns in the movement phase do not show marked carry-over to adjacent phases, a notable consistency in coding exists between the waiting and delivery periods. This consistency is also expressed when using the reward phase for encoding and the waiting period for decoding, especially in the early phases of learning. As regards the lack of consistency or carry-over between the earlier trial phases (odor sampling, movement and waiting period), it is of note that our data are at variance with a view where reward-magnitude information would be maintained by the OFC throughout various trial phases in a
working memory-like manner, as exemplified by dorsolateral prefrontal neurons in primates showing enhanced firing during delay intervals (Leon and Shadlen 1999; Fuster et al., 2000; Wallis et al., 2001).

**Figure 9.** Decoding of reward size during the waiting period when spike vectors from the reward delivery period were used for encoding. A reconstruction ensemble of 26 neurons was used and the decoding time window was 1.5 s. For all four two-trial blocks (see also Fig. 7), a decoding score was obtained significantly above chance level. The fourth trial block showed a significantly lower decoding score with respect to the previous three trial blocks. Error bars indicate SEM values.

**Contribution of individual cells to coding of reward magnitude**

Decoding expressed as a function of cell number (Fig. 5) does not reveal individual contributions of cells to the population code of reward magnitude since in this calculation decoding scores were averages across groups of randomly selected neurons. To assess the contribution of individual cells to the decoding success, we computed for all recorded neurons the difference in the percentage of decoded information when a specific cell was added to a group of five neurons randomly selected from the same session. Apart from the consideration that single cells may contribute reasonably to coding by such a relatively small group, this size was chosen arbitrarily. For each cell this calculation was done a hundred times, each time with a new randomly selected group of five additional neurons. This procedure showed that during the movement period, 32% of the cells \((n = 219)\) made a minimal contribution to the decoding capacity of the population \((-0.5 \text{ and } 0.5\%)\). Thirty five percent of the cells \((n = 239)\) made a positive contribution \(>0.5\%) to successful reconstruction \((\text{average contribution} \pm \text{SEM}: 6.5 \pm 0.4\%)\), whereas 33% \((n = 225)\) contributed negatively \(<-0.5\%) with an average of 4.3 \pm 0.4\%, implying that addition of these cells led to a decrease in correct decoding. The positive contribution was found to be significantly higher than the negative one \((\text{unpaired t-test}; P = 0.000)\). Three percent of the cells \((n = 20)\) made a contribution of \(>15\%) to the decoding score \((\text{average: } 20.1 \pm 0.9\%)\), whereas 1.5% \((n = 10)\) made a large negative contribution \(<15\%) that was significantly lower than the positive contribution, namely 17.2 \pm 0.7\% \((\text{unpaired t-test}; P < 0.05)\).
During the waiting period, 26% of the cells (n = 178) made a minimal contribution to the decoding capacity of the population, whereas 40% of the cells (n = 273) made a positive contribution (> 0.5%) to successful reconstruction (average contribution ± SEM: 5.1 ± 0.3%). Thirty four percent (n = 232) contributed negatively with an average of 4.5 ± 0.3%. The average positive contribution, however, did not differ significantly from the negative one. Examination of the number of cells that made a contribution of more than 15% to the decoding score (either positive or negative) showed that 2% of the cells (n = 14) contributed positively to the decoding score with an average of 21.4 ± 1.9%, whereas 0.6% (n = 2) made a negative contribution that was significantly lower than the positive contribution, namely 20.5 ± 1.4% (unpaired t-test; P < 0.05).

During the reward period, 15% of the cells (n = 102) made a minimal contribution, whereas 50% (n = 342) and 35% (n = 239) contributed in a positive (average contribution ± SEM: 6.7 ± 0.4%), and negative (4.0 ± 0.4%) manner, respectively. The positive contribution was found to be significantly higher than the negative one (unpaired t-test; P = 0.000). Six percent of the cells (n = 41) contributed more than 15% to the decoding score in a positive manner (23.4 ± 1.3%), whereas 0.4% (n = 3) was found to contribute more than 15% in a negative manner (19.0 ± 1.3%). Also in this case, the negative contribution was significantly lower compared to the positive one (unpaired t-test; P = 0.000).

In addition, cells that showed the largest variability in their response towards reward magnitude (i.e., parameter variability between 0.9 and 1.0; Fig. 4) were removed from the entire population of neurons and the decoding score for the remaining population was recalculated (number of neurons removed: movement period: n = 69; waiting period: n = 62; reward period: n = 79). This resulted in decoding scores that did not differ significantly from the decoding curves obtained using the entire population.

In summary, for all three task phases we found large subgroups of cells making modest to large contributions to the population representation of reward magnitude, both positive and negative. These results confirm the notion (see also Fig. 5) that reward magnitude is coded in a distributed and redundant manner. Although subsets making positive contributions generally outweighed the ‘negative contributors’, it is striking to note that so many ‘negative contributors’ were found, which implies that many cells generated firing patterns that were inconsistent with respect to reward size across en- and decoding trials. This notion is especially relevant when framed in the context of the problem, how target areas of the OFC ‘read out’ its population activity (cf. Pouget et al., 1998; 2000). If one considers previously acquired results indicating single OFC cells to code information on e.g. expected reward size or delay up till reward delivery, one may be inclined to think that such single-cell signals can be easily “read out” and used by a target area (such
as the striatum) for further computations, but this is a simplification of the problem if one reflects on the current results. Namely, there appear to be many other OFC neurons in the vicinity of the specifically tuned one that may contribute ‘noise’ and or even make a negative contribution to the population code, relative to that neuron. Despite this read-out problem, the present results also show that target areas of the OFC — if endowed with proper processing circuitry — can extract significant information on predicted and actual reward sizes from OFC population output, even in early phases of learning.

Erroneous go-trials: quinine responses versus sucrose responses

During the task, animals learned to avoid negative response outcomes, as visible by the low number of go-responses to quinine (‘false alarms’). We examined whether reconstruction of reward size could be achieved by using the quinine-reinforced Go trials for encoding and positively reinforced trials for decoding. If, as hypothesized, the decoding success during the waiting period is truly attributable to the expectation and processing of reward information, reconstruction for quinine versus sucrose trials should deliver a random decoding score (i.e. close to 33.3% correct). Alternatively, however, some correct reconstruction may occur due to the expected and actual volume of liquid ingested, regardless of the quality of the reinforcer (the volume of quinine solution was equal to the middle-sized sucrose reward, viz. 150 microliter). If this is the case, a significant decoding success is predicted to occur during both the waiting and reward period. For this calculation, all available quinine trials were used for encoding, whereas for decoding the sucrose trials 1-9 were used. Fig. 10 shows that for both the waiting and reward period the decoding score for reward size was around chance level regardless of ensemble size, confirming that decoding success in Fig. 5-8 is attributable to reward quality. Using only trials with 0.15 ml of sucrose reward to calculate the decoding score when quinine was used for encoding yielded a similar result, namely a percentage correct at chance level. This furthermore indicates that variations in population activity observed during these two trial periods are not due to licking behavior of the animal, since in both trial types reinforcement was consumed by the animals (see also Van Duuren et al., 2007 for consistent single unit results in relation to licking).

Discussion

The involvement of the OFC in decision-making has been primarily examined by lesion experiments or single unit recordings made during task performance. Based on these studies, the OFC has been proposed to integrate learned associative information with action plans to guide behavior in a flexible manner (Baxter et al.,
Figure 10. Decoding scores for reward size (sucrose solution) when spike vectors from quinine trials (amount of 0.15 ml) were used for encoding. Results for (A) the waiting and (B) reward delivery period. During both task phases the decoding success for reward size is around chance level, which indicates that the population activity represents reward quality and not volume of liquid reinforcer. When only trials with 0.15 ml of sucrose reward were used for decoding, a similar result was obtained: the decoding score was around chance level. To maintain clarity, standard errors of proportion are not shown, but these values were generally comparable to those in Fig. 4.

2000; Dias et al., 1997; Fellows and Farah, 2003; Gallagher et al., 1999; O’Doherty et al., 2001; Plassmann et al., 2007; Ramus et al., 2000; Schoenbaum et al., 1998, 2006; Tremblay and Schultz, 1999). However, it has remained uncertain how behaviorally meaningful information is represented by OFC population activity when adaptive behavior is required. By performing ensemble recordings while animals learned odor-reinforcer associations, we showed that information about reward magnitude could be decoded from OFC population activity during three trial phases, viz. when animals generated movements directed at reward, during waiting for reward, and when reward is received. Comparing reconstruction success between these periods, the decoding score was highest for the reward period; apparently, population activity is more discriminative with respect to varying reward amount when based on the actual presence of reward than based on predictive information derived from odor cues. It can be argued that during the reward period OFC receives other information besides valuation inputs, viz. sensorimotor feedback related to liquid ingestion per se, but this does not explain the random decoding score when quinine trials were used for encoding and sucrose trials for decoding. Despite this argument, population activity during the reward period may be determined by other processes besides reward-appraisal, e.g. tasting, which is inextricably linked to it. The lower decoding scores observed in the waiting period may be explained by the fact that animals were in the process of learning novel odor-reward contingencies, which is supported by the rapidly rising decoding scores during progressive trial blocks (Fig. 7A). Another, not mutually exclusive explanation holds that no specific task requirement to obtain a high reward amount was present, possibly leading to a
relatively weak demand on the system’s computational resources to represent reward magnitude. Although animals are immobile during the waiting period, we cannot rule out the possibility that for example small changes in posture may influence neural activity within this period. Changes in attention, intimately linked to reward expectation, might also occur, but it is unlikely that the results would be confounded by attentional factors, because OFC has not been implicated in attention per se, and attentional modulation of firing responses would not explain the difference between coding for sucrose versus quinine.

**Development of population representation of reward magnitude**

Despite a limited number of trials, we were able to demonstrate for the first time an increasing success in reconstructing reward size during task progression. Across trials, animals learned to discriminate between different reward sizes, and hence this increase in correct predictions of reward size may reflect learning-related changes. Indeed, testing the alternative hypothesis of a non-specific ‘drift’ in ensemble response over trials indicated a learning effect, because the forward curve (late encoding) exhibited a steeper rise and a modest decay as trials progressed compared to the more gradually rising backward curve (early encoding). Increasing decoding scores were also found for the reward period, where the forward curve rose more steeply than the backward curve. This may be explained by an adaptation in ensemble activity correlating to an additional learning process. An initial number of trials in which the various rewards are actually assessed by the animal may be needed to train the population to distinguish them. Admittedly, however, our task did not include requirements to assess this type of learning explicitly.

**Variability and redundancy of the representation of reward magnitude**

Mean firing rates across the population showed a high degree of variability, consistent with a great variability in activity levels during specific trial phases. Despite this, individual cells showed a vast range of variability values towards reward magnitude. Thus, the great overall variability in neural activity in the population was only partially matched by differences in individual neuronal responses to reward size (‘tuning’), suggesting a limited degree of the variability in reward-magnitude representations.

Decoding scores were weakly dependent on the ensemble size used for reconstruction (Fig. 5), and removing the fraction of cells demonstrating high parameter variability did not significantly alter the score, implying that the absence of these cells did not result in a significant loss of reward-predictive information. Considering individual cell contributions to the performance of a small ensemble (n = 5), only a few cells made a substantial (> 15%) positive contribution to the decoding score, whereas a large number of cells made small to moderate (0.5 - 15%)
contributions. These results are consistent with the concept of redundant coding, meaning that reward magnitude is coded in a broadly, distributed manner within OFC, and losses of considerable subpopulations will not lead to a strongly degraded output signal. That the amount of cells making either a negative or positive contribution in the waiting period is about equal, is compatible with a gradual development of reward-predictive population activity during learning and a moderate decoding success for this phase. Such negative contributions may easily arise when spike patterns show a high variability across trials inconsistent with variations in upcoming reward size.

Temporal specificity of reward coding within and across trial phases
Besides decoding reward size within specific temporal phases of learning trials, we also studied whether representations remain detectable at a finer time resolution. When reconstruction time windows in the reward phase were gradually increased, decoding performance tended to stabilize after about 3 s, indicating that significant representations can be detected at a time resolution finer than entire trial phases. For the waiting period, however, 1.5 s was needed to reach a significant decoding score (Fig. 6A and B). Analysis of decoding performance across successive time segments of both trial phases (Fig. 6C and D) showed that decoding success did not greatly vary with the temporal position of the segment in each phase. However, this preserved decoding capacity for small intervals within trial phases does not imply that the same OFC ensemble would code reward information consistently throughout the trial, as shown by a decreased success rate when decoding frames were shifted in time (Fig. 8). For the waiting period a shift towards the right resulted in a slow decrease in decoding success, but a shift in the reverse direction quickly brought decoding to chance level. Apparently, cell groups representing reward-predictive information during this period do not generate similar representations during earlier trial phases, e.g. odor sampling. A similar lack of carry-over was found for the movement phase. In addition, no consistent decoding of reward-frames was found during the inter-trial interval, arguing against a maintenance of reward representations by the same neurons between consecutive trials.

Population coding of reward information in OFC: functional implications
Only a few studies have related ensemble activity to reward processing in OFC. In an odor-discrimination task, Schoenbaum and Eichenbaum (1995b) found that during odor sampling ensembles coded odor identity and expectation of reward presented in the following trial. This pioneering study differs from ours because we focused on outcome prediction on the basis of a cue within the same trial, on temporal coding specificity within trials, and on dynamic adaptation during learning of novel associations. Gutierrez et al. (2005) demonstrated population activity
discriminating water and sucrose rewards, both when animals anticipated reward and when they tasted it. In this experiment, predictive cues were absent and animals were allowed to drink water, followed by a session of sucrose-solution intake. Such a coarse time indicator for outcome anticipation may be confounded by drifts in ensemble coding over session time, and differential anticipatory activity may be confounded by differences in motor preparation. An advantage of a parametrically varied reward is that such possible confounds can be controlled for since animals perform similar behavior to obtain different reinforcers.

The overall concept of OFC population coding emerging from the present findings holds that representations of reinforcer quality and magnitude are broadly distributed across ensembles and are characterized by a high, sub-second time resolution. These properties cannot be deduced by other techniques such as single-cell recording or fMRI imaging. Combining the current results with single-unit results suggesting task-phase specific coding (Simmons and Richmond, 2008), neural coding in OFC appears highly specific and well-articulated for the temporal phase an animal is in towards achieving a goal, while reward parameters as magnitude are expressed more as modulatory signals, rather than being main determinants of firing rate. When considering how such parallel-distributed signals may be read out by neuronal populations in target areas of OFC, it is noteworthy that significant readout of reward-predictive information can occur by processing output from small groups of neurons and of narrow time segments within trials, enabling rapid decision-making. If such target populations assume a particular functional organization as a continuous attractor, they have the natural capacity, at least in principle, to read out OFC output efficiently by template matching or Bayesian reconstruction (Zhang et al., 1998, Pouget et al. 1998, Wu and Amari, 2005).

Acknowledgements

This work was supported by NWO Grant 903-47-084, NWO grant 918.46.609 and BSIK (SenterNovem) grant 03053. We would like to thank David Redish and Peter Lipa for providing the cluster cutting software, Els Velzing for help with graphical illustrations and our colleagues for their comments on the manuscript.
Chapter 4

Single cell and population coding of expected reward probability in the orbitofrontal cortex of the rat
Abstract

Based on behavioral and imaging studies, the orbitofrontal cortex (OFC) has been strongly implicated in decision-making under conditions of uncertain outcome. However, it is still unknown whether and how information about the probability of reward is actually coded by single orbitofrontal cells and neuronal ensembles. To this end, ensemble recordings in rat orbitofrontal cortex were performed while animals learned an olfactory discrimination task in which four different odor stimuli were predictive of different reward probabilities (0, 50, 75 and 100%). Firing patterns correlating to the probability of upcoming reward occurred during the execution of the behavioral response towards the reward site and within an immobile waiting period prior to reward delivery. In the reward period, neurons were found to respond either to the presence or absence of reward, or to both. The subset of neurons that responded to omission of an expected reward did not show this response during similar behavior outside task trials, suggesting a correlate to expectancy violation. The neuronal population was characterized by a wide divergence in the firing-rate variability attributable to reward probability. Using template matching as a reconstruction method, population analysis revealed that reward probability was significantly represented in ensemble activity during the movement and waiting phases of the task, with a decoding score that was only weakly dependent on neuronal group size, with variable contributions by individual neurons. These results are consistent with a distributed representation of expected reward probability within OFC. Comparison of the population responses between the P = 0% condition and the other three reward probabilities confirmed that predictive information coded by the population was quantitatively related to reward probability and not to reward uncertainty.
Introduction

One of the key factors in decision-making is the probability of future rewards resulting from voluntary actions. Behavioral studies in humans have shown that a certain reward is generally preferred over an uncertain or probabilistic reward of the same amount: in a process called probability (or odds) discounting, the value of probabilistic rewards is degraded as the reinforcer becomes more uncertain (Rachlin et al., 1991). The choice between small, likely rewards and large, unlikely rewards has been found to activate the orbitofrontal cortex (OFC) (Ernst et al., 2004; Rogers et al., 1999), an area of the prefrontal cortex that has been strongly implicated in the assessment of reward value (O’Doherty et al., 2001, 2003) and in the planning of actions leading to immediate rewards (Tanaka et al., 2004). Additional support for an involvement of the OFC in decision-making under uncertainty comes from studies with humans suffering orbitofrontal damage. These patients perform poorly on tasks involving uncertainty, such as the Iowa gambling task, by continuing to choose high-risk decks of cards, whereas normal subjects bias their choice behavior towards low-risk decks (Bechara et al., 1996, 1997).

Furthermore, recent brain imaging studies have shown that visual stimuli associated with higher uncertainty elicit increasing activations in orbital areas (Tobler et al., 2007), and suggest a positive correlation between orbitofrontal activity and the unpredictability of reward (Hsu et al., 2005).

Findings in rodents support the notion of OFC mediating choice behavior in tasks that involve risky decision-making, although the results are contradictory. Using a probability discounting paradigm, Mobini et al. (2002) demonstrated that orbitofrontal-lesioned rats preferred the smaller, certain reinforcer over the larger, but infrequent reward. More recently, Pais-Vieira et al. (2007), using an alternative probability discounting paradigm more similar to the gambling tasks used in humans, showed that animals with orbitofrontal lesions preferred the larger, but uncertain reward, which is in accordance with stronger risk-taking behavior as demonstrated by data from patients with prefrontal lesions. The contrasting results of these two rodent studies are likely caused by differences in experimental design and methods, but altogether these animal studies and the human imaging data do suggest that the OFC is involved in assessing the value of rewards on the basis of their certainty. Despite this body of evidence, it is still unknown whether and how the expected probability of reward is coded in the OFC. Studying this specific topic may also shed light on the more general question how neuronal populations represent uncertainty of any behaviorally relevant variable, be it sensory, motor or motivational (Knill and Pouget, 2004; Daw et al., 2005). To examine how the firing activity of orbitofrontal neurons is affected by expectation of a varying reward probability, we trained rats on a probabilistic olfactory discrimination ‘go-no/go’ task, in which
odors were predictive of the probability of a pellet reward. Single-cell and population activity was examined in several trial phases, including a movement and an immobile waiting period in which animals anticipated reward, as well as after reward delivery.

**Materials and Methods**

**Subjects**

All experiments were approved by the Animal Experimentation Committee of the Royal Netherlands Academy of Arts and Sciences and were carried out in accordance with the National Guidelines for Animal Experimentation. Data were collected from 4 male Wistar rats (Harlan CPB; Horst, The Netherlands), weighing 375-425 g at the time of surgery. Animals were socially housed in standard type 4 macrolon cages, weighed and handled daily, and kept under a reversed 12 hr light/dark cycle (dimmed red light at 7:00 AM). Animals were maintained on 90% of their free-feeding body weight (16 g standard rat food chow per day per rat), with water available *ad libitum*. After surgery the animals were housed individually in a larger cage (1 x 1 x 1 m) under the same conditions.

**Behavior**

**Apparatus**

The recording chamber (40 x 37 x 41.5 cm), placed in a sound-attenuated and electrically shielded box, had a black interior with straight walls. The front panel contained on the right side a light signaling trial onset and an odor sampling port, on the left side a food trough. Behavioral events and data collection were controlled and registered by a computer. Both sampling port and food trough contained an infrared beam transmitter and detector port inside to detect the responses made by the animals. Odor delivery was controlled by a system of solenoid valves and flow meters (van Duuren et al., 2007) with separate delivery lines for each odor to prevent mixture of odors in the system. Two pellet dispensers were present (ENV-203 Magazine Type, 45 mg MED Associates), one for pellet delivery (45 mg sucrose pellets; Bioserve), and one empty dispenser used to mimic the sound of the dispenser during unrewarded trials. The odorants (Tokos BV; Noordscheschut, The Netherlands) were separated into different families, i.e. fruity, floral, herbal, woody and citrus. For each discrimination session, four distinct odors were used, each odor from a different family. Furthermore, no single family of odors was preferentially associated with a particular trial outcome.
**Behavioral paradigm**

After habituation, animals were progressively trained on the behavioral procedure of the four-odor probability task. Four new odors were used in each discrimination session, each odor associated with a specific reward probability, i.e. $P = 100\%$, $P = 75\%$, $P = 50\%$ and $P = 0\%$. Animals were initially trained to make a nose poke in the odor sampling port, which was sufficient to immediately obtain reinforcement by visiting the food trough. In the next phase animals learned to make an odor poke with a minimal duration of 1.5 s. In the final stage of shaping a waiting period of 1.5 s was introduced upon the poke in the food trough and before the pellet was delivered.

Once animals were familiar with the behavioral procedure of the task, two different 4-odor discrimination problems were consecutively presented to the animal to provide additional training. After rats learned new odor-reward probability associations (as visible by withholding responses towards or at the food trough after sampling the odor predictive for the null reward probability), they were implanted with a headstage containing an array with individually movable tetrodes ("hyperdrive") and recordings started. During each recording session, a new set of four odors was presented, which were chosen pseudo-randomly. During the task, trial onset was indicated by the trial light switching on, after which the animal had 15 s to make an odor poke. If no odor poke was made, the trial light turned off and the inter-trial interval (with a variable duration of 10-25 s) started. Whenever a prolonged odor poke was made, the trial light switched off after 0.25 s, followed 0.25 s later by the presentation of an odor. This period was included to prevent the animal from moving during cue sampling. Odor sampling itself was required to last at least 1 s. After retraction of the animal’s nose out of the odor sampling port or whenever a maximal duration for odor sampling (10 s) was exceeded, odor presentation was terminated. Premature retraction from the odor sampling port (odor pokes shorter than the minimal duration of 1.5 s) resulted in the start of the inter-trial interval. Following the waiting period in the food trough of 1.5 s, a pellet was delivered during the reward trials, and 5 s later the inter-trial interval started. The behavioral sequence comprising the departure from the sampling port to the food trough, including nose entry and waiting period in the food trough, will be referred to as the ‘go’ response.

**Surgery and electrophysiology**

Animals were anaesthetized with 0.08 ml/100 g Hypnorm i.m. (0.2 mg/ml fentanyl, 10 mg/ml fluanison) and 0.04 ml/100 g Dormicum s.c. (midazolam 1 mg/kg) and mounted in a Kopf stereotaxic frame. After exposure of the cranium 5 small holes were drilled to accommodate surgical screws, one of which served as ground. Another hole was drilled over the OFC in the left hemisphere (centre of the
hole 3.6 mm anterior, 3.2 mm lateral to bregma according to Paxinos and Watson, 2005). The dura was opened and the exit bundle of the hyperdrive was lowered onto the exposed cortex, after which the hole was filled with a silicone elastomer (Kwik-Sil, World Precision Instruments, Sarasota, Florida), and the hyperdrive was anchored to the screws with dental cement. The hyperdrive, which was custom built, contained an array of 12 individually drivable tetrodes and 2 reference electrodes (13 µm nichrome wire; Kanthal, Palm Coast, Florida), spaced apart by at least 310 µm (Gothard, 1996; Gray et al., 1995). Immediately after surgery all tetrodes and reference electrodes were advanced 1 mm into the brain; in the course of the next three days the tetrodes were gradually lowered until the OFC was reached. Animals were allowed to recover at least 7 days before the start of the recordings. In order to record different units during each recording session, all tetrodes were lowered at the start of a recording day with increments of 40 µm. Once the tetrodes were lowered the animal was left to rest in his home cage for at least 2 hours in view of unit recording stability, after which the experimental session started.

Electrophysiological recordings were performed using a Cheetah recording system (Neuralynx, Tucson, Arizona). Signals from the individual leads of the tetrodes were passed through a low noise unity-gain field-effect transistor preamplifier, insulated multi-wire cables and a 72 channel commutator (Dragonfly, Ridgeley, West Virginia) to digitally programmable amplifiers (gain 5000 times; band-pass filtering 0.6-6.0 KHz). Amplifier output was digitized at 32 KHz and stored on a Windows NT station. The occurrence of task events in the behavioral chamber was recorded simultaneously.

After finishing experiments with a given rat, tetrode positions were marked by passing a 10 s, 25 µA current through one of the leads of each tetrode. Animals were perfused transcardially approximately 24 hours after the lesions were made, using a 0.9% saline solution followed by 10% formalin. After removal from the skull, brains were stored in a 10% formalin solution for several days before sectioning. Brain sections of 40 µm were cut using a vibratome and were Nissl-stained to reconstruct the tracks and final positions of the tetrodes. This showed that recording sites ranged from 2.7 mm to 4.7 mm anterior to bregma, and were limited to the ventral and lateral orbital regions of the OFC. Recording depth ranged from approximately -3 mm to -5.5 mm (Paxinos and Watson, 2005; Fig. 1).

Data analysis

Behavior

Behavioral data was analyzed using SPSS for Windows (version 11.0). Unless otherwise stated, results are expressed as mean ± SEM values. Movement Time was defined as the interval between nose retraction from the odor port and nose entry
Figure 1. Localization of tetrode recording sites. As indicated by rectangles, recordings in all rats were localized in the ventral and lateral regions of the OFC, between 2.7 and 4.7 mm anterior from bregma. Recording depth ranged from approximately -3 mm to -5.5 mm (Paxinos and Watson, 2005). Several tetrode tracks are visible, as indicated by the black arrows. Black asterisks mark the lesion sites showing the final position of three tetrodes.

into the food trough, whereas the Overall Response Time was defined as the duration of the behavioral sequence starting with odor sampling and ending with the nose poke in the food trough. The mean response times per reward probability were obtained from all trial types associated with a particular probability from all sessions. These measures were compared across different trial types with the non-parametric Kruskall-Wallis test (P < 0.05), followed by a post-hoc Mann-Whitney U test (P < 0.05).

Single units
Single units were isolated and analyzed as previously described by Van Duuren et al. (2007). In short, spike sorting was done off-line using standard cluster cutting procedures (BBClust/MClust 3.0). Peri-event time histograms (PETHs) were constructed to examine correlations between events in the task and changes in firing rate. Neural responses during trials were statistically assessed with the non-parametric Wilcoxon matched-pairs signed-rank (WMPSR) test (P < 0.01). They were considered significant if firing rates during trials, quantified per bin, were significantly different from a fixed control (baseline) period during the inter-trial
interval (-5 to -4.5 s before trial onset). This control period consisted of five
consecutive bins, and any of the bins in the trial period tested for a significant
change in firing was required to differ significantly from each of these five control
bins. In addition, responses had to be significant for two binsize resolutions (i.e. 100
and 1500 ms) to be considered as such. Once the WMPSR test indicated a
significant deviation in firing rate with respect to baseline, the non-parametric
Kruskall-Wallis test (P < 0.05) and a post-hoc Mann-Whitney U test (P < 0.05)
were used to compare the different PETHs pertaining to the various odor-
probability pairs. For these and all other calculations, only three reward probabilities
were taken into consideration (i.e. P = 100, 75 and 50%), because the null
probability did not yield enough trials to perform robust analyses, except for a test
of neural coding of reward uncertainty (see Results section).

Variability in the population code for reward probability

To examine the variability in responses within the population towards reward
probability in more detail, we calculated two different measures of response
variability (Perez-Orive et al., 2002; Rolls and Tovee, 1995). Parameter variability
($V_{par}$), which is indicative of a single cell’s response variability attributable to
differences in reward probability, was calculated by:

$$V = \frac{N}{N-1} \frac{\bar{r}^2 - (\bar{r})^2}{\bar{r}^2}$$

(Eq.1)

with

$$\bar{r} = \frac{1}{N} \sum_{j=1}^{N} r_j$$

and

$$\bar{r}^2 = \frac{1}{N} \sum_{j=1}^{N} r_j^2$$

where N indicates the total number of reward probabilities (N = 3) and $r_j$ the
mean firing rate per cell per probability. In addition, we calculated the population
variability ($V_{pop}$), which is indicative of the variability in the mean firing rate of
single cells across the population, irrespective of the probability of reward. This
measure was calculated in a similar fashion, but $r_j$ now indicates the mean firing rate
of neuron $j$ during a particular trial phase, averaged across all three reward
probabilities, and N the number of units recorded in a given session. Thus, $\bar{r}$ now
represents the mean firing rate in the population and $\bar{r}^2$ the mean squared firing rate.
Values ranged between 0 and 1, with 1 representing the maximal variability attainable.

**Ensemble analysis of reward probability coding**

Representation of expected reward probability by ensemble activity was examined using template matching as reconstruction method. The sessions were divided into two blocks: the initial 3/4 of the trials was used for decoding, whereas the final 1/4 of the trials was used for encoding. Two vectors were constructed for each reward probability, denominated as \( x = (x_1, x_2, \ldots, x_N) \) and \( y = (y_1, y_2, \ldots, y_N) \), containing the spike counts within a specified time window for the encoding (\( xx \)) and the decoding block (\( yy \)), with \( i_x \) and \( i_y \) indicating the spike count of cell \( i \) averaged across trials. Thus, one vector is used for the encoding part of the procedure (which determines the “template”, i.e. the response profiles or ‘tuning curves’ of the cells towards reward probability; the response profiles consist of a list of the spike counts of all cells pertaining to the different reward conditions, calculated across trials 10-15). The other vector is used for the decoding part of the procedure, where the spike counts, specific for reward probabilities, are taken from the same cells, but now from the first part of the session, trials 1-9). The decoding vector is then compared to the encoding vector. Thus, these vectors are used to calculate the decoding score, which is the percentage of correctly identified reward probabilities in the decoding phase, based on the activity patterns found in the encoding phase. Hence the ensemble code for reward probability is made up of the different firing rates of all recorded cells combined in the en- and decoding phase in relation to reward probability. Note that, besides mean firing rate per trial phase, other aspects of firing patterns, such as related to spike timing, may make additional contributions to ensemble coding in general (cf. Narayanan et al., 2005).

A standard time window was used for the various trial phases for which reward probability was reconstructed, corresponding with the duration of that particular phase within the trial. The decoding time frame used for the period in which the animal moved from the odor sampling port to the food trough (the ‘movement period’) was 1 s, and the time frame for the waiting period at the food trough was 1.5 s. For the reward phase the decoding time frame was 5 s, unless otherwise noted.

Template matching was used as previously described by Lehky and Sejnowski (1990) and Zhang et al. (1998). The similarity (‘matching’) between the two vectors containing the spike count in the defined time window for the encoding and decoding block was calculated by computing the cosine of the angle between them. A value of 1 represents an exact similarity between the two vectors and -1 the exact opposite, whereas 0 (i.e. orthogonal) indicates no similarity between the two vectors. First, the inner product of \( x \) and \( y \) was calculated by:
\[
\sum_{i=1}^{N} x_i y_i \quad \text{(Eq.2)}
\]

where \( x_i \) and \( y_i \) indicate the average firing rate of neuron \( i \) from a total of \( N \) cells within the specified time window for the encoding and decoding block, respectively. The cosine value was calculated by:

\[
\cos \theta = \frac{\sum_{i=1}^{N} x_i y_i}{\|x\| \|y\|} \quad \text{(Eq.3)}
\]

with the denominator representing the product of the absolute vector lengths. Whenever the decoding spike vector belonging to a particular reward probability provided the highest cosine value with respect to the encoding vector, then that particular probability was selected as the reconstructed likelihood.

The decoding score (i.e. the percentage of trials in which the probability of reward was correctly reconstructed) was expressed as a function of time and of the size of the ‘reconstruction ensemble’, i.e. the group of neurons which was subsampled from the entire population and used for the calculations. The maximum size of the reconstruction ensemble was 27, which corresponds to the lowest amount of cells recorded in the sessions used for this analysis. Thus, all ensembles used for our population coding study contained at least 27 units. Calculations were made for each recording session separately, after which decoding scores were averaged across sessions. For the assessment of decoding as a function of size of the reconstruction ensemble, the decoding score was calculated a hundred times for each group size, each time with neurons randomly picked from the population recorded in that particular session. Decoding as a function of time was calculated similarly with a reconstruction ensemble of 27 neurons: the decoding score was calculated a hundred times per time window, each time with randomly picked neurons. The decoding curves were analyzed further by applying linear regression analysis (\( P < 0.05 \)) and a one-way ANOVA test with, if appropriate, a Bonferroni correction (\( P < 0.05 \)).

Besides template matching, we applied Bayesian reconstruction as a method to study population coding (Lehky and Sejnowski, 1990; Zhang et al., 1998). The decoding performance obtained with this method, however, were generally similar or slightly lower than for template matching, and therefore these results will not be discussed here.
Results

Behavior

For the analysis we used data from 19 recording sessions, obtained from 4 rats. Animals performed on average 32 trials for the three highest reward probabilities (P = 100%; 32.2 ± 2.2; P = 75%; 32.6 ± 2.9; P = 50%; 32.3 ± 2.1). For probability P = 0%, animals performed significantly fewer trials compared to the other three probabilities, on average 9.0 ± 2.3 (paired sampled t-test, for all three comparisons: P < 0.001; note that each odor-probability coupling was novel at the beginning of each session). Movement Time (the interval between nose retraction from the odor port and nose entry into the food trough) showed no significant difference between the probabilities P = 100, 75 and 50% (respectively 0.69 ± 0.01 s, 0.68 ± 0.01 s and 0.72 ± 0.01 s), but the Movement Time for each of these reward probabilities was significantly shorter than for the null probability (1.15 ± 0.05 s). Furthermore, examination of the Overall Response Time (the duration of the behavioral sequence starting with odor sampling and ending with the nose poke in the food trough) revealed that animals responded significantly faster on P = 100% and 75% trials as compared to P = 50%; no significant difference was found between P = 100% and P = 75% (P = 100%; 2.43 ± 0.03 s, P = 75%; 2.39 ± 0.03 s and P = 50%; 2.61 ± 0.04 s). Thus, learning within this task was evident from the shorter Overall Response Time for the two highest reward probabilities, as well as from the lower amount of trials and slow responding for the P = 0% reward condition.

Electrophysiology

Single units: neural correlates of task events

During the 19 recording sessions, a total of 541 single units was recorded in the OFC, with a firing rate of 1.30 ± 0.07 spikes/s (mean ± SEM). Of these 541 units, 129 (24%) showed 177 statistically significant responses during the task, which implies that a considerable proportion of cells exhibited more than one correlate. Task-related modulations were observed in neurons that responded during sampling of odors, during the behavioral period in which animals moved from the odor sampling port towards the food trough, during the waiting period at the trough, and after pellet delivery (Fig. 2).

During odor sampling, 38 neural correlates (21%) were found (Fig. 2A). These responses consisted mainly of increments in firing rate (95%, n = 36); 2 responses (5%) consisted of a decrement. About half of the cells displaying an increase in activity (47%, n = 17) started firing already within the 500 ms period prior to odor presentation (when the nose was already in the sampling port). This activity probably reflects preparatory behavior related to odor sampling or
anticipation of odor delivery. For this group of 38 correlates we did not examine whether the activity changes were affected by (upcoming) reward probability, since the design of our task did not provide the possibility to determine whether modulations in firing activity were actually due to odor identity or to the associated reward probability.

After successful odor sampling, animals moved towards the food trough to obtain a reward. During this movement period, neurons displayed significant changes in firing rate as well: 18 correlates (10%) consisting of an increase in firing rate were found (Fig. 2B), whereas cells did not show decrements during this period. During the subsequent waiting period, when animals anticipate reward in an immobile state, 53 responses (30%) were found, of which 50 (96%) displayed an increase in firing activity (Fig. 2C), whereas the remaining 3 neurons (4%) showed a decrease.

Figure 2. Overview of behavioral correlates of neural activity changes observed during task performance. Peri-event time histograms and raster plots showing examples of the observed task-related behavioral correlates. Examples from four different units recorded in four different sessions demonstrating correlates related to (A) odor sampling (synchronized on onset of odor presentation during P = 100% trials) (B) movement activity preceding nose entry into the food trough (synchronized on entry of the food trough during P = 50% trials), (C) the waiting period of 1.5 s with nose in the food trough, synchronized on onset of waiting and (D) pellet delivery (both during P = 100% condition). These as well as the following histograms (Fig. 3, 4 and 5) are presented with a binsize of 100 ms. In all raster plots, individual consecutive trials are represented as horizontal lines, with the first trial at the top row. Horizontal scale denotes time (s), vertical scales firing rate (Hz).
For the reward delivery period, neural correlates were examined in relation to rewarded or unrewarded trials. During this period, 69 significant neuronal responses (39%) were observed (Fig 2D), which could be divided into three subgroups. The first subgroup consisted of 32 neurons (47%) that specifically responded during the rewarded, but not unrewarded trials (Fig. 3A; note that additional correlates besides the main reward-response could be present). Within this group, 7 neurons (22%) showed a decrement in firing activity, whereas the remaining 25 neurons (78%) displayed an enhancement. The second subgroup consisted of 30 neurons (44%) that responded during both rewarded and unrewarded trials with a significant increase in firing activity (Fig. 3B). Comparison of these neuronal responses with a Mann-Whitney U test (P < 0.05) indicated that within this group, 19 neurons (63%) demonstrated differential firing activity towards the rewarded and unrewarded condition: responses were found to be either larger for the rewarded condition (37%; n = 7) or unrewarded condition (63%; n = 12).

Figure 3. Differential firing after reward delivery during the rewarded and unrewarded condition. Activity in rewarded trials is synchronized on pellet delivery, in unrewarded trials at the same time point as pellet delivery in rewarded trials. A. Example of a unit demonstrating an increase in firing activity solely during the rewarded trials. B. Activity of a different unit showing a double correlate: both during the reward and waiting period (starting at -1.5 s) this neuron showed a significant increase in firing activity. During the reward phase, the increase in firing rate during rewarded trials was significantly larger as compared to unrewarded trials. No difference was found between the two conditions in the waiting period. C. This unit increased firing specifically during unrewarded trials during the task. Activity during the inter-trial interval (ITI) was absent. Horizontal scale denotes time (s), vertical scale firing rate (Hz).
For the remaining 11 neurons (37%) of this second subgroup, no difference in firing activity between the two conditions was found. The third group of responses observed in the reward period consisted of 7 neurons (9%) that increased their firing activity during unrewarded, but not rewarded trials (Fig. 3C). To examine whether these neurons may encode a negative error in reward prediction (cf. Schultz et al., 1997) we tested whether they exhibited the same response during nose-poking behavior during inter-trial intervals (ITIs). Because no odor cues were provided and nose-poking is presumably habit-driven, we predicted that reward expectation would be absent or at least less pronounced during the ITI period. Indeed, none of these neurons increased their firing rate in the absence of reward in the inter-trial interval (Fig. 3D). This indicates that the activity of these neurons reflects the omission of reward within the task context and not motor behavior associated with visiting and departing from the reward site. To examine the possibility that the observed firing responses during the reward phase covary with motor activity due to food consumption, we observed the chewing behavior of the animals during recordings by means of visual oscilloscope inspection. Chewing was found to outlast the duration of the observed neural responses after pellet delivery. Hence, these responses may be triggered by reward consumption, but do not appear to correlate with the overall phase of food intake.

**Single units: modulation of firing rate by expected reward probability**

Whether neurons fired differentially in anticipation of the three reward probabilities was examined for the movement period preceding the food poke and for the waiting period. Neural activity within these two periods is likely to reflect expectation of reward (Schoenbaum et al., 1998; Van Duuren et al., 2007), and hence this activity may be modulated by the probability of upcoming reward.

Of the 71 correlates observed during both these two task phases, a total number of 17 (24%) demonstrated statistically significant differences in firing towards the three reward probabilities. During the movement period, 3 neurons out of the 18 neurons showing a response in this period (17%) appeared to discriminate between two different reward probabilities: firing activity was found to be largest for either the highest (n = 2) or lowest reward probability (n = 1) (Fig. 4A and C). We also examined whether the observed correlates did not reflect ‘go’ movements of the animal per se, but were actually related task-related. To this end, we compared firing-rate changes observed during trials with activity changes when animals executed the same behavior in the inter-trial interval. A total number of 16 responses (89%) were found to be task-related, because these neurons did not show a significant correlate during the inter-trial interval (Fig. 4B). In 2 units (11%), a significant response occurred during the inter-trial interval at similar time points as during correct ‘go’ trials.
Figure 4. Differential firing in relation to expected reward probability during the movement period. A. Example of a single unit demonstrating an increase in firing rate during the movement period, i.e. the period after leaving the odor port and before nose entry into the food trough. Activity is synchronized on nose entry into the food trough ('food poke'). During movement, this single unit discriminated significantly between the $P = 50\%$ condition and the $P = 75$ and $100\%$ condition, showing the largest increase in firing activity for $P = 50\%$. No significant difference was found between the $P = 75$ and $100\%$ condition. In addition, this cell demonstrated a correlate in the reward period specifically during rewarded trials, visible in the histogram as a peak in firing rate at $\sim 2$ s after food poke onset. B. Activity of another neuron during the movement period as part of task performance ($P = 75\%$ condition) and during the same behavior in the inter-trial interval (ITI). Activity is synchronized on nose entry into the food trough; horizontal scales denote time (s), vertical scales firing rate (Hz). C. Overview of the three significantly different profiles related to reward probability in the movement period. Different units are represented by different symbols. On the horizontal scale reward probability (%) is plotted; the vertical scale displays the peak firing rate of individual units in association with different reward probabilities, normalized to the rate in the $P = 100\%$ condition.

Considering the absence of notable changes in sensory input and the animal’s overall immobility, the waiting period is a task phase suitable for studying whether neurons may code predicted outcome probability as well. Within the group of 53 neurons that showed a significant response during waiting, 14 neurons (27%) demonstrated differential firing towards the probability of reward, which was either
between two ($n = 8; 57\%$) or three reward probabilities ($n = 6; 43\%$). Neural activity was found to increase ($n = 4; 28\%$) or decrease ($n = 5; 36\%$) with increasing probability, or units displayed the largest or smallest response to the middle probability ($n = 5; 36\%$) (Fig. 5).

**Figure 5.** Differential firing in relation to expected reward probability during the waiting period. A. Example of a unit showing differential firing towards all three reward probabilities during the waiting period. Activity is synchronized on nose entry into the food trough. This unit discriminated between all three probabilities, with the highest response towards the $P = 75\%$ condition and the lowest response to $P = 50\%$. Horizontal scales denote time (s), vertical scales firing rate (Hz). B. Overview of the significantly different firing rate profiles towards reward probability found during the waiting period. Different units are represented by different symbols. On the horizontal scale reward probability (%) is plotted; the vertical scale displays the peak firing rate of individual units in association with different reward probabilities, normalized to the peak in the $P = 100\%$ condition.

**Variability of the representation of reward probability**

To examine the extent to which firing-rate changes throughout the various trial phases and across the recorded population are attributable to reward probability we calculated two measures, parameter variability ($V_{par}$) and population variability ($V_{pop}$). The time windows used for this calculation were 1 s for the movement period and 1.5 s for both the waiting and reward delivery period. The mean $V_{par}$, which is indicative of the response variability of the individual neurons associated with
variations in reward probability, was 0.29, 0.22 and 0.27 for the movement, waiting and reward delivery period, respectively (Fig. 6A - C). The mean $V_{\text{pop}}$, expressing the variability in firing rate across the population regardless of reward probability, was 0.76, 0.71 and 0.71 for the movement, waiting and reward delivery phase, respectively (Fig. 6D - F). These results demonstrate that there is a high variability in firing rates present across the population, but also that firing rates of individual neurons within OFC are modulated by predicted reward probability to a generally modest degree. However, in all three trial periods a subset of neurons was present that showed a very high degree of modulation by reward probability (parameter variability, range 0.9-1.0; Fig. 6A - C).

![Figure 6](image)

**Figure 6.** Distribution of population and parameter variability for the movement period (A and D), the waiting period (B and E) and the reward period (C and F). On average, variability related to parameter covered a broader range of values than population variability, spanning the whole range from 0.0 to 1.0 (average 0.29 and 0.22 and 0.27 for the movement, waiting and reward period, respectively). Variation in population variability was less, with values ranging from 0.5 to 1.0 (average 0.76, 0.71 and 0.71 for the movement, waiting and reward period, respectively).

**Population coding of expected reward probability**

The results described above indicate that activity of a restricted subset of orbitofrontal neurons reflects expectancy of reward coming up with a specific probability. We next asked whether not only single neurons, but also the whole population of recorded neurons codes information regarding this reward parameter. Answering this question may shed light on how target regions of the OFC may read
out information represented at the population level. While the analyses above suggest a probability representation at the single cell level, ensembles may not show a robust representation when, for instance, noise or other types of response variability obscure single cell contributions. To this end, we decoded reward probability from the population activity for the three trial phases under examination, using template matching as reconstructing algorithm. For this analysis, 8 sessions with the largest amount of simultaneously recorded cells were used (n = 338), with a minimum of 27 cells. It should be noted that the study of probability coding in the reward delivery period primarily subserves the purpose of comparison to the motor and waiting periods. If an orbitofrontal neuron fires more vigorously on rewarded versus non-rewarded trials (which is often the case as indicated by the single unit data), its accumulated spike counts will naturally come to correlate with reward probability because often-rewarded trial types will elicit more spikes than rarely rewarded types, whereas no specific coding of probability can be said to exist.

The probability of reward could be reconstructed from ensemble activity during all three trial periods, with a percentage correct significantly above the 1/3 chance level (one-way ANOVA: P < 0.001 in all three cases). Plotting the decoding score as a function of size of the reconstruction ensemble showed that for all trial phases, performance improved with an increasing amount of cells, with the slope of the decoding curve being significantly positive (linear regression; in all cases P < 0.001). The highest decoding scores obtained within these periods were 46% for the movement period (at n = 25 cells), and 48% (n = 27) and 44% (n = 23) for the waiting and reward period, respectively (Fig. 7). Whereas the curves for the movement and waiting periods both showed a gradually rising decoding success when ensemble size increased, the curve for the reward period rose more steeply at low cell counts, after which decoding success saturated around ensembles sizes of about 8 and higher. This difference suggests a higher redundancy of neural coding in the reward period as compared to the anticipatory phases.

Temporal resolution of ensemble coding

To examine the temporal resolution of ensemble coding in more detail during these three trial phases, the decoding score was calculated using time windows of increasing duration, but shorter than the standard time frames used above. This provides a measure of how decoding performance increases as more and more time within each trial phase is taken into account. When decoding was calculated as a function of time, again with 27 cells in the ensemble, the maximal scores obtained were 47% for both the movement and waiting period and 45% for the reward delivery phase (Fig. 8).
Figure 7. Decoding of reward probability with an increasing number of neurons for the (A) movement, (B) waiting and (C) reward period. Decoding time windows used were 1, 1.5 and 5 s, respectively. The horizontal axis indicates the size of the reconstruction ensemble, the vertical axis the percentage of trials in which reward probability was correctly decoded. The horizontal dashed line indicates chance level (33.3%) and dotted lines flanking the curves represent the 95% confidence interval (two times the standard error of proportion).
Linear regression analysis indicated a significant improvement in decoding score as the time window increased, both for the waiting and reward phase (P = 0.013 and P = 0.028, respectively), but not for the movement phase. A one-way ANOVA indicated a significant decoding score with respect to chance level for the time window of 1 s in the movement period, whereas for the waiting period this was obtained for a width of 1.25 and 1.5 s. Finally, for the reward period, all time frames had decoding scores significantly above chance level, except for the time frames of 0.25, 0.75 and 1 s. In addition, in all three phases the differences between the various time frames failed to reach significance. Overall, these results indicate that information on anticipated reward probability gradually accumulates during trial phases, and reaches significance only after relatively long time segments of trial phases have elapsed, viz. about 1 s or longer.

The null reward probability condition

Reward probability and reward uncertainty, although being considered fundamentally different reward parameters (Dreher et al., 2006; Fiorillo et al., 2003, Tobler et al., 2007), are intimately linked in the sense that with the two extreme probabilities (P = 0 and 100%) uncertainty is absent, whereas with the intermediate probabilities uncertainty increases, being maximal in the P = 50% condition. In order to examine whether the observed neural activity during the movement and waiting period reflects reward probability or reward uncertainty, we examined probability reconstruction success by using unrewarded trials (P = 0%) for encoding. A first hypothesis holds that whenever reward probability is coded by OFC ensembles, encoding by unrewarded trials and decoding by P = 50% trials should yield decoding above chance, because the global difference in reward probability is smaller for these 2 trials types than P = 0% versus P = 75 and P = 100% trials. In case OFC would code uncertainty, however, one expects that encoding by unrewarded trials and decoding by P = 100% trials result in a decoding score above chance, because these two reward conditions are more alike in terms of uncertainty than P = 0% versus P = 75 % and P = 50 % trials. A third hypothesis holds that in this procedure decoding scores for P = 50, 75 and 100% should be random (33.3% success) because the P = 0% condition is set within a different trial type (as signaled by a distinct odor), without carrying over any quantitative information about reward probability to other trial types.

As illustrated in Fig. 9, decoding for the movement and waiting periods was similar in that P = 100% trials yielded significant below-chance scores, while P = 50% trials were significantly above chance (one-way ANOVA; P < 0.001). However, for the P = 75% condition decoding was at a random level in the movement period but gradually decreased below this level in the waiting period. When decoding scores for these two trial periods was averaged across all reward probabilities, performance was
around chance level (data not shown). These results indicate that the observed variations in population activity during these trial phases are attributable to reward probability rather than uncertainty.

**Figure 8.** Decoding scores using time windows of increasing duration within (A) the movement, (B) waiting and (C) reward period. The size of the reconstruction ensemble was 27 neurons and onset of each window coincided with the onset of that period. The horizontal axis shows the width of the time window (s) from which spikes were taken for reconstruction, the vertical axis the percentage of correctly decoded trials. As the duration of consecutive trial phases was different, step sizes and maximal widths of time windows varied per trial phase. Black asterisks mark significance with respect to chance level (33.3%; one way ANOVA). The horizontal dashed line indicates chance level and dotted lines flanking the curves represent the 95% confidence interval (two times the standard error of proportion).
As illustrated in Fig. 9, decoding for the movement and waiting periods was similar in that $P = 100\%$ trials yielded significant below-chance scores, while $P = 50\%$ trials were significantly above chance (one-way ANOVA; $P < 0.001$). However, for the $P = 75\%$ condition decoding was at a random level in the movement period but gradually decreased below this level in the waiting period. When decoding scores for these two trial periods was averaged across all reward probabilities, performance was around chance level (data not shown). These results indicate that the observed variations in population activity during these trial phases are attributable to reward probability rather than uncertainty.

**Figure 9.** Decoding scores for the various reward probabilities with an increasing number of neurons during (A) the movement period and (B) the waiting period when spike vectors from unrewarded trials ($P = 0\%$) were used for encoding. In both task phases decoding for $P = 50\%$ trials was significantly above chance level (33.3%), whereas the score for $P = 100\%$ trials was below chance level. Decoding performance for $P = 75\%$ was either at chance level (movement period) or decreased below chance with increasing ensemble sizes (waiting period). When the decoding performance was averaged across all 3 probabilities, curves were at chance level for both trial phases (not shown).
Contribution of individual cells to coding of reward probability

In calculating the performance of probability reconstruction from population activity, decoding scores are averages across 100 groups of randomly selected neurons. This, however, does not provide insight in the contribution of individual neurons to an ensemble code for reward probability. To acquire more insight into the redundancy versus sparsity of coding, we calculated for all neurons the difference in the percentage of decoded information when a specific cell was added to a group of five neurons randomly selected from the same session. For each cell this calculation was done a hundred times, each time with a new randomly selected group of five additional neurons. Apart from the consideration that single cells may contribute reasonably to coding by such a relatively small group (Fig. 7), this size was chosen arbitrarily.

Since for all three trial phases this analysis yielded similar results, we only provide the data for the waiting period. During this period, 25% of the cells (n = 85) made a minimal contribution to the decoding success (between -0.5 and 0.5%); 40% (n = 135) made a positive contribution (> 0.5%) (average ± SEM: +5.5 ± 0.4%), whereas 35% (n = 118) made a negative contribution (< 0.5%) to the reconstruction, with an average of -5.2 ± 0.5% (i.e. addition of these cells led to a decrease in correct decoding). The average positive contribution did not differ significantly from the negative one, as examined with an unpaired t-test. This lack of significance agrees with the absence of a net positive slope in the reconstruction curve at an ensemble size of 5 (Fig. 7B). We also examined the percentage of cells showing an extremely strong contribution (>15% or < -15%). This showed that only 2% of the cells (n = 7) made such extreme (positive or negative) contributions, with an average contribution of +20 ± 1.4% and -20 ± 2.0%, respectively. Further inspection of the distributions of single cell contributions confirmed that there was no particular subset of cells contributing especially to the coding of reward probability, and that the positive and negative contributions were nearly symmetrically distributed around zero, which is altogether consistent with a distributed representation across populations that contain cells making highly variable contributions.

Discussion

The OFC has been strongly implicated in decision-making under uncertainty (Bechara et al., 1996, 1997). Here we examined to our knowledge for the first time whether and how single-cell and population activity within OFC is affected by likelihood of future reward. The behavioral results showed animals to respond significantly faster during trial types with the two highest reward probabilities.
compared to the lower reward probabilities, indicating that they responded differentially depending on predicted reward probability.

During the waiting period, 27% of the cells demonstrating expectancy-related activity showed differential firing towards the various reward probabilities. This discriminatory activity was represented at the population level as well; predicted reward probability could be reconstructed from ensemble activity significantly above chance level. Although the overall decoding scores may seem rather low, the task required the animals to learn novel odor-probability associations each session, and in order to estimate reward probability, animals will need to accumulate experience across many trials. Moreover, there was no task requirement necessitating the animal to discriminate expected probabilities, as the chance of obtaining reward was not contingent upon the speed of ‘go’ responses.

In addition to the waiting period, expectancy-related activity might also occur earlier within the trial (Van Duuren et al., 2007), for example when animals moved from the odor sampling port to the food trough. During this period, 13% of the neurons showing a task-related correlate were found to discriminate between different reward probabilities. That this activity is most probably reward-related and not confounded by motor activity is supported by the absence of a neural response when the same behavioral sequence was executed during the inter-trial interval (Fig. 4). The lack of a difference in Movement Time between the reward probabilities also supports the idea that this differential activity is attributable to the different reward contingencies instead of differential motor activity during ‘go’ responses. Reward probability could be reconstructed from ensemble activity in this trial phase as well, with a decoding score comparable to the score obtained during the waiting period (Fig.7).

In the reward delivery period, 10% of the neurons were found to respond during either the rewarded trials or unrewarded trials, or during both. A remarkable subgroup consisted of cells that specifically responded upon omission of reward within trials. Given that responses were absent upon food pokes outside the trial context, we propose that these units code a violation of a positive reward expectancy. Whether they would encode an error in reward prediction in general (cf. Schultz et al., 1997) remains to be investigated, using e.g. unexpected reward delivery. The probability of reward could be reconstructed from population activity in this trial period with a performance comparable to the other two trial phases (decoding score of 44%). However, when instead of reward probability the availability of reward was reconstructed from population activity, the decoding score went up to 89% (n = 27) (data not shown). This indicates that during the reward period the presentation of a reward is coded more reliably than the overall reward probability. This need not be surprising given that neural activity during this period may be determined by
processes other than ‘tracking’ actual reward probability, for example taste sensations or ingestion, which are closely related to processes of reward appraisal.

**Variability and distribution of the representation of reward probability**

Although single units showed differential firing towards varying reward probability, this leaves unanswered the question whether probability of reward is represented in a sparse or redundant manner within OFC, i.e. by a few highly specifically tuned cells or in a broadly distributed way. Therefore, we examined the firing-rate variability attributable to probability, and found that this measure was relatively modest in comparison to the overall variability in mean firing rate across the population. Secondly, decoding scores were found to depend only weakly on the size of the reconstruction ensemble, and removal of cells displaying the largest variability in their response towards probability (i.e. parameter sparseness between 0.9 and 1.0; Fig. 6) resulted in a decoding score for both periods that did not differ significantly from the decoding curves obtained using the entire population (data not shown). Thirdly, considering the widely dispersed single cell contributions to the decoding score, with nearly symmetrical distribution of positive and negative values, these results indicate that reward probability is coded in a broadly distributed manner within OFC. However, because the decoding score did not rapidly saturate when cell count increased in pre-reward periods (fig.7A-B as opposed to C), coding does not appear to be highly redundant in these phases, but instead cells make variable contributions to it.

It is important to address whether and how a distributed representation of probability may be utilized by other brain structures targeted by the OFC to guide behavior and attention. How these structures integrate population signals into adaptive behavioral decision-making is largely unknown, but is of note that feedforward or recurrent networks of units with broad tuning curves can extract sensory, motor or motivational variables from a source population of noisy neurons (Deneve et al., 1999; Zhang et al., 1998). Such networks may be implemented in target structures of OFC such as higher associational cortical areas or corticostriatal circuits (Uylings et al., 2003; Voorn et al., 2004). Notably, network architectures capable of sustaining continuous attractors can read out population activity by a natural form of template matching (Wu and Amari, 2005). Prefrontal output also reaches mesencephalic DA cells (Phillipson 1979; Uylings et al., 2003; Van de Werd and Uylings, 2008), potentially supporting the generation of phasic reward-prediction errors and of more tonic signals representing reward uncertainty (Fiorillo et al., 2003).
Functional implications

It is still unclear whether parameters related to expected reinforcement or utility are coded by neural activity subserving the role of a ‘common currency’ within the OFC or elsewhere in the brain, i.e. whether neurons code a lumped measure of expected utility in which all relevant parameters (such as delay, magnitude, uncertainty) have been counted in (Montague and Berns, 2002; Kalenscher and Pennartz, 2008). As suggested by Roesch et al. (2006), coding of time-discounted rewards in rat OFC seems independent of the coding of absolute reward value. In contrast, earlier findings by Roesch and Olson (2004) in primate OFC indicated that neurons do code reward value in a common currency: single unit activity elicited by visual cues associated with differently delayed or sized rewards was shown to covary with both parameters. As demonstrated by the current study and a previous analysis of reward size (Van Duuren et al., 2007), the probability of future reward is coded in a similar fashion within OFC as is the case for reward magnitude. Both parameters are represented in a distributed fashion by neurons that display a large diversity in tuning and parameter sensitivity. If independent parameter coding would turn out predominant for single neurons, it is still possible that larger OFC ensembles act as functional entities coding a ‘common currency’. The finding that both reward probability and magnitude exert modest modulatory effects on single cells, and that parameter information appears to be represented in a widely distributed form, suggests that the ensemble level is at least as relevant for studying the ‘common currency’ problem as the single unit level. Whenever the OFC is important for probability discounting, as suggested by behavioral studies (Bechara et al., 1996, 1997; Mobini et al., 2002; Pais-Vieira et al., 2007), one should expect codes for probability and magnitude to co-vary at least at the population level, since this provides the OFC with the opportunity to take into account the effect of reward probability on the value that a particular reward represents.

In a recent fMRI study by Dreher et al. (2006), it was demonstrated that in a task in which reward probability was systematically varied, activity in human OFC covaried with a negative error prediction signal at the time when an expected reward was omitted. The idea that OFC may code errors in reward prediction is supported by our observation that some orbitofrontal neurons responsive in the reward period increased their firing rate solely during trials where an expected reward was omitted. Furthermore, regarding the distinction between probability and uncertainty of reward, fMRI findings by Tobler et al. (2007) suggested that reward uncertainty rather than probability is coded within orbital areas. This can be interpreted as contrasting with our finding that reward probability, but not uncertainty, is coded in OFC (Fig. 9). However, BOLD signals as observed with fMRI are not considered to reflect the spiking output of a particular brain area, but rather the external synaptic inputs and local intracortical processing (Logothetis et al., 2001). Hence, changes in
BOLD signal are taken to globally correspond to changes in input activity, but not to the actual recruitment of neurons in the area under investigation. Whereas in the current study neurons were found to either decrease or increase their firing activity during trial periods in a parallel-distributed way, the limited sensitivity and spatiotemporal resolution of BOLD signals may explain why such modulations have thus far not been detected with fMRI.

**Acknowledgements**

This work was supported by NWO Grant 903-47-084, ZonNW (NWO) grant 912.02.050, NWO grant 918.46.609 and BSIK (SenterNovem) grant 03053. We would like to thank David Redish and Peter Lipa for providing the cluster cutting software, Els Velzing for help with graphical illustrations and Francesco Battaglia, Jadin Jackson and Tobias Kalenscher for their comments on the manuscript.
Chapter 5

Pharmacological manipulation of neuronal ensemble activity by reverse microdialysis in freely moving rats: a comparative study of the effects of Tetrodotoxin, Lidocaine and Muscimol
Abstract

To be able to address the question how neurotransmitters or pharmacological agents influence activity of neuronal populations in freely moving animals, the combidrive was developed. The combidrive combines an array of 12 tetrodes to perform ensemble recordings with a moveable and replaceable microdialysis probe to locally administer pharmacological agents. In this study, the effects of cumulative concentrations of tetrodotoxin (TTX), lidocaine and muscimol on neuronal firing activity in the prefrontal cortex were examined and compared. These drugs are widely used in behavioral studies to transiently inactivate brain areas, but little is known about their effects on ensemble activity and possible differences between them. The results show that the combidrive allows ensemble recordings simultaneously with reverse microdialysis in freely moving rats for periods at least up to two weeks. All drugs reduced neuronal firing in a concentration dependent manner, but differed in the extent to which firing activity of the population was decreased and in speed and extent of recovery. At the highest concentration used, both muscimol and TTX caused an almost complete reduction of firing activity. Lidocaine showed the fastest recovery, but resulted in a smaller reduction of firing activity of the population. From these results it can be concluded that whenever during a behavioral experiment a longer lasting, reversible inactivation is required, muscimol is the drug of choice, since it inactivates neurons to a similar degree as TTX, but does not, in contrast to TTX, affect fibers of passage. For a short-lasting, but partial inactivation, lidocaine would be most suitable.
Introduction

Until recently, neurophysiological analysis of information processing in the brain was primarily based on the examination of firing activity of single cells during behavior, as measured with repetitive presentations of stimuli (Gerstein and Kiang, 1960). This however could not provide an answer to the question how information is represented by the pattern of activity distributed across a population of neurons. With the emergence of techniques to record large numbers of neurons simultaneously ("ensemble recordings"), it became possible to examine information coding at the level of cell populations (Wilson and McNaughton, 1993). However, an issue that has not been addressed thus far is how neurotransmitters influence activity of these cell populations. To gain more insight in the interaction between neurotransmitters or pharmacological agents and neuronal firing activity, we sought to develop a method in which drugs could be locally administered while performing ensemble recordings in freely moving rats. As drugs should ideally be delivered with a constant concentration throughout the experimental session within the entire recording area, reverse microdialysis is preferred over either local injections, since with injections additional fluid is introduced into the brain causing a change in pressure, or iontophoresis, with which only a very small area can be reached.

The combination of (reverse) microdialysis with extracellular electrophysiological recordings in vivo was initially applied in research concerning hypoglycaemia, the pathophysiology of cerebral ischaemia and epilepsy. In those studies, performed in freely moving or anaesthetized animals, a single recording electrode for monitoring the EEG was glued next to the dialysis probe (Vezzani et al., 1985; Sandberg et al., 1986; Ludvig et al., 1992), inserted in the inflow tubing (Obrenovitch et al., 1991) or placed within the proximity of the dialysis probe (Tossman et al., 1985). Later on, microdialysis/electrode devices were developed that were suitable to perform single unit recordings in freely moving rats, cats and monkeys (Ludvig et al., 1994, 2000; Dudkin et al., 1994; Sakai and Crochet, 2000). The devices used in rats consisted of a fixed microelectrode array positioned next to a guide in which the dialysis probe was fitted (Ludvig et al., 1994; Brazhnik et al., 2004). Although these studies did not present data concerning the effect of perfusion per se on neuronal activity, the results did show the possibility of recording single units and influencing their activity by drug administration. A recent study in anaesthetized animals in which reverse microdialysis was combined with intracellular recordings furthermore demonstrated that effects of dialysis on the membrane properties, excitability and ongoing synaptic activity of neurons in the vicinity of the probe are minimal (West et al., 2002). Hence, this suggests that the technique of reverse microdialysis is suited to be combined with a multi-tetrode
array (Gothard et al., 1996) to conduct local pharmacological interventions during ensemble recordings.

To this end the combidrive was developed, a multi-tetrode array consisting of a circular row of 12 individually movable tetrodes and 2 reference electrodes surrounding a movable microdialysis probe. Unlike the already existing recording devices, this design should allow usage for several weeks after implantation, since the dialysis probe can be replaced if necessary. Furthermore, the electrophysiological recordings are performed with a multi-tetrode array instead of single electrodes, keeping the advantages of tetrodes in isolating single-units and yielding high numbers of cells (Gray et al., 1995; McNaughton et al., 1983; Recce and O’Keefe, 1986). During assessment of combidrive functioning several issues were addressed, including whether firing activity of neurons in the prefrontal cortex would be affected by perfusion per se and if the dialysis probe could be replaced without loss of recording capacity. The combidrive was applied in a comparative study of three drugs that all exert an inhibitory effect on neuronal activity, but differ in mechanism of action and physical-chemical properties, namely the sodium-channel blockers lidocaine and tetrodotoxin and the GABA\(\alpha\) agonist muscimol. Although these drugs are widely applied in behavioral studies to transiently inactivate selective brain areas (Albert & Mah, 1973; Ivanova & Bures, 1990; Brioni et al., 1989), little is known about the dynamics of their inhibitory effect in relation to population activity in awake animals. For example, these drugs are known to differ in duration of the inhibitory effect (Boehnke and Rasmusson, 2001), but it remains unclear to what extent neurons within a population and the neuronal population as a whole respond to the various drugs. Furthermore, a comparison of the effects of the drugs on neuronal activity within freely moving animals has not been made until now.

Materials and Methods

Subjects

All experiments were approved by the Animal Experimentation Committee of the Royal Netherlands Academy of Arts and Sciences and were carried out in agreement with Dutch Law (Wet op de Dierproeven, 1996) and European regulations (Guideline 86/609/EEC). Data were collected from 4 male Wistar rats (Harlan CPB; Horst, The Netherlands), weighing 360-425 g at the time of surgery. Animals were housed in standard type 4 macrolon cages, weighed and handled daily, and kept under a reverse 12 hr light/dark cycle (dimmed red light at 7:00 AM) with free access to food and water (standard rat chow; Hope Farms, The Netherlands). After surgery animals were housed individually in a larger cage (1 x 1 x 1 m) under the same conditions.
Construction of the combidrive and microdialysis probes

The combidrive presented here was custom-built at the NIN and was adapted from the multi-electrode drive array ("hyperdrive") as described by Gothard et al. (1996). The bundle of this hyperdrive, which contains 12 individually movable tetrodes and 2 reference electrodes, was modified to fit a microdialysis probe in the center. This resulted in a design in which the dialysis probe is surrounded by a circular row of 12 tetrodes and 2 reference electrodes, with a distance of 550 µm between the surface of the dialysis probe and the centre of a tetrode. The total weight of the combidrive was 32 g, the height (without guide tube for the probe) 4.4 cm, and the diameter at the site where the connectorboard was attached 3.3 cm (Fig.1).

Tetrodes were constructed as previously described by Gray et al. (1995). Briefly, four nichrome wires (diameter 13 µm; Kanthal, Palm Coast, Florida, USA) were twisted together, and a microbundle was formed by melting the polyimide coating with a heat gun. Electrode tips were goldplated with the use of a goldcyanide solution (Select Plating, Meppel, The Netherlands) to achieve an impedance range of 0.1-1.0 MΩ.

Figure 1. The combidrive (design based on Gothard et al., 1996) A. A close-up of the combidrive without its protective covers; a) screw which drives the up- and downward movement of the guide for the microdialysis probe, b) one of 14 drive screws that allow individual (vertical) movement of the tetrodes, c) individual wires of a tetrode connecting to the printed circuit board (d), e) bundle through which tetrodes and microdialysis probe exit the combidrive. B. Overview of the entire combidrive showing; e) the bundle, f) protective cover, g) screwed joint connecting the microdialysis probe to the guide, h) in- and outlet of the microdialysis probe, i) guide for the microdialysis probe. C. Magnification of the bundle showing 14 tetrodes, surrounding the microdialysis probe.
To be able to reach all subareas of the prefrontal cortex (including the ventral part), concentric dialysis probes with a length of 8 cm were constructed from two pieces of fused silica (i.d. 0.075 mm, o.d. 0.150 mm) that were inserted into a wider piece of fused silica (i.d. 0.320 mm, o.d. 0.430 mm) (Aurora Borealis Control, Schoonebeek, The Netherlands). Both pieces protruded approximately 1 cm on one side of the outer fused silica to serve as in- and outlet, which were both protected by 25G needles. On the opposite side, the part of the probe entering the brain, one piece of the inner fused silica protruded 3 mm, whereas the other piece remained 1 cm inside. A Hospal AN69 membrane (i.d. 0.240 mm, o.d. 0.320 mm) with a recovery over the membrane of about 10-15% was closed with (two-part epoxy) glue and fitted over the inner and in the outer fused silica. An exposed length of approximately 2 mm was used for dialysis. To fit the dialysis probe in the combidrive, a seven cm long stainless steel guide tube (i.d. 0.5 mm, o.d. 0.9 mm) was inserted along its central axis. To attach the dialysis probe to this guide tube, a coupling nut (o.d 2.5 mm) was glued to the outer fused silica just beneath the 25G needles; together with the thread on the guide tube this nut formed a screwed joint. By means of a worm-gear transmission, the guide tube can be lowered and raised, resulting in the down- and upwards movement of this guide. Whenever the tube is maximally lowered it fits in the central channel of the bundle and does not exceed the length of the bundle. Hence, the guide tube is prevented from entering the brain, whereas in this position the probe has reached its maximum depth in the brain. Replacement of the probe is achieved by raising the probe out of the brain and removing it from the guide, after which a new probe can be inserted and lowered again.

**Surgery and electrophysiology**

Animals were anaesthetized with 0.08 ml/100 g Hypnorm i.m. (0.2 mg/ml fentanyl, 10 mg/ml fluanison) and 0.04 ml/100 g Dormicum s.c. (midazolam 5 mg/ml) and mounted in a Kopf stereotaxic frame. After the incision additional local anaesthesia (Xylocaine spray; 10%, Astra) was applied to the skull as well. Body temperature was maintained at 37.5 ºC using a heating pad. After exposure of the cranium, 6 small holes were drilled into the cranium to accommodate surgical screws, one of which served as ground. Another larger hole was drilled over the prefrontal cortex in the left hemisphere (centre of the hole 3.6 mm anterior, 3.2 mm lateral to bregma according to Paxinos and Watson, 2005). The dura was opened and the bundle of the combidrive was lowered onto the exposed cortex, after which it was anchored to the screws with dental cement. To protect the brain from the dental cement, the hole was first filled with a silicone elastomer (Kwik-Sil, World Precision Instruments, Sarasota, Florida). Immediately after surgery all tetrodes were advanced 1.5 mm into the brain, whereas the reference electrodes were lowered 1 mm. The
microdialysis probe was slowly lowered into the brain (5 mm below cortical surface) over a time course of 30 min. In the course of the next three days tetrodes were gradually lowered until they were within range of the dialysis membrane, after which experiments started.

Electrophysiological recordings were performed using a Cheetah recording system (Neuralynx, Tucson, Arizona). Signals from the individual leads of the tetrodes were passed through a low noise unity-gain field-effect transistor preamplifier, insulated multi-wire cables and a fluid-enabled 72 channel commutator (Dragonfly, Inc. Ridgeley, West Virginia) to digitally programmable amplifiers (gain 5000 times; band-pass filtering 0.6-6.0 KHz). Amplifier output was digitized at 32 KHz and stored on a Windows NT station. A 1 ms data sample was taken whenever the signal crossed a preset voltage boundary, so that the width of a recorded spike was captured in 32 data points.

Pharmacological agents and fluid connections

Muscimol hydrobromide and lidocaine hydrochloride were obtained from Sigma-Aldrich (Germany), tetrodotoxin (TTX) was obtained from Tocris (England). For muscimol and TTX, stock solutions of 1.00 mM in milliQ water were made and stored at -80 °C; they were further diluted with phosphate-buffered artificial cerebrospinal fluid (aCSF), containing 143 mM NaCl, 1.2 mM CaCl$_2$, 2.7 mM KCl, 1.0 mM MgCl$_2$, 0.26 mM NaH$_2$PO$_4$ and 1.74 mM Na$_2$HPO$_4$ (pH = 7.4). Lidocaine was dissolved in aCSF before every experiment. All solutions were controlled for osmolality (Micro Osmometer, model 3300, Advanced Instruments, Massachusetts, USA; allowed range 270-290 mOsm) by correcting the amount of NaCl in the aCSF based on the Merck index for sodium equivalents and, if necessary, adjusted for pH (allowed range 7.2-7.6).

A Univentor 801 microinfusion syringe pump (Univentor, Malta) was used to pump the solution through 167 cm long PEEK-tubing (o.d. 0.51 mm; i.d. 0.13 mm; Aurora Borealis Control, Schoonebeek, The Netherlands) that ran via one channel of a quartz-lined dual-channel swivel (Pronexus, Skärholmen, Sweden) and the central channel of the commutator towards the inlet of the dialysis probe (flow rate 2 µl/min). All connections were made of PVC tubing (i.d. 0.38 mm). Switching between the different solutions was done by detaching the PEEK tubing and connecting it to a second, pressurized syringe pump. Between sessions all tubing was rinsed with milliQ water and methanol.

Experiments

Control experiments

At the start of an experimental day the animal was connected to the system and the flow through the dialysis probe was checked. In order to record novel
neurons during each recording session, tetrodes were lowered with increments of 40 µm under continuous perfusion of aCSF. Once the tetrodes were lowered the animal was placed in the recording chamber (40 x 37 x 41.5 cm) and left for at least 1 hour to stabilize unit recordings, after which the actual recording started. The recording chamber was placed in a sound attenuated and electrically shielded box and fitted with a motion detector to monitor locomotor activity of the animal.

Control experiments were performed to determine whether the design of the combidrive actually allowed electrophysiological recordings during reverse microdialysis and whether aCSF perfusion per se would influence firing activity of single units. To this end, recordings of baseline firing activity were made during a 20 min period with no flow, followed by the pump being switched on (referred to as ‘pump switch’) and an aCSF perfusion period of 20 min. Furthermore, to examine the effect of replacement of the probe on the activity of the surrounding neurons, the dialysis probe was changed in a separate session under continuous recording of neuronal activity. Probe replacement was done by hand over a time course of 60 min (for both raising and lowering 30 min).

Pharmacological interventions

After finishing the control experiments, the effects of several drugs on neuronal firing were examined. One experimental session was carried out per day with a single drug, with each drug tested at least two times in at least two animals. The general procedure within these recording sessions was to record baseline neural activity for 20-30 min during aCSF perfusion, followed by perfusion of various drug concentrations for 30 min each. Cumulative concentration effect curves were made for all drugs, consisting of 0.074, 0.74, 7.4 and 74 mM (0.002 - 2%) solutions for lidocaine (2% is the highest concentration of lidocaine within the physiological range for which osmolality could be controlled for), for muscimol of 1.0, 3.0, 10.0 and 30.0 µM solutions and for TTX of 0.01, 0.03, 0.1, 0.3 and 1.0 µM solutions. The session ended with a washout period during which aCSF was perfused. This period was variable in duration for the various drugs since the (qualitative) criterion to end washout was the recovery of neural firing as visible on the oscilloscope (not necessarily back to baseline firing rate). Whenever the dialysis probe was clogged it was replaced, with a maximum of 2 new probe insertions per animal to prevent extensive tissue damage. After the insertion of a new probe recordings continued the next day to allow the tissue to recover.

In addition, since the recording sessions in which drug perfusions occurred were long lasting (up to 6 hours), control sessions of a similar duration were performed, either with or without aCSF perfusion. These control sessions served to exclude the possibility that observed changes in neuronal firing could be ascribed to natural
changes in the firing activity of single units, instead of being the result of the drug perfusions.

**Data analysis**

Single units were isolated by off-line cluster cutting procedures (BBClust/MClust-3.0). Before a cluster of spikes was accepted as belonging to a single unit, several parameters were checked visually, namely the averaged waveforms across the four leads, the cluster plots showing spike parameter distributions such as peak amplitudes across the four dimensions, the autocorrelogram and the spike interval histogram. Since the absence of spike activity during the refractory period (2 ms) is indicative of good isolation, units of which the autocorrelogram and the spike interval histogram revealed any activity during this period were removed from the analysis.

**Control Experiments**

For the statistical analysis, only single units were included that had a baseline firing rate of at least 0.1 Hz and that were active throughout the entire recording session. To determine the effects of aCSF perfusion on neuronal firing activity, the normalized firing rate was calculated in blocks of 5 min. The final 5 min block of the baseline period served as control (100%) value. The final 5 min block of both conditions (i.e. baseline without flow versus aCSF perfusion) was compared using a repeated measures ANOVA (P < 0.05) with time (block) as within-subject variable (SPSS for Windows version 12.0.1). A possible effect of the pump switch was assessed in a similar fashion, although in this case the final 5 min block of the baseline condition was compared with the first 5 min block of the aCSF perfusion period. If indicated by Mauchly’s test of sphericity, a Huynh-Feldt correction was applied to adjust the number of degrees of freedom. To examine whether the replacement of the probe influences recording capacity, the number of single units recorded before and after replacement of the probe was compared.

**Pharmacological interventions**

Single units that were removed from the analysis included neurons with a baseline firing rate less than 0.1 Hz and units that exceeded the 99% confidence interval of the mean firing rate of the population (i.e. 3 standard deviations from mean baseline firing). Furthermore, neurons that did not show any activity in the last 30 min of the washout period were discarded as well; this was done to exclude cells that stopped firing for reasons other than drug perfusion. To examine drug effects, the final 5 min block of each perfusion period was compared with the activity in a 5 min block at the corresponding time point during the aCSF control sessions. A repeated measures ANOVA (P < 0.05) was performed with group (i.e. drug and
aCSF) as between-subjects variable and time as within-subjects variable. Whenever a group or interaction effect was found, additional t-tests were performed to determine which 5 min blocks were different. To examine possible effects of time within groups, an ANOVA with repeated measures was performed over the separate groups as well ($P < 0.05$); whenever a time effect was found, a simple (first) contrast was performed to examine which block differed from baseline firing activity. Based on the number of comparisons the $\alpha$ was adjusted by a Bonferroni correction. Furthermore, the number of degrees of freedom was adjusted by a Huynh-Feldt correction when indicated by Mauchly’s test of sphericity. Based on firing rate and valley width, putative interneurons were initially separated from pyramidal cells. However, due the small amount of interneurons recorded per drug (lidocaine: 12, muscimol: 2, TTX: 3), all data was pooled.

To assess whether the observed drug responses for the different drugs were identical over all sessions, which were spaced across non-consecutive days and were recorded in different animals, a repeated measures ANOVA ($P < 0.05$) was performed over all sessions with sessions as between-subjects variable.

A comparison of the relative reduction of firing activity between the three drugs was made by comparing the cumulative reduction at the highest concentration. Two-sample Kolmogorov–Smirnov tests for equality of distribution were used to assess whether each of the drugs caused different degrees of inhibition.

**Histology**

The final position of the tetrodes was marked by passing a 10 s, 25 µA current through one of the leads of each tetrode in order to induce a lesion and initiate gliosis. After 24 hours the animal was perfused transcardially using a 0.9% saline solution followed by 10% formalin. After removal from the skull the brain was stored in 10% formalin for several days before sectioning. Brain sections (40 µm) were cut using a vibratome and Nissl-stained to identify the location of the probe and to reconstruct the tracks of the tetrodes and their final position.

**Results**

**Histology**

Histological verification of the positions of both tetrodes and microdialysis probe showed that the recording sites and probes in all animals were located in the orbital and lateral areas of the prefrontal cortex. The placement of the probe in each of the four animals is shown in Fig. 2, together with a representative section showing the location of the probe and the end point of three different tetrodes (Paxinos and
Watson, 2005). In two animals, in which the probe was replaced once, the extent of tissue damage was comparable to the animals in which a single probe was inserted.

![Figure 2](image.png)

**Figure 2.** Localization of the microdialysis probe and tetrode recording sites in the four animals (A), and (B) a representative histological slide showing probe location and three marked endpoints of different tetrodes (indicated by the white arrows). The white area in the centre of the section marks the location of the probe tip. Recordings in all rats were localized in the prefrontal cortex, primarily the orbital and lateral areas (Paxinos & Watson, 2005).

**Control experiments**

During all experiments in the recording chamber, animals did not show any sign of distress due to the implanted combidrive or the attachment to the recording equipment, and were able to behave normally. The effect of aCSF perfusion and pump switch on firing activity of neurons was examined in a single session in one rat in which 23 cells were recorded, 17 of which passed the criteria for statistical assessment. Mean firing rates ranged between 0.11 and 3.83 spikes/s. No effects of the aCSF perfusion or pump switch on the baseline firing activity of these cells were found (baseline: firing rate (average ± SEM): 1.00 ± 0.27 Hz; pump switch: 0.99 ± 0.26; F(1,16) = 0.598, P = 0.451; aCSF perfusion: 0.71 ± 0.21; F(1,16) = 0.292, P = 0.597).

The probe was replaced under continuous recording, during which a total number of 59 single units was recorded. Examination of the spike waveforms before and after the probe change suggested that 48 units (81%), recorded before the probe replacement started, were still present when replacement was finished and the new probe had been inserted. Furthermore, 8 units (14%) stopped firing during probe
movement and were lost, whereas 3 units (5%) newly appeared during the probe change. Hence, a total number of 56 units was recorded before, and 51 units recorded after replacement of the probe. Fig. 3 shows an example of neuronal activity before and after the probe was raised.

**Figure 3.** Plots showing peak amplitudes of neuronal activity as recorded on a single tetrode before and after probe movement. A. Before movement of the probe two different clusters (1 and 2) representing two different single units (B) were recorded, which were still present when probe movement was finished as suggested by similarity in waveform and lack of major shifts in the cluster plots. A possible new cluster appeared (cluster 3) which, however, was still close to the noise (N) (not shown in B). B. Average waveforms across the four leads of the tetrode of cluster number 1 and 2.

**Pharmacological interventions**

During the recordings with drug perfusions and the aCSF control sessions (n = 15) a total number of 348 single units was recorded. No effect of drug perfusion on the motor activity of the animals was observed. During the 4 aCSF control sessions 69 single units were recorded, of which 29 passed the criteria for the statistical analysis.

During 4 sessions with muscimol perfusion a total number of 161 neurons was recorded, of which 92 were statistically assessed (Fig. 4). A main effect of time ($F_{(3.586,426.731)} = 9.969$, $P = 0.000$) and group ($F_{(1,119)} = 14.002$, $P = 0.000$) was found, as well as a group/time interaction ($F_{(3.586,119)} = 5.999$, $P = 0.000$). An additional t-test revealed a significant reduction of neuronal firing activity as compared to the aCSF control during perfusion of the 10 and 30 µM muscimol solutions (10 µM: $t = -2.560$, $P = 0.000$; 30 µM: $t = -4.148$, $P = 0.016$), but not
at the end of the washout (Fig. 5A). Furthermore, analysis of the individual groups showed a main effect of time for muscimol \( (F_{(2.420,220.201)} = 37.209, P = 0.000) \) but not for the aCSF control: firing activity during perfusion of 3, 10 and 30 µM muscimol solutions and at the end of the washout (after 3 hours) was significantly reduced as compared to baseline.

![Figure 4](image)

**Figure 4.** Example trace of the activity of a single unit during muscimol perfusion. A. Average waveform across the four leads of the tetrode (width = 1 msec). B. Graph showing the distribution of the spike intervals (ISI). C. Activity of the neuron during the cumulative perfusion of aCSF and 1, 3, 10 and 30 µM muscimol. During perfusion of 30 µM, activity gradually disappeared and recovered during the washout period when aCSF was perfused. On the horizontal axis time (min), on the vertical axis peak amplitude across the four different leads (µV). Each color represents one lead of the tetrode. The arrow indicates entrance in the brain.

A total number of 71 single units was recorded during 5 recording sessions with lidocaine perfusion. Of these neurons, 44 were statistically assessed. A main effect of time \( (F_{(2.914,206.861)} = 10.251, P = 0.000) \) and a group/time interaction was found \( (F_{(2.914,206.861)} = 3.947, P = 0.009) \); the post-hoc t-test revealed a significant reduction in neuronal firing activity as compared to the aCSF control during the perfusion of 74mM lidocaine solution \( (t = -2.457, P = 0.016) \), but not at the end of the washout (Fig. 5B). The analysis of the individual groups showed a main effect of time for lidocaine \( (F_{(1.551,66.681)} = 16.137, P = 0.000) \) but not for the aCSF control:
compared to baseline, neuronal firing activity was significantly reduced during perfusion of 74 mM lidocaine and at the end of the washout (after 30 min). During 2 sessions in which TTX was perfused a total number of 47 single units was recorded, of which 20 were used for the statistical assessment. A main effect of time was found ($F_{(4,058,190.711)} = 8.194$, $P = 0.000$) as well as a group/time interaction ($F_{(4,058,190.711)} = 2.879$, $P = 0.023$). An additional t-test showed a significant reduction in neuronal firing activity as compared to the aCSF control during the perfusion of the 0.1 $\mu$M, 0.3 $\mu$M and 1.0 $\mu$M solutions and at the end of the washout period (0.1 $\mu$M: $t = -2.569$, $P = 0.013$; 0.3 $\mu$M: $t = -2.249$, $P = 0.036$; 1.0 $\mu$M: $t = -2.249$, $P = 0.033$; washout: $t = -3.262$, $P = 0.003$) (Fig. 5C). Analysis of the individual groups showed a main effect of time for TTX ($F_{(1,638,31,121)} = 14.297$, $P = 0.000$), but not for the aCSF control. During TTX perfusion firing activity was significantly reduced during the 0.1 $\mu$M, 0.3 $\mu$M and 1.0 $\mu$M TTX applications and at the end of the washout period (after 2.5 hours) as compared to baseline.

Figure 5. Cumulative concentration-effect curves for muscimol (A; $n = 92$), lidocaine (B; $n = 44$) and TTX (C; $n = 20$). The normalized firing activity during aCSF control (white bars) and drug (grey bars) perfusion is depicted across concentrations (mean ± SEM). The washout periods were 3 h (A, muscimol), 30 min (B, lidocaine) and 2.5 h (C, TTX). The † - symbol indicates a significant difference compared to baseline firing activity. * indicates a significant difference between groups (aCSF control and drug evaluated at a similar time point). Three symbols $P = 0.000$, two symbols $P < 0.01$, one symbol $P < 0.05$. On the horizontal axis the condition of perfusion, on the vertical axis normalized firing activity.
These results indicate that, in addition to the aCSF perfusion period of 20 min that was used in the control experiments, longer aCSF perfusion periods that were performed to serve as control for the drug perfusions did not significantly affect firing activity either since no time effect was found, in contrast to the drug perfusions. It can be concluded that duration of perfusion does not affect stability of neuronal activity, but it should be noted that firing activity of the population tended to decrease during these aCSF control sessions. A possible explanation for this could be the biological state of the animals, for example a diminished arousal due to the fact that they stayed 6 hours in the same recording environment.

The replaceable dialysis probe allowed the use of animals for repeated recording sessions since the probe could be changed when needed: in two animals the probe needed replacement once, for both animals after the fourth recording session. For the other two animals no probe replacement was needed.

![Figure 6](image.png)

**Figure 6.** Relative reduction in firing activity at the highest concentration for each drug. Depicted is the proportion of the population (y-axis) that shows the indicated reduction of firing activity (x-axis) for muscimol (n = 92), lidocaine (n = 44) and TTX (n = 20). *** indicates a significant difference in distribution, (Kolmogorov Smirnov, P = 0.000). The distributions of TTX and muscimol differ significantly from the distribution displayed by lidocaine, but do not differ from each other.

Figure 6 shows the distribution of the reduction in firing rate of neurons within the population at the highest concentration for each drug. For both TTX and muscimol firing of about 40% of the cells was completely abolished, whereas for lidocaine this was about 15% of the cells. A repeated two-sample Kolmogorov–Smirnov test revealed that the distributions of both TTX and muscimol were significantly different from lidocaine, (respectively Z = 2.242, P = 0.000 and Z = 3.380, P = 0.000) but did not differ from each other (Z = 0.846, P = 0.472). This result indicates the relatively low homogeneity in neuronal response of the population to lidocaine as compared to TTX and muscimol, as illustrated in figure 6. However, examination of the activity of individual neurons during drug perfusion...
revealed that for all drugs, including muscimol and TTX, variability in neuronal responses existed among neurons. Differences in the duration after which the firing activity of neurons decreased during perfusion were observed when neurons were recorded on different tetrodes and even when they were recorded on the same tetrode. This is illustrated in figure 7, in which responses of individual neurons during muscimol perfusion are shown, which were recorded on either the same tetrode (Fig. 7A) or on different tetrodes (Fig. 7B).

![Figure 7](image.png)

**Figure 7.** Examples of different neuronal responses during perfusion of several muscimol concentrations within a single session. A. Responses of three individual neurons recorded on the same tetrode. B. Responses of single neurons that were recorded on different tetrodes. On the horizontal axis the condition of perfusion, on the vertical axis normalized firing activity.

Since drug sessions were spaced across non-consecutive days and were recorded in different animals with different dialysis probes, the similarity between the observed neuronal responses for all different drugs over all sessions was examined as well. This revealed no group or interaction effect, meaning that the effect of a particular drug on neuronal firing activity was identical with different dialysis probes across all animals.
Discussion

The present study demonstrated that the combidrive allows ensemble recordings simultaneously with reverse microdialysis in freely moving rats. Perfusion of the microdialysis probe with aCSF for several hours did not significantly affect the basal firing activity of single units. The concentration-dependent reduction in firing activity observed during the local administration of lidocaine, TTX and muscimol showed that tetrodes are within the diffusion range of the probe. In addition, the probe could be used over extended periods of time and could be replaced. Similar pharmacological effects were obtained with multiple probes within a single animal, while the duration of periods during which recordings were made ranged between 7 and 13 days and included up to ten recording sessions per animal. Based on these results it can be concluded that the combidrive is suitable to be applied in behavioral studies, especially during more time-demanding learning tasks.

The drugs tested showed differences in the extent to which they affected activity of the population and in speed of recovery. At the highest concentration used, muscimol and TTX caused a reduction of firing activity of respectively 97.5 and 98%. Lidocaine showed the fastest recovery, but also resulted in a smaller reduction of mean firing activity of the population, namely 80%.

Drug perfusions

The concentration effect curve for lidocaine demonstrated a significant reduction in firing activity during perfusion of a 74 mM (2%) solution. Based on experiments in which a single 2% solution of lidocaine was perfused (to test probe functioning), the onset of the effect was determined at 5-6 min after brain entrance (data not shown), similar to the onset observed during recordings with the concentration effect curve. Neuronal activity (partially) recovered after 30 min, which is in accordance with previous findings after local injections (Albert and Madryga, 1980; Boeijinga et al., 1993; Tehovnik and Sommer, 1997) or reverse dialysis (Boehnke and Rasmusson, 2001).

Compared to lidocaine, the effects of TTX and muscimol perfusion were stronger and more persistent. During perfusion of 0.1 µM TTX, firing activity was significantly reduced, followed by an even larger reduction during perfusion of 0.3 and 1.0 µM TTX. Activity partially recovered 2.5 hours after perfusion was finished, but was still significantly different from the aCSF control. For muscimol, firing activity was significantly reduced during perfusion of 10 µM as compared to aCSF control. However, compared to baseline, perfusion of 3 µM already caused a significant reduction in firing activity. Activity partially recovered 3 hours after end of drug perfusion.
The neuronal population did not respond homogeneously to the drugs. Perfusion of the highest drug concentrations of TTX and muscimol completely abolished firing activity of 40% of the cells, whereas for lidocaine this was 15% (Fig. 6). The fact that not all neurons were fully responsive to the drugs, even when recorded on the same tetrode, might have been caused by the spatial location of recorded neurons with respect to the dialysis probe. Tetrodes record cells located within a region with an estimated radius of ± 65 µm (Gray et al., 1995), meaning that, under the assumption of a straight vertical descent, the recording area ranged between 485 and 615 µm from the dialysis probe. Although the exact location of the neurons cannot be reconstructed, cells located between the dialysis probe and tetrodes would be expected to respond faster and/or to lower drug concentrations than neurons located at the other side of the tetrodes, causing the cells within the population of neurons to respond differentially to the drugs (Fig. 7). However, based on the observed differences in the percentage of cells showing a certain reduction in firing between lidocaine as compared to TTX and muscimol, a more likely explanation is that drug-specific differences, e.g. dissociation constants, determine the neuronal response. This is supported by Boehnke and Rasmusson (2001), who showed that even with a lidocaine concentration of non-physiological osmolality (10%), neuronal activity could not be completely abolished.

The amount of single units showing no recovery at all after drug perfusion was largest for TTX, namely 36% (n = 17), for lidocaine and muscimol respectively 0% and 4% (n = 7). Together with the difference in speed of recovery between TTX and lidocaine, this can be explained by the fact that TTX binding to sodium channels is stronger and longer-lasting than lidocaine binding (Hille, 1992). Therefore, it is expected that if washout was even more prolonged after TTX perfusion these cells ultimately would have recovered their firing activity.

Although data concerning the effects of TTX, lidocaine and muscimol on single units as applied by microdialysis is generally lacking, the results of this study are consistent with the few existing reports using related techniques. For example, Tehovnik and Sommer (1997) observed a lidocaine effect in monkey prefrontal cortex within 5 min after an injection at a distance of 1 mm from the electrodes and a recovery within 30 min. Boehnke and Rasmusson (2001) examined the effect of lidocaine (10%) and TTX (10 µM) on evoked potentials (evoked by stimulation of the forepaw digits) at various distances from the microdialysis probe in raccoon somatosensory cortex. Recovery of activity at a distance of 0.5 mm from the dialysis probe required approximately 40 min after lidocaine application, but was not observed within two hours after TTX. The apparent discrepancy with the current study (30 out of 47 (64%) neurons partially recovered from TTX when recorded with tetrodes positioned at 0.55 mm from the dialysis probe) can be explained by the fact that evoked potentials require higher levels of activity to be detected against
a noisy background. Edeline et al. (2002) reported a blocking effect of muscimol injections on neuronal firing activity in an area of drug diffusion of 2-3 mm which was confirmed with autoradiography. This is consistent with the effect after 10 µM muscimol perfusion as observed in the present study. The results are furthermore in agreement with Sakai and Crochet (2001), who observed a weak effect of 50 µM muscimol on single units in cat brain stem (distance between dialysis probe and electrodes 1 mm), whereas concentrations of 100 or 500 µM caused respectively an almost complete and complete suppression of firing activity.

Diffusion of drugs

One uncertain factor in the method concerns the spatio-temporal dynamics of drug diffusion. Diffusion can be calculated (Fick’s law), but when the diffusion coefficient in brain tissue (D_BR) of a drug is unknown, only an estimate of the concentration delivered at the recording site can be made. Diffusion constants of most drugs have not been determined, although data for TTX (Zhuravin and Bures, 1991) and some neurotransmitters are available (Rice et al., 1985). Estimates can be calculated on the basis of diffusion constants in liquids and recovery data in vivo (Lindfors et al., 1989), which are relatively simple to measure, but the penetration into the brain tissue has to be determined experimentally for each individual drug as factors such as binding to macromolecules and receptors and uptake into cells cannot be estimated from diffusion constants in liquids or recovery data (Benveniste, 1989). This was clearly demonstrated by the study of Boehnke and Rasmusson (2001), where TTX and lidocaine, despite comparable molecular weights, differed in spreading and time of duration of the effect. Alternative strategies for measuring effective spread are autoradiography (Edeline et al., 2002) or dual probes (Höistad et al., 2000). However, for the drugs tested in the current study, all available data indicated that they can bridge the 0.55 mm distance between microdialysis probe and recording tetrodes.

Overall, findings in this study demonstrated that both muscimol and TTX caused a comparable reduction in neuronal firing, whereas the effect of lidocaine was less strong. Hence, when during behavioral experiments the contribution of a brain area to learning is examined by inactivation of that particular area, and a longer lasting, but reversible inactivation is required, muscimol should be used. Firing activity was almost completely abolished for a longer period of time, and, although reduction of neuronal activity was comparable with TTX, muscimol has the advantage that it inactivates neurons locally, i.e. at the soma and the dendrites, whereas TTX also prevents the occurrence of action potentials in fibers of passage. Whenever an inactivation is required for a shorter period of time, lidocaine would be most suitable, although it should be taken into account that the population response was highly variable and not fully blocked. Regardless of the precise drugs used, the
results also imply that behavioral studies relying on local injections should take into account the variability in the neuronal response to the drug, even in a confined region around the injection site.

The results furthermore demonstrated that the combidrive can be used to combine reverse microdialysis with ensemble recordings in freely moving animals, and can be applied in future studies to examine how neurotransmitters exert their effect on the activity of neuronal populations during behavior. This provides information regarding the interplay between neurotransmitters and the activity of neuronal populations during for example processing of reward-related information in behavioral learning tasks, and would clarify which neurotransmitters are actually involved in these kinds of processes.

**Acknowledgements**

This work was supported by NWO Grant 903-47-084 and ZonMW-TOP 912-02-050. We would like to thank David Redish and Peter Lipa for providing the cluster cutting software, Paul Evers for histological processing of the brains, Ton Put for help with illustrations and our colleagues at the mechanical workshop for their excellent technical assistance.
Chapter 6

General Discussion
The main aim of the research described in this thesis was to examine whether and how reward-related parameters such as the magnitude and probability of reward are coded within the orbitofrontal cortex. To this end, ensemble recordings were performed while animals performed an olfactory discrimination learning task. Neural activity was examined to determine whether predictive information regarding the magnitude and probability of reward were coded by individual cells and populations of neurons, and to gain knowledge about how these parameters are actually represented within the orbitofrontal cortex in terms of specificity for trial phases, evolvement over time within these specific trial phases, and to get insight in how these representations are built up as learning progresses. In addition, to be able to manipulate the activity of these orbitofrontal ensembles during behavior, the combidrive was developed.

This discussion will first provide an overview of the findings by chapter, followed by a discussion focusing on the contribution of the orbitofrontal cortex to reversal learning, since deficits in this type of learning are the most prominent feature after orbitofrontal damage. Finally, the results of the current thesis are integrated in a general view of orbitofrontal functioning during adaptive goal-directed behavior.

Summary of results

In chapter 2, the coding of reward magnitude by single units in the orbitofrontal cortex was examined. The orbitofrontal cortex is known to be involved in the representation of the motivational significance of stimuli and in applying this information to the guidance of goal-directed behavior (Gallagher et al., 1999; Lipton et al., 1999; O'Doherty et al., 2003; Schoenbaum et al., 1999, 2003; Thorpe et al., 1983; Yonemori et al., 2000). Orbitofrontal neurons in primates were found to encode predictive information regarding upcoming reinforcers, demonstrating differential firing activity to various types or amounts of reinforcers (Hikosaka and Watanabe, 2000; Ichihara-Takeda and Funahashi, 2006; Padoa-Schioppa and Assad, 2006; Roesch and Olson, 2004; Roesch et al., 2006; Simmons and Richmond, 2008; Tremblay and Schultz, 1999; Wallis and Miller, 2003). In addition, findings in rat orbitofrontal cortex suggested predictive neural coding of both appetitive and aversive outcomes (Schoenbaum et al., 1998). However, it was not known whether reward-predicting information was quantitatively represented in the rat orbitofrontal cortex. This would be expected, since it is known that the magnitude of a primary reinforcer exerts a profound effect on the selection and speed of behavioral responses (Black, 1968; Bohn et al., 2003; Boysen et al., 2001; Brown and Bowman, 1995; Campbell and Seiden, 1974) and hence represents great motivational value. To examine whether information regarding reward prediction is quantitatively represented in the rat orbitofrontal cortex, ensemble activity was
olfactory discrimination ‘go/no-go’ task in which five different odor stimuli were predictive for various amounts of a rewarding sucrose solution or an aversive reinforcer, i.e. quinine. Ensemble recordings were made using an array of tetrodes (Gothard et al., 1996; Gray et al., 1995), a technique which provides the advantages of yielding high numbers of well-isolated cells. During the task, animals could obtain reinforcement after odor sampling by making a nose poke in the fluid delivery well. Upon the nose poke in the fluid well, there was a waiting period of 1.5 s, after which the reinforcer was delivered. During this waiting period, neural activity reflecting the expectancy for the upcoming reinforcement was predicted to occur.

The results obtained in this study showed that predictive information regarding reward magnitude is represented by single units in rat orbitofrontal cortex in multiple trial phases. During the task, animals were able to discriminate between the various amounts of expected reward (as visible by shorter response latencies for the larger amounts of reward), and neural correlates related to both actual and expected reward magnitude were observed. Responses related to the expectation of the upcoming amount of reward were found to occur within the waiting period prior to reinforcement delivery and during the execution of the behavioral response towards the fluid well after odor sampling (the ‘movement period’). About one-half of the neurons showing a behavioral correlate in these periods demonstrated differential firing towards the different reward sizes. They showed a variety of different tuning curves to reward magnitude, which is as compatible with coding in a parallel-distributed network as is a monotonic relationship between firing activity and reward magnitude.

To elaborate further on the neural coding of reward magnitude in the orbitofrontal cortex, Chapter 3 addressed the question whether this reward parameter could also found to be represented by the activity of orbitofrontal ensembles. Single cell studies can not show how predicted or actual rewards are dynamically represented at the population level in the orbitofrontal cortex, which is especially relevant for understanding how other, connected brain areas may read out population activity from this structure. Although two earlier studies described ensemble activity in the orbitofrontal cortex (Gutierrez et al., 2005; Schoenbaum and Eichenbaum, 1995b), it remained unknown how actual and predicted rewards are represented by ensemble activity within a specific trial phase during an operant task, and whether such a population code would be specific for different trial phases. Using two different reconstruction algorithms, the Bayesian method and template matching, the magnitude of reward could be decoded from population activity during the movement and waiting period and after reward delivery. The decoding score was only weakly dependent on the size of the neuronal group participating in the reconstruction, consistent with a redundant, distributed representation of reward information. Furthermore, decoding was found to be largely specific for trial periods,
meaning that to a fair extent the ensemble activity is specific for these particular trial phases, and carry-over of reward information to the next trial phase occurs only to a limited extent.

Since animals learned to discriminate between at least some of the various amounts of expected reward during the task, the dynamics by which consistent coding develops as the learning task progressed was also assessed. The decoding performance was found to increase steeply across the first few trials of the behavioral session, an effect that could not be explained by a nonspecific drift in response strength across trials. Finally, when the population responses to quinine as negative reinforcement were compared to the appetitive sucrose reinforcement, ensembles were demonstrated to discriminate between these two different outcomes, meaning that coding in the delivery phase is related to the quality of reward.

Altogether, these findings show that representations of reinforcer quality and magnitude are broadly distributed across ensembles with a high, sub-second time resolution.

In Chapter 4 we examined whether a different reward parameter, i.e. probability of reward, was represented in rat orbitofrontal cortex as well. As is the case for reward magnitude, the probability of reward is a key factor in decision-making as well. For example, in a process called probability discounting, the value of probabilistic rewards is downgraded as the reinforcer becomes more uncertain (Rachlin et al., 1991). Human brain imaging studies indicated a role for the orbitofrontal cortex in decision making under uncertainty. For example, Hsu et al. (2005) demonstrated a positive correlation between orbitofrontal activity and the level of unpredictability of reward. In addition, animals with orbitofrontal lesions preferred the larger, but uncertain reward, which is in accordance with the stronger risk-taking behavior as demonstrated by humans with orbitofrontal damage (Pais-Vieira et al., 2007). However, no neurophysiological evidence was available to demonstrate that single unit activity in the orbitofrontal cortex can actually code reward probability. To this end, ensemble recordings were performed during a probabilistic olfactory discrimination ‘go-no/go’ task, in which odors were now predictive of the probability of a pellet reward. The task design was similar to the task used to assess the coding of reward magnitude: after odor sampling, there was a waiting period of 1.5 s in the food trough, after which the pellet was delivered or not. The probability of expected reward was found to be coded in the orbitofrontal cortex in a fashion comparable to the magnitude of reward: during multiple trial phases, predictive information regarding the probability of reward was found to be represented by both single units and populations of neurons. During both the movement and waiting period, single unit activity was found to be modulated by reward probability, and this parameter could also be reconstructed from population activity significantly above chance level for both trial phases. Neurons that
discriminated between different expected probabilities showed a variety of different tuning curves. Of particular interest was the observation that a subset of neurons active in the reward delivery phase was found to respond specifically during unrewarded trials, and not during rewarded trials. These cells might signal the absence of reward when it is expected and hence may provide an error signal representing the violation of a positive reward prediction. Furthermore, we were able to demonstrate in this study that it was indeed probability of reward and not reward uncertainty coded by population activity - two distinct, but closely related reward parameters.

Finally, chapter 5 described a new technological development in the field of ensemble recordings. With the ability to record ensemble activity, additional questions arise, for example how neurotransmitters influence activity of cell populations. To gain more insight in the interaction between neurotransmitters or pharmacological agents and neural firing activity, the combidrive was developed, in which the technique of (reverse) microdialysis was implemented in a multi-tetrode array similar to those in use to perform ensemble recordings (Gothard et al., 1996; Gray et al., 1995). Hence, the combidrive combines a multi-tetrode array consisting of a circular row of 12 individually movable tetrodes and 2 reference electrodes to perform ensemble recordings with a movable and replaceable microdialysis probe to locally administer pharmacological agents.

Assessment of the combidrive showed that this device indeed allowed ensemble recordings simultaneously with reverse microdialysis, and that the firing activity of neurons in the prefrontal cortex was not affected by perfusion through the microdialysis probe per se. In addition, the combidrive was applied in a comparative study to examine the effects of cumulative concentrations of tetrodotoxin (TTX), lidocaine and muscimol on neural firing activity in the prefrontal cortex. Although these drugs are widely applied in behavioral studies to transiently inactivate selective brain areas (Albert & Mah, 1973; Brioni et al., 1989; Ivanova & Bures, 1990), until now little was known about the dynamics and reversibility of their inhibitory effect in relation to population activity in awake animals and the possible differences between them. The results of this comparative study showed that all drugs reduced neural firing in a concentration-dependent manner, but differences were observed in the extent to which firing activity of the population was diminished and in the speed and extent of recovery. Both muscimol and TTX caused an almost complete cessation of firing activity for longer periods of time at the highest concentration used, whereas the perfusion of lidocaine resulted in a smaller reduction of firing activity of the population and in the fastest recovery of firing activity after washout. The overall results from this study indicated that when during a behavioral experiment a longer lasting, but reversible inactivation is required, muscimol is the drug of choice. Although the effects of TTX were comparable to muscimol, the
latter has the additional advantage that it inactivates neurons locally, i.e. at the soma and the dendrites, whereas TTX also prevents the occurrence of action potentials in fibers of passage. Whenever an inactivation is required for a shorter period of time, lidocaine would be most suitable, although this drug was not able to fully block the population response. The results also implied that during behavioral studies it is important to take into account the variability in the neural response to the drug that is used, even in a confined region around the injection site. Furthermore, the combidrive turned out to be a useful tool to be applied in future studies to examine how neurotransmitters exert their effect on the activity of neuronal populations, as well as to clarify which neurotransmitters are actually involved in local cognitive or computational operations that take place in the brain during behavior, for example the contribution of serotonin or dopamine to processes of learning and attention.

The next question is whether the knowledge obtained as described in this thesis about the processing of reward-related information by orbitofrontal ensembles might provide a closer understanding of the involvement of the orbitofrontal cortex in reversal learning, in which subjects have to learn that previously rewarded stimuli are no longer rewarded whereas previously unreinforced events are now paired with reward. Impaired reversal learning is the most prominent, and probably most frequently reproduced feature of orbitofrontal damage (Bohn et al., 2003; Chudasama and Robbins, 2003a; Dias et al., 1996; Fellows and Farah, 2003; Ferry et al., 2000; Hornak et al., 2004; Izquierdo et al., 2004; Kim and Ragozinno, 2005; McAlonan and Brown, 2003; Meunier et al., 1997; Rolls et al., 1994; Schoenbaum et al., 2002, 2003b). Reversal learning itself was not experimentally examined in this thesis, since the first step in understanding the functionality of the orbitofrontal cortex should be to explore what kind of information is processed in this area and in what manner, which is most straightforward during task acquisition. The question that is central in the part below is how reversal learning depends on the orbitofrontal cortex, and we will explore how the results obtained in this thesis during task acquisition can add to the understanding of the functional role of this area in reversal learning.

The orbitofrontal cortex and reversal learning

Response inhibition

As already mentioned, damage to the orbitofrontal cortex results in the inability to rapidly learn reversals of previously acquired stimulus-reinforcer associations in reversal tasks. Initially, this impairment was considered to be caused by a failure in response-inhibition (Jones and Mishkin, 1972; McEnaney and Butter, 1969). In these studies, monkeys with orbitofrontal damage who performed an
object or place discrimination learning task demonstrated difficulties in reversing the previously acquired stimulus-reinforcer association. According to Jones and Mishkin (1972), the most plausible explanation for this deficit was a perseveration of the already established associations, although they did not exclude the possibility that this impairment was caused by difficulties in the formation of new stimulus-reinforcer associations. Similar observations came from more recent experiments that involved a variety of reversal tasks in multiple species (Bohn et al., 2003; Chudasama et al., 2003a; Dias et al., 1996; Fellows and Farah, 2003; Ferry et al., 2000; Izquierdo et al., 2004a; Kim and Ragozzino, 2005; McAlonan and Brown, 2003; Meunier et al., 1997; Rolls et al., 1994; Schoenbaum et al., 2002, 2003b). In most of these studies, a deficit in withholding responses became apparent after the reversal, but no difference between normal and lesioned animals was found in the acquisition of the initial discrimination, meaning that during this phase of the task animals learned to inhibit their responses at a similar rate as control animals. However, this observation is not accounted for in the explanation that the orbitofrontal cortex is involved in the inhibition of responding.

The idea that inhibition of prepotent responding is not a hallmark of orbitofrontal functioning is further supported by a number of more recent experiments. In a reversed reward contingency task, monkeys with orbitofrontal lesions were able to inhibit responses normally (Chudasama et al., 2007). In this task, animals selected a reward quantity by touching or reaching towards the rewards, without being able to get them. The selection of the smaller of two food quantities resulted in the receipt of the larger quantity, and vice versa. Reward magnitude represents strong motivational value, and the choice for the largest reward can be considered a strong prepotent response. However, orbitofrontal lesioned monkeys learned to suppress this natural response tendency in order to obtain the largest reward at the same rate as the control monkeys. Furthermore, Ostlund and Balleine (2007) showed that rats with orbitofrontal lesions were able to stop responding normally (i.e. stop pressing a lever) after reward devaluation during instrumental action-outcome learning in a lever-press task, indicating that the OFC is not critically involved in encoding the outcome of instrumental conditioning. Interestingly, testing of outcome-specific pavlovian-instrumental transfer (i.e. the facilitatory influence of pavlovian learning over instrumental performance) in the same task showed that orbitofrontal lesions made after, but not before training, abolished pavlovian-instrumental transfer (PIT). This finding provides additional evidence against the explanation for the deficits observed during reversal learning and reward devaluation paradigms (all tasks in which the target response depends at least partially on pavlovian conditioning; Roberts, 2006), that the orbitofrontal cortex has a general function in response inhibition: the effect on PIT was attributable to a failure to increase performance (i.e. increasing the number of lever presses in the
presence of a pavlovian stimulus), which is inconsistent with the idea of a deficit in response inhibition.

Altogether, these studies demonstrate that the orbitofrontal cortex is not required for the inhibition of responses, at least not in all situations. But when a deficit in response inhibition has to be ruled out as the cause of the observed impairments in reversal learning after orbitofrontal damage because this area is not critical for response inhibition, how does reversal learning then depend on the orbitofrontal cortex?

**Electrophysiology**

Schoenbaum and colleagues have provided electrophysiological data regarding orbitofrontal functioning during reversal learning in rats (see Schoenbaum et al., 2007 for review). The basic design of the dual-odor discrimination ‘go/no-go’ task used in these studies is that animals learn to make a ‘go’ response to the reward site to obtain reward after sampling an odor predictive of a reinforcing outcome (sucrose), but that this response will result in a punishment when made after sampling a second odor predicting an aversive outcome (quinine). Both control animals and animals with orbitofrontal lesions learn to respond to the positive odor, and not to respond after the negative one. During the reversal, in which the contingencies are switched such that the positive odor now predicts the negative outcome and the negative odor the positive outcome, lesioned animals are impaired in acquiring the new associations (Schoenbaum et al., 1999; 2003b). Performing electrophysiological recordings in both basolateral amygdala and orbitofrontal cortex showed that during this task, neurons within these two areas develop cue selective-activity (Schoenbaum et al., 1999; 2003b; see also Rolls et al., 1996; Thorpe et al., 1983), either in combination with a response towards the response outcome (referred to as cue-selective outcome-expectant neurons), or solely to the cue (referred to as cue-selective neurons). These neurons were furthermore found to either reverse their cue-selectivity during reversal learning, or to become non-selective. For example, during the initial learning phase, a particular neuron becomes selective to the positive odor, but after reversal the neuron switches its response to the previously negative odor, now that this odor predicts the positive outcome. Hence, these cells fire more in response to a stimulus predicting the positive response outcome. Neurons displaying the opposite pattern, which is responding more to the negative odor, are also found. Based on this finding, one could propose that the orbitofrontal cortex rapidly encodes reversals of cue-outcome associations. However, examination of the population response of the neurons displaying cue-selectivity in orbitofrontal cortex showed that the group of neurons referred to as cue-selective neurons actually loose their selectivity after reversal, and that after reversal a new cue-selective population develops (Stalnaker et al., 2006). Furthermore, it was
demonstrated in the same study that the probability of observing the reversal of cue-selectivity in the orbitofrontal cortex is actually inversely related to the rate of reversal learning, meaning that when more neurons are reversing their cue-selectivity, performance during reversal learning worsens. Hence, the flexibility of associative encoding in orbitofrontal cortex is inversely related to the speed of reversal learning. These findings indicate that flexible encoding in the orbitofrontal cortex is not an exclusive mechanism for reversal learning, since in that case one should expect that the population of neurons reverses cue-selectivity, together with an opposite relationship between performance during reversal learning and the probability of observing reversal of cue-selectivity, meaning a better behavioral performance with an increasing amount of reversing neurons.

An alternative hypothesis for the contribution of the orbitofrontal cortex during reversal learning was proposed by Schoenbaum et al. (2006). According to this hypothesis, the coding of expected outcomes within the orbitofrontal cortex may provide a signal that can be used to compare with the actual outcome and hence drive new learning, or adaptations of existing associations. When an actual outcome does not correspond with the expected outcome (which is obviously the case during reversal learning), this comparison can be used to alter, or update, the acquired representations. As a consequence, the orbitofrontal cortex would support reversal learning indirectly by facilitating changes in associative encoding in other brain areas, such as the basolateral amygdala, instead of directly, by rapidly encoding associations. However, one can object that during the acquisition phase prior to the reversal, in which stimulus-reward associations are presented that are unfamiliar to the animal, new learning occurs as well. In this respect, the proposal of Schoenbaum et al. (2006) still does not explain the ability of lesioned animals to normally inhibit responses during this phase.

Furthermore, it should be mentioned that the idea of the orbitofrontal cortex being responsible for reward driven learning was originally implemented in a computational model of reinforcement learning as proposed by Pennartz (1997). In this model, glutamatergic projection neurons of orbitofrontal cortex and basolateral amygdala are responsible for the processing of reward related information during reward-driven sensorimotor learning. Synaptic weights of these neurons come to reflect a stimulus-specific value of mean previous reward, and this reward value then acts as a predictor of future reward: when there is a constant performance upon the presentation of this specific stimulus, the mean reward value associated with this stimulus is also expected in future trials. The difference between the actual and the mean previous reward equals the error in the reward prediction, and these errors give rise to synaptic modifications in the sensorimotor network (i.e. the neocortex and striatum) that can drive new learning (Pennartz et al., 2000).
Relationship with the basolateral amygdala

As was demonstrated by Schoenbaum et al. (1999), neurons in the basolateral amygdala develop cue-selectivity as well, and, in contrast to the orbitofrontal cortex, this population does reverse its response after reversal learning (Stalnaker et al., 2007b). Hence, this population might be suggested to code the reversed cue-outcome associations. Furthermore, this associative coding was shown to depend on the orbitofrontal cortex, which is in line with the proposal that the orbitofrontal cortex would support reversal learning by facilitating associative coding in the basolateral amygdala. Cue-selective neurons in basolateral amygdala were unable to reverse selectivity in rats with orbitofrontal lesions (Saddoris et al., 2005) during reversal learning, which is probably caused by the absence of signaling of expected outcomes. Hence, the impairment observed during reversal learning seems to be mediated by the inflexibility of associative encoding in the basolateral amygdala, which is in turn caused by damage to the orbitofrontal cortex. Furthermore, when the orbitofrontal cortex supports reversal by detecting errors so that downstream areas can modify their representations, one should predict that lesions of the basolateral amygdala diminish the effects during reversal lesions after orbitofrontal lesions. In a recent study by Stalnaker et al. (2007a), it was demonstrated that lesions of the basolateral amygdala, which apparently do not have an effect on acquisition or reversal by themselves (Izquierdo and Murray, 2007, but see Schoenbaum et al., 2003a), correct the reversal impairment caused by bilateral orbitofrontal lesions. This suggests that the persistent coding of associations within basolateral amygdala, due to the absence of a correcting signal from the orbitofrontal cortex, impairs reversal learning. But still the issue remains how lesioned animals are able to withhold responses during the acquisition phase prior to the reversal.

Alternatively, we propose a different model for the neural mechanism underlying reversal learning, which is slightly analogous to the model for extinction learning as proposed by Quirk and colleagues (see for review Quirk and Meuller, 2008). During extinction learning, conditioned responding to a conditioned stimulus decreases when a reward is omitted. Extinction, like all other forms of learning, is thought to occur in three phases: acquisition, consolidation and retrieval, which are processes that each depend on a specific structure, namely the amygdala, the prefrontal cortex and the hippocampus. Based on experimental findings, it is hypothesized that the amygdala stores both memories for conditioning and extinction. Information about the conditioned stimulus enters the amygdala, hippocampus, and the infralimbic part of the medial prefrontal cortex (IL). The IL integrates the information about the conditioned stimulus with contextual information from the hippocampus in order to determine extinction retrieval. Furthermore, during extinction, the IL inhibits output from the amygdala to reduce fear. However, outside the extinction context, the output from the amygdala remains
uninhibited. The IL may therefore emit a ‘safety signal’ that can overrule a primary, prepotent response tendency elicited by a (formerly) fear-inducing stimulus (Milad and Quirk, 2002).

For the neural mechanism underlying reversal learning, we propose a model in which the primary association between stimulus and reward during reversal learning is formed within the amygdala, whereas after the reversal the orbitofrontal cortex is able to overrule this primary association. For example, in the acquisition phase of a conditioning task, primary associations between stimulus A predicting reward, and stimulus B predicting punishment, are formed in the amygdala. This associative information is projected to the orbitofrontal cortex, which also has the capability by itself to form stimulus-reward associations. After reversal, when the initial contingencies are altered, the primary association is overruled by the orbitofrontal cortex, since the alterations in the known stimulus-reward associations require an adaptation in behavioral responding. In this process, the primary association in the amygdala is not erased, but is still retained in memory (comparable to what happens during extinction learning; see Quirk and Meuller, 2008), and the orbitofrontal cortex guides the required alterations in behavior based on the secondary, higher-order association, a result of ‘meta-rule’ learning. This model also explains why animals with orbitofrontal lesions are still able to learn the initial associations during reversal learning, since the primary association is formed in the amygdala. Hence, the orbitofrontal cortex subserves a dual function in reversal learning: together with the amygdala it is involved in the formation of the stimulus-reward associations, but simultaneously it also controls activity within the amygdala. It is important to stress here that the primary associations can also be formed elsewhere in the brain, i.e. outside the orbitofrontal-amygdalar circuitry (e.g. in the medial PFC and ventral and dorsal striatum). Hence, orbitofrontal cortex and amygdala would not be the only brain regions involved in reversal learning, which does not contradict the hypothesis of the orbitofrontal cortex mediating higher-order learning per se. Confirming the non-exclusive involvement of orbitofrontal cortex and amygdala in reversal learning, it was previously demonstrated that monkeys with lesions of the rhinal cortex are impaired in object reversals, which is a type of reversal learning in which the identity of an object is predictive for future reward (Murray et al., 1998). Considering the fact that this is a different type of reversal learning that makes use of a visual discrimination task, whereas Schoenbaum and colleagues use an olfactory discrimination task, one might expect that this particular task may rely on different brain systems. Nevertheless, the rhinal cortex is known to have widespread connections with the neocortex and the amygdala (Insausti et al., 1997; Shi and Cassell, 1999), so there may well be a general role for this brain area in reversal learning as well.
Interesting in this respect is the recent finding that complete, bilateral lesions of the amygdala in the monkey did not impair object reversal learning in monkeys, but did impair performance in a reward devaluation task, which means that monkeys were unable to shift their choices for objects on the basis of changes in the associated reward value (Izquierdo and Murray, 2007). As suggested by the authors, these results indicate that the amygdala makes a specific contribution to the process of associating stimulus and reward: the amygdala would not be critical for guiding choices based on the representation of reward contingency (at least not in object reversal tasks), but for guiding choices after changes in reward value, as evident from the observed impairments in the reinforcer devaluation task. However, one can object that this reversal task, since it is a visual discrimination task, can also be learned by visually-based performance rules in which the reward provides information independent of its reinforcing value (Gaffan et al., 1985). Hence, learning in this task would then be independent of the formation of associations between stimulus and reward.

**Single unit and ensemble activity: towards a unified electrophysiological view of orbitofrontal functioning**

Results obtained with the electrophysiological single unit studies within orbitofrontal cortex and basolateral amygdala as described above did not involve population analysis of neural activity as described in the present thesis, but were based on the firing activity of individual cells. Although ensemble analyses can not clarify the role of the orbitofrontal cortex in reversal learning, they do provide additional knowledge about orbitofrontal functioning during olfactory discrimination ‘go/no-go’ tasks, in particular how dynamic the coding of reward-predictive information actually is. According to the proposal by Schoenbaum et al. (2006), and in agreement with the neural-network model by Pennartz (1997), the orbitofrontal cortex supports reversal learning by providing a signal that represents the discrepancy between the actual and expected outcome, which is subsequently used to alter associative coding in downstream areas. If this is indeed the case, rapid and flexible encoding of this information is required, since the ability to rapidly make decisions when environmental circumstances are changing can be vital. We demonstrated that the representation of reward predictive information by orbitofrontal ensembles is indeed flexible by nature, at least during initial task acquisition. Small subsets of orbitofrontal neurons coding reward predictive information are already sufficient for providing (parallel-distributed) signals that allow significant read-out in target structures, such as the basolateral amygdala. In addition, this information is available on short time scales (Chapter 4), which is another requirement to be able to rapidly make decisions. In the context of our alternative hypothesis for orbitofrontal functioning in reversal learning, i.e. exerting
higher-order control over the amygdala, striatum and other structures, a similar flexibility in forming secondary associations and in representing reward predictive information is required.

Of special interest is that reward predictive information was found to be coded in various stages in the process of decision making, including an early stage during which the ‘go’ response is executed (Chapter 2). The presence of predictive information in this phase supports a role for the orbitofrontal cortex in the control of conditioned responses and in encoding action-outcome associations. This is evident in reward devaluation tasks, in which animals with orbitofrontal lesions are unable to decrease conditioned responding to a conditioned stimulus according to an updated representation of the devaluated response outcome. This ability requires that predictive information regarding response outcomes becomes integrated with the decision towards and execution of the response, a process that may also require the orbitofrontal cortex (Baxter et al., 2000).

**Future perspectives and concluding remarks**

One aspect not included in this thesis concerns the activity of orbitofrontal ensembles during reversal learning. Since the orbitofrontal cortex and basolateral amygdala seem to form a complementary system involved in this type of learning, it would be of great interest to perform ensemble recordings simultaneously in orbitofrontal cortex and basolateral amygdala during reversal learning. Whenever an output signal from the orbitofrontal cortex is indeed necessary to update established associations in the basolateral amygdala (Schoenbaum et al., 2006), one would predict that the formation and alteration of these associations will occur with different time courses within these two structures. Furthermore, since several studies imply that more brain areas are involved in reversal learning, the next step in revealing the brain circuitry mediating reversal learning in its entirety could be recording neural activity in the rhinal cortex during this task, or inducing lesions in this area in combination with recordings in output areas, such as the amygdaloid complex. To test how generally our hypothesis of the orbitofrontal cortex mediating higher-order learning can be applied one can examine, when an animal has learned an association between a stimulus and a reward in a specific context, whether the orbitofrontal cortex is needed to learn that the same association is not valid in a different context.

In conclusion, the overall concept of population coding in the orbitofrontal cortex emerging from this thesis is that representations of various reward parameters, such as reinforcer quality, magnitude and probability are broadly distributed across ensembles and are characterized by a high, sub-second time resolution. The finding
that signals related to reward expectation are coded across different temporal phases along the process of decision making furthermore supports the hypothesis that orbitofrontal neurons collectively code a matrix of reward parameters as a function of the delay towards the moment of response outcome. Reward parameters are expressed more as modulatory signals rather than being main determinants of firing rate, and such a matrix provides the framework through which behavior can be altered upon environmental changes, regardless of whether the behavioral alterations are made voluntarily or not.

Although Schopenhauer concluded already in 1839 that man does not possess ‘free will’, he reached this conclusion solely by means of his own thinking, and without feeling the need of acquiring scientific evidence to support his thoughts and ideas. Nowadays, however, besides philosophers, neuroscientists are trying to solve this issue as well, seeking empirical evidence for the existence of ‘free will’. But the elucidation of the neural mechanisms of decision-making requires enormous effort, as hopefully became apparent from this thesis, and is only a small step towards solving the issue of how ‘free will’ is possibly exercised, especially when one considers that decision-making might be influenced by processes that do not even reach consciousness.

However, I do think that the work presented in this thesis will help in revealing the process of decision-making, and ultimately to solving the question of whether ‘free will’ can exist.
References

Bechara, A., Damasio, A.R., Damasio, H. and Anderson, S.W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 50: 7-15
Bechara, A., Tranel, D., Damasio, H. and Damasio, A.R. (1996) Failure to respond autonomically to anticipated future outcomes following damage to the prefrontal cortex. Cerebral Cortex 6: 215-225


Chudasama, Y., Kralik, J.D. and Murray, E.A. (2007) Rhesus monkeys with orbitofrontal cortex lesions can learn to inhibit prepotent responses in the reversed reward contingency task. *Cerebral Cortex* 17: 1154-1159


cortex and limbic structures (eds. HBM Uylings, CG van Eden, JPC de Bruin, MGP Feenstra and CMA Pennartz), Progress in Brain Research Vol. 126, 3-28


Ivanova, S.F. and Bures, J. (1990) Acquisition of conditioned taste aversion in rats is prevented by tetrodotoxin blockade of a small midbrain region centered around the parabrachial nuclei. *Physiol Behav* 48: 543-549


sampling depends upon input from basolateral amygdala. *Neuron* 39:855-867


generated patterns of behaviour with poor environmental modulation. Neuropsychologia 31: 1379-1396


Dutch Summary

Een factor die in sterke mate bepalend is voor gedrag van mens en dier is het kunnen verkrijgen van een beloning als gevolg van een handeling. Bepaalde hersengebieden, waaronder de prefrontale cortex, het striatum en het ventrale mesencephalon, zijn betrokken bij het voorspellen van toekomstige beloningen op basis waarvan gedragshandelingen gestuurd kunnen worden. Hierbij spelen factoren als de beloningsgrootte, de kans op de beloning en de emotionele waarde ervan een belangrijke rol. Verstoringen in dit systeem van hersengebieden leiden onder andere tot ziektebeelden als schizofrenie, depressie en drugs- en gokverslaving. Deze stoornissen hebben gemeenschappelijk dat een aantal aspecten van het verwerken van beloningsinformatie is verstoord, zoals de perceptie en het voorspellen van een beloning, maar ook het plannen tot het verkrijgen van de beloning.

In de experimenten beschreven in dit proefschrift is onderzocht hoe in de orbitofrontale cortex, een subgebied van de prefrontale cortex betrokken bij het doelmatig uitvoeren van gedragshandelingen, verschillende aspecten van beloningen worden verwerkt. De orbitofrontale cortex is onder meer nodig om snel en adequaat te reageren wanneer een bepaalde handeling niet meer door een beloning wordt gevolgd, maar plotseling wordt bestraft. De verandering in het resultaat van de handeling vergt een aanpassing in het gedrag, en om deze aanpassing te kunnen maken is de aanwezigheid van (voorspellende) informatie omtrent de uitkomst van de handeling een vereiste. Om te bestuderen hoe voorspellende informatie over beloningsparameters als grootte en kans worden verwerkt in de orbitofrontale cortex werden ratten in een gedragstaak getraind om een geur te ruiken en vervolgens naar een naastgelegen beloningsbakje te lopen om, na een korte wachtperiode met de snuit in het bakje, een beloning te krijgen. Er zijn twee varianten van deze taak ontworpen: in de eerste variant wordt de geur gekoppeld aan een lekkere suiker oplossing met een specifieke grootte of aan een vaste hoeveelheid bittere quinine oplossing, in de tweede variant aan een specifieke kans op een beloning. Bij het uitoefenen van deze taak leert het dier dat de geur een voorspellende waarde heeft voor de beloning die zal volgen. Bijvoorbeeld: in de kannstaak leert het dier welke geur altijd, maar ook welke geur nooit wordt gevolgd door een beloning, zodat hij naar verloop van tijd zal stoppen met het benaderen van het beloningsbakje na het ruiken van de geur die nooit wordt beloond. Terwijl het dier deze geurassociaties leert, wordt de vuuractiviteit van vele hersencellen (neuronen) tegelijkertijd gemeten (‘ensemble recordings’), en kunnen de neuronale correlaten van het voorspellen van beloningen in de wachtperiode bestudeerd worden, zowel op het niveau van de enkele cel als op het niveau van groepen cellen.

Uit de resultaten is onder andere gebleken dat voorspellende beloningsinformatie omtrent identiteit, grootte of kans op beloning verwerkt wordt
in de orbitofrontale cortex: neuronale activiteit representeert de voorspelling van de uitkomst van het bezoek aan het beloningsbakje. De neuronale correlaten van deze informatie zijn niet alleen terug te vinden op het niveau van de individuele cellen, maar de activiteit van groepen orbitofrontale cellen blijkt deze informatie ook te herbergen. Het feit dat deze beloningsinformatie door populaties van cellen wordt geregistreerd geeft aan dat deze informatie belangrijk is: de representatie van de beloningvoorspelling is op deze manier dermate robuust dat andere hersengebieden deze weer verder kunnen verwerken, zodat deze kennis uiteindelijk geëffectueerd kan worden in de gedragshandelingen. Enigszins verassend was de vinding dat voorspellende beloningsinformatie niet alleen aanwezig is tijdens de korte wachtperiode voordat de beloning gegeven wordt, maar ook tijdens het moment dat het dier naar het beloningsbakje toeloopt. De aanwezigheid van voorspellende informatie in beide fases van de taak laat zien dat de orbitofrontale cortex deze informatie continu paraat heeft, wat maakt dat gedrag snel kan worden aangepast wanneer veranderingen in de omgeving daar om vragen. Uit de analyses is verder gebleken dat de representatie op populatie niveau in de orbitofrontale cortex redundante is, wat betekent dat voorspellende beloningsinformatie door vele cellen gedistribueerd over de orbitofrontale cortex wordt gecodeerd, waarbij verlies van cellen niet direct leidt tot verlies aan informatie. Voorts is gebleken dat de beloningsvoorspelling zich ontwikkelt gedurende de uitvoer van de taak, parallel aan het leren van de geurassociaties. Dit betekent dat in het begin van de taak, wanneer het dier de geurassociaties nog moet leren, de komende beloning net significant boven kansniveau uit de activiteit van de populatie kan worden voorspeld, maar dat met ervaring van het dier in de taak deze predictie allengs beter wordt.

Ook is in dit proefschrift aandacht besteed aan het verder ontwikkelen van de techniek van ensemble recordings. Hiertoe is de combidrive ontwikkeld: een instrument dat ensemble recordings combineert met de techniek van microdialyse, waardoor het mogelijk wordt om tijdens het meten van de activiteit van groepen cellen de neuronale activiteit van dezelfde neuronen te beïnvloeden door het lokaal toedienen van drugs. Op deze manier wordt het mogelijk om niet alleen de effecten van drugs, maar ook de interactie tussen neuronen en de natuurlijke chemische boodschapperstoffen in de hersenen (i.e. neurotransmitters), te bestuderen. Bij het valideren van de combidrive is gebleken dat deze twee technieken goed kunnen worden gecombineerd; om dit te bekijken zijn de effecten van drie verschillende drugs (TTX, lidocaine en muscimol) op de activiteit van neuronen bestudeerd en met elkaar vergeleken. Het is bekend dat al deze stoffen de activiteit van neuronen blokkeren, maar met een ander werkingsmechanisme. Ook in deze studie blokkeerden alle stoffen de activiteit van neuronen, maar de mate waarin dit gebeurde verschilde tussen de drugs. Het effect van TTX en muscimol op de celactiviteit was vergelijkbaar: een complete, langdurige blokkade van de activiteit,
terwijl het effect van lidocaine minder sterk was en korter van duur. Deze resultaten leiden tot de conclusie dat wanneer men bijvoorbeeld een hersengebied wil inactiveren om te onderzoeken wat de bijdrage van dit hersengebied is aan het leren van een bepaalde gedragstaak, voor een langdurige inactivatie muscimol de beste drugs is om te gebruiken (gezien wat andere, meer nadelige eigenschappen van TTX). Voor een kortstondige inactivatie blijkt lidocaine de meest geschikte drugs te zijn. Naast het onderzoeken van de effecten van drugs op neuronale activiteit verschaf de combidrive de mogelijkheid om in toekomstige experimenten de invloed van neurotransmitters op de plasticiteit van de activiteit van neuronen te bestuderen. Het leren kennen en begrijpen van dergelijke interacties tijdens leertaken zoals de twee taken die hierboven zijn beschreven zijn van groot belang voor het verder begrijpen van de cellulaire en neurochemische basis van aandoeningen waarbij bijvoorbeeld het verwerken van beloningsinformatie is verstoord – inzichten die kunnen leiden tot een meer effectieve behandeling van deze affectieve stoornissen.
List of Publications


Van Duuren, E., van der Plasse, G., Joosten, R.N.J.M.A., Feenstra, M.G.P. and Pennartz, C.M.A. Single cell and population coding of expected reward probability in the orbitofrontal cortex of the rat. Submitted
Famous Last Words

Op deze plaats wil ik graag alle mensen bedanken die op een positieve manier hebben bijgedragen aan het maken van dit proefschrift. Verder gaat mijn speciale dank uit naar de mensen die al die jaren de technische en praktische ondersteuning hebben geleverd: zonder jullie werk zou dit proefschrift er niet zijn gekomen.

Esther

Amsterdam, 7 juli 2008
Per correre migliori acque alza le vele
omai la navicella del mio ingegno,
che lascia dietro a sé mar sí crudele

Om koers te zetten over still waatren
hijst ’t scheepje van mijn geest nu blij de zeilen,
de zee van ’t bitterst leed voorgoed ontvarend

Dante, 1321
Neural representation of reward information: coding by single cells and populations in rat orbitofrontal cortex

Esther van Duuren