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

Citation for published version (APA):
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
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_pur), which is indicative of a single cell’s response variability attributable to differences in reward probability, was calculated by:

\[ V = \frac{N}{N-1} \frac{\overline{r^2} - (\overline{r})^2}{r^2} \]  

(Eq.1)

with

\[ \overline{r} = \frac{1}{N} \sum_{j=1}^{N} r_j \]

and

\[ \overline{r^2} = \frac{1}{N} \sum_{j=1}^{N} r_j^2 \]

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, \( \overline{r} \) now represents the mean firing rate in the population and \( \overline{r^2} \) 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 (\( x \)) and the decoding block (\( y \)), with \( x_i \) and \( y_i \) 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:
\[ \sum_{i=1}^{N} x_i y_i \]  
(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:

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

**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 ~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.** 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 \((> 15\% \text{ or } < -15\%)\). This showed that only 2% of the cells \((n = 7)\) made such extreme \((\text{positive or negative})\) 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.