Opinion TRENDS in Cognitive Sciences Vol.8 No.6 June 2004

Neuronal representations of cognitive state: reward or attention?

John H.R. Maunsell

Howard Hughes Medical Institute & Baylor College of Medicine, Division of Neuroscience, One Baylor Plaza, S-603, Houston, TX, USA

The effects of spatial or featural attention on the activity of neurons have been studied in many experiments that have used a variety of neurophysiological approaches. Other experiments have examined how expectations about reward are represented in neuronal activity in various brain regions. Although attention and reward are distinct concepts, I argue here that many neurophysiological experiments on attention and reward do not permit a clean dissociation between the two. This problem arises in part because reward contingencies are the only parameter manipulated in any of these experiments. I describe how attention and reward expectations have been confounded, giving rise to uncertainty about how signals related to attention and reward are distributed in the brain.

Characterizing the cognitive information conveyed by the activity of neurons is more challenging than characterizing sensory or motor signals. Sensory and motor signals can be explored by examining how changes in neuronal activity relate to changes in specific physical attributes of a sensory stimulus or motor response. For cognitive signals, however, we must depend on our concepts of cognitive processing, which might ultimately prove to be poor descriptions of the signals used by the brain. It is therefore encouraging to see how much progress has been made investigating neuronal signals related to cognitive processes such as attention, memory and motivation.

Many studies have examined the effects of attention at the level of single neurons, and reward is also being examined at this level with increasing frequency. Although attention and reward are distinct behavioral parameters and have been the subjects of essentially independent lines of research, careful consideration of the methods used and the results obtained suggests that some studies of attention and reward might have been looking at exactly the same neuronal signals. This possibility has not been widely appreciated, yet it has important implications for understanding the types of cognitive signals that control behaviors. Understanding of higher brain function will depend on accurate descriptions of the behaviorally relevant information encoded in the brain. The functional relationships between different populations of neurons cannot be unraveled without good terminology for describing the different signals they contain.

Attention and reward each embody large ranges of phenomena, and I will be considering only limited subsets of each. Studies of neuronal representations of reward distinguish between the immediate detection or perception of rewards received, expectations about future rewards (immediate and long term), and discrepancies between expected and actual rewards [1,2]. Studies of neuronal representations of attention distinguish between arousal or vigilance, selective processing of particular sensory stimuli, and supervisory systems that control motor plans that unfold over extended periods [3]. The specific topic I address here are problems related to distinguishing neuronal signals related to expectations about future rewards from those related to attention to particular locations, stimuli or stimulus features. I will focus on single unit studies in monkeys, although the issues apply equally to studies using any neurophysiological approach and any species.

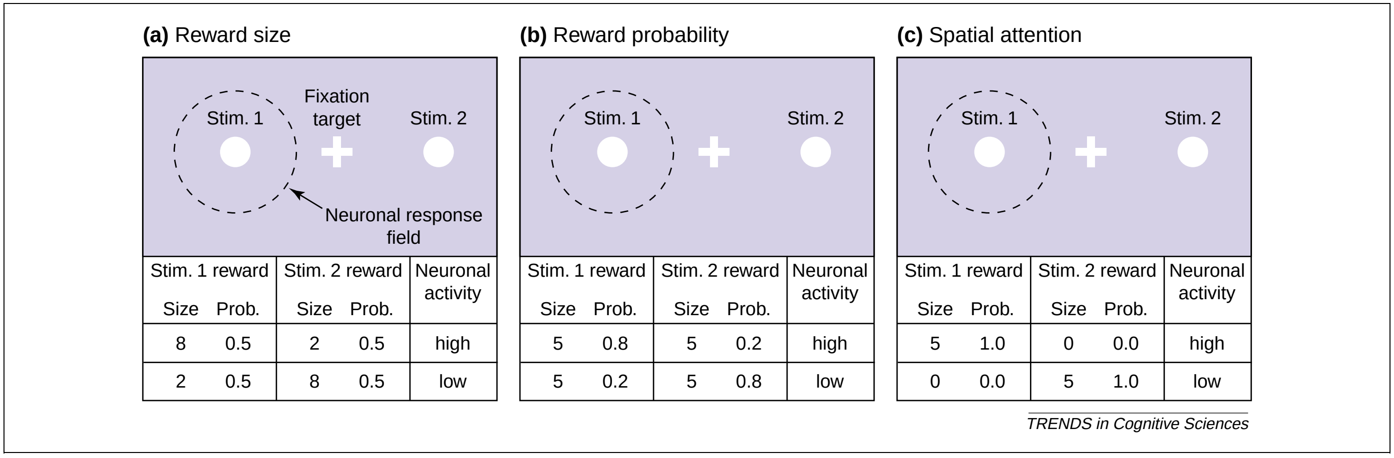
Effects of reward expectation on neuronal activity

Several studies have examined how a monkey's expectations about reward can influence the activity of individual neurons. In most such experiments, expectations are varied by having the animal do blocks of trials in which reward conditions are kept constant within blocks (allowing stable expectations to develop) but varied between blocks (allowing comparison of the effects of different expectations). Rewards are typically manipulated by changing either the amount of reward delivered for a correct response (Figure 1a) or the probability that the reward will be associated with a particular location or stimulus (Figure 1b). Both manipulations affect the activity of individual neurons, with greater reward magnitude or probability typically associated with more action potentials.

For example, Platt and Glimcher [4] examined the activity of neurons in the lateral intraparietal area (LIP), a cortical area involved in guiding saccadic eye movements. Monkeys were given a task in which the amount of reward associated with different visual stimuli was varied. The animal had to fixate on a central spot while two stimuli were presented, one inside the response field of the LIP neuron being record, and one outside. The animal then received a cue telling it which stimulus it should make a saccade to, but had to wait for a go signal before making its response. As had been shown previously, LIP neurons were more active when the animal was cued to make a saccade to the stimulus in the response field.

Corresponding author: John H.R. Maunsell (maunsell@bcm.tmc.edu).

www.sciencedirect.com 1364-6613/$ - see front matter © 2004 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2004.04.003



(a) Reward size

Stim. 1 reward		Stim. 2 reward		Neuronal activity
Size	Prob.	Size	Prob.	
8	0.5	2	0.5	high
2	0.5	8	0.5	low

(b) Reward probability

Stim. 1 reward		Stim. 2 reward		Neuronal activity
Size	Prob.	Size	Prob.	
5	0.8	5	0.2	high
5	0.2	5	0.8	low

(c) Spatial attention

Stim. 1 reward		Stim. 2 reward		Neuronal activity
Size	Prob.	Size	Prob.	
5	1.0	0	0.0	high
0	0.0	5	1.0	low

Figure 1. Schematics of representative reward and attention tasks. These schematics do not describe any specific experiments, but illustrate the essence of the stimuli and reward contingencies used in different types of experiments. **(a)** Representative reward size task. The upper half of the panel represents a visual display comprising a central fixation target and two other stimuli, one of which lies within the response field of a neuron being recorded. When reward size is manipulated, either stimulus is equally likely to be selected as a response target on a given trial. In some blocks of trials, correct responses to one target receive a large reward (8), whereas correct responses to the other target receive a small reward (2). In other blocks of trials, the size of rewards is reversed. The activity of many neurons is modulated by reward size, with more activity during trials in blocks with large rewards associated with the stimulus in the response field. **(b)** Representative reward probability task. The configuration of the visual stimuli and response field is the same as in (a). When reward probability is manipulated, rewards are always the same size. In some blocks of trials one stimulus is more likely to be selected as a response target (0.8), whereas in other blocks of trials that stimulus is less likely to be selected as a target (0.2). The activity of many neurons is modulated by reward probability, with more activity during trials in which the stimulus in the response field is more likely to be selected. **(c)** Representative spatial attention task. The configuration of the visual stimuli and response field is the same as in (a). In typical spatial attention experiments, targets appear on both sides, but rewards are given only for responses to one side. Responses to targets on the wrong side (distractors) are unrewarded. The rewarded side alternates between blocks of trials. Neuronal activity is stronger during blocks in which the stimulus in the response field is being rewarded. The structure of the three tasks shown is very similar, and in all cases neuronal activity is stronger when the stimulus in the response field is associated with more reward.

In one version of the task, Platt and Glimcher varied the reward size associated with each stimulus between blocks of trials (Figure 1a). In addition to the previously established effect, they saw that the activity of LIP neurons was greater in blocks in which the stimulus in a neuron's response field was associated with larger rewards and less in blocks in which the stimulus was associated with smaller rewards. In a second version of the task, they kept the size of the reward fixed, but between blocks they changed the probability that the stimulus in the response field would be rewarded from 20% of the trials to 80% of the trials (Figure 1b). Most LIP neurons were more active when it was more likely that a saccade to the response field would be rewarded. The modulations induced by changes in reward probabilities were similar to those evoked by changes in reward size. Platt and Glimcher interpreted these changes in activity as showing that LIP neurons reflect reward-related variables associated expected gain and outcome probability.

Many other reports have described similar reward-related modulations in LIP and other brain regions. Another study of modulations of LIP activity resulting from reward manipulations described those effects as representing 'experienced value' [5]. Preliminary recordings from LIP suggest that increasing the reward associated with all targets has effects that are similar to increasing the probability that a stimulus in response field will be associated with a reward [6]. Effects of expected reward size or reward type similar to those described in LIP have also been shown in the prefrontal cortex [7–11] and superior colliculus [12], both of which contribute to guiding eye movements. Several studies have described neuronal signals related to expected reward in the basal ganglia using similar designs [13–17].

Each of these studies interpreted neuronal modulations as associated with reward contingencies and the animal's expectations of the amount of reward it was likely to receive. Although these interpretations are parsimonious given the manipulations performed, alternative interpretations are possible. Specifically, the phenomena described in experiments of reward manipulation may be closely related to those seen in experiments that examined neuronal mechanisms related to attention.

Reward expectation or attention?

The effect of attention on neuronal responses has been studied extensively (reviewed in [18–21]). Single unit studies in visual cortex have shown that increased attention to a stimulus typically increases the responses of neurons that represent that stimulus [22]. Many such studies have examined the effect of attending to a particular stimulus or location in space, but it has been shown that attention to a particular visual feature, such as a direction of motion, can also selectively enhance the responses of neurons that respond to that feature, whether or not their receptive field is in the location being attended [23–25]. Although most studies of attention have examined its effect on responses to sensory stimuli, attention also has been shown to modulate the activity of neurons when no stimulus is present (e.g. [26–28]).

At first glance, studies of attention seem to address issues that are distant from those addressed in experiments that examine the effects of reward expectations. Nevertheless, the structure of attention and reward experiments is similar (Figure 1). In particular, the only tool used to control attention is the manipulation of rewards. For example, when the effects of spatial attention are examined, subjects are motivated to direct attention to one location or another only by expectations about which location is more likely to be associated with a reward (Figure 1c). Such reward manipulations reliably lead to shifts in attention, as documented by changes in behavioral performance (e.g. [29,30]), and such studies invariably interpret changes in neuronal activity as related to allocation of attention. However, these experiments typically provide no basis for assigning changes preferentially to attention or to expectations about reward. In most cases, attention-related modulation could equally well be described in terms of expectation about rewards because the two are inextricably confounded.

Conversely, the experiments described above that examined the effects of reward can be equally well described in terms of attention. It is natural to expect that subjects will allocate more attention to those stimuli or locations that are more likely to be rewarding, and the tasks used in studies of reward expectation are not designed in a way that would permit a distinction between the two.

The neurophysiological and behavioral consequences of shifting attention and changing reward expectations do not provide a clear basis for distinguishing between them. Increasing the reward associated with a stimulus or increasing the amount of attention allocated to a stimulus will usually (but not always) produce stronger neuronal responses to that stimulus. Attention and reward-contingency both can increase the responses of neurons that have receptive fields overlapping the attended or reward-contingent location even when no stimulus is present. Behavioral performance, measured by reaction times or detection thresholds, is superior for attended stimuli (reviewed in [31]), and similar improvements in behavioral performance are seen for stimuli associated with larger rewards [10,11,15–17,32,33].

Although there are differences between the designs of most attention and reward experiments, these do not provide a basis for attributing affects to one or the other. For example, most attention studies manipulate attention in an all-or-none way by rewarding one target reliably and the others not at all, whereas some reward studies have adjusted reward parametrically. The latter show that neuronal modulations vary continuously with expected reward (e.g. [4]). However, the few neurophysiological studies that have varied the difficulty of a spatial attention task have shown that neuronal modulations by attention vary depending on task demands [29,30,34]. Another potential difference between reward and attention studies is that reward studies sometimes vary the rewards for responses to all stimuli [6]. The reward effects seen in these studies therefore might not be spatially specific like those shown in attention experiments, and instead more related to arousal or vigilance. Although that is possible, increased rewards might instead cause subjects to focus more attention to both the reward targets, at the expense of other distractions in the visual field.

Favoring one interpretation

Most studies of attention or reward expectations have not addressed the potential for confusion between them, and have interpreted neuronal signals as either attention-related or reward-related without considering alternatives. However, a few reports did consider the problem and argued specifically for one interpretation. Although such views might be correct, the data do not provide strong support.

Kawagoe and colleagues [15] sought to distinguish their reward-related neuronal modulations from spatial attention. They suggested that their results could not be explained by attention because their subjects were required to attend to every stimulus, whether it was rewarded or not. Although their animals did have to attend to every stimulus, there is no reason to expect that monkeys would attend equally to rewarded and unrewarded stimuli. The results are consistent with the subjects paying more attention to stimuli that were associated with rewards. Humans distribute more attention to targets that are more likely to be associated with reward [31].

Platt and Glimcher also argued that the LIP signals they recorded were related to reward rather than attention [4]. They suggested that attention is unlikely to influence LIP because in some situations LIP neurons appear insensitive to the behavioral relevance of stimuli that are not the targets of eye movements. They showed this in a different experiment that used a task in which a stimulus in the response field was irrelevant in some trials, and in other trials was the cue that instructed the animal when to make a saccade [35]. Because no difference in neuronal responses was seen between these conditions, they concluded that attention was unlikely to play a role in determining the activity of LIP neurons.

There are two reasons why a failure to see attention-related modulation in this experiment cannot rule out spatial attention in all conditions. First, the animal might not have changed its distribution of spatial attention between the two conditions. With the particular task design that was used, the animal could have successfully completed all trials by always monitoring the distractor and the fixation spot and making a response whenever either went out. Second, even if the animal directed more attention to the distractor during the trials in which it was relevant, the task might not have been sufficiently demanding to command much attention. The amount neurons are modulated by spatial attention depends on the effort the animal must direct to the task [29,30], and detecting the disappearance of an isolated light might have been insufficiently challenging to produce a readily detectable attentional modulation. Notably, responses of LIP neurons to irrelevant distractor stimuli are affected by behavioral state under other conditions [36].

Platt and Glimcher also suggested that the reward-related neuronal activity they saw was inconsistent with modulation of visually evoked activity by attention because the changes in neuronal activity occurred before the stimuli appeared. This is true, but, as noted above, attention can modulate the activity of neurons throughout visual cortex even when no stimulus is present.

Distinguishing reward and attention

Recent experiments do not distinguish whether neuronal signals are related to attention or to expectations about reward. Can changes in neuronal activity be assigned unequivocally to one or the other? A complete treatment of the issues involved in defining and distinguishing cognitive signals such as these is beyond the scope of this article. Nevertheless, is it helpful to consider the range of possibilities for signals related to attention and reward, and to suggest approaches that might serve to distinguish them.

It is possible the brain does not have distinct neuronal signals related to attention and reward expectation. This possibility becomes more likely if we take the broadest definition of reward. A broad definition of reward would include not only the immediate primary rewards (e.g. apple juice on a given trial), but also all other factors that motivate performance, such as preference for a novel location or stimulus, the satisfaction of performing well or the desire to complete a day's work. If reward is defined to include all motivating factors, then there may be no differences between attention and expectation of reward: the allocation of attention might be an exact representation of the subject's current assessment of what is likely to be rewarding.

Although it is possible that attention might not be distinct from a broadly defined expectation of reward, all the experiments discussed above took a narrower definition of reward. Specifically, they examined effects of the amount of reward delivered at the each of each trial: the immediate primary reward. It seems likely that different neurons and brain regions are affected to differently degrees by attention and such reward expectations (at least in primates and probably in all mammals).

Carefully designed experiments might distinguish the relative contributions of immediate primary reward and attention. An obvious approach would be to manipulate reward and attention independently. Changes in task difficulty can affect the amount of attention allocated to different stimuli in ways that can be measured behaviorally [29,30]. Although these studies did not explicitly keep immediate primary rewards constant when adjusting task difficult, if differences in neuronal activity were seen when attention was varied and immediate primary reward was held constant it would support the view that the effect was associated specifically with attention. This approach might be extended to test for signals related to expectations about immediate primary reward. An increase in task difficulty that was coupled with a decrease in immediate primary reward might provide a means to dissociate the two factors, and thereby identify the relative influence of attention and reward expectations for different neurons and structures. For example, a neuron with activity that was positively correlated with task difficulty (and therefore negatively correlated with immediate primary reward) could be described as more influenced by attention, whereas the opposite outcome would provide evidence for greater influence by immediate primary reward.

Although I have focused this discussion on single unit studies in monkeys, the issue exists for any neurophysiological approaches to these aspects of attention and reward. For example, recent imaging studies have examined the effect of reward expectations using monetary or taste rewards [32,37]. Such reward manipulations can be expected to affect the allocation of attention in space and time. Many other imaging studies in humans have described activity associated with attention (reviewed in [21]). These experiments typically do not use an immediate primary reward, so the potential confound with reward expectation is not as pronounced as in the animal studies described above. Nevertheless, in these experiments attention to the correct stimulus must be coupled to an expectation of a less narrowly defined reward.

Here I have considered only attention and reward, but there is also the potential for similar confounds in the interpretation of other cognitive signals. Because reward contingencies are the fundamental tool for controlling behavior in neurophysiological experiments, signals seen in very different task conditions could have common elements. For example, in the monkey superior colliculus, another structure involved in the control of eye movements, the activity of some neurons also varies with the probability that a particular eye movement will be required [38–40]. Although this modulation has been interpreted as a motor preparation signal, it might be more closely related to a representation of reward expected at that site, or the allocation of attention to that site. The possibility that these signals could be equally well ascribed to spatial attention or motor preparation has previously been raised by Sparks [41], and Roesch and Olson [7] have discussed how signals described as reward-related might instead be signals of motivation-related modulation of motor preparation. Many signals that have been attributed exclusively to attention, behavioral salience, motivation, intention, motor preparation and reward expectation might ultimately prove to comprise a mixture of different signals.

Conclusion

In summary, I have argued that some experiments examining neuronal representations of attention or reward expectation might in fact have been measuring either of those parameters. Given our limited understanding of the organization of the neuronal representations that support cognition, clarifying our terminologies and descriptions of neuronal activity is crucial for identifying the signals that are most important for different cognitive functions. Efforts to distinguish the relative contribution of attention and reward in different neuronal populations and brain regions will be a valuable step towards establishing a precise description of the neuronal signals that support higher brain functions.

Acknowledgements

I thank John A. Assad, William H. Bosking, Raika Pancaroglu and David L. Sparks insightful discussions and helpful comments on earlier versions of this manuscript. Supported by NIH R01 EY05911. J.H.R.M. is an Investigator with the Howard Hughes Medical Institute.

References

1 Berridge, K.C. (1996) Food reward: brain substrates of wanting and liking. *Neurosci. Biobehav. Rev.* 20, 1–25

2 Schultz, W. (2000) Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207

3 Pashler, H.E. (1998) *The Psychology of Attention*, The MIT Press

4 Platt, M.L. and Glimcher, P.W. (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238

5 Newsome, W.T. (2003) Decision-making and the neural representation of 'experienced value'. *Soc. Neurosci. Abstr. (Abstract Viewer)* Program No. 436

6 Bendiksby, M.S. and Platt, M.L. (2003) Motivation focuses attention and enhances neuronal selectivity in parietal cortex. *Soc. Neurosci. Abstr. (Abstract Viewer)* Program No. 385.5

7 Roesch, M.R. and Olson, C.R. (2003) Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J. Neurophysiol.* 90, 1766–1789

8 Watanabe, M. (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382, 629–632

9 Watanabe, M. (1990) Prefrontal unit activity during associative learning in the monkey. *Exp. Brain Res.* 80, 296–309

10 Leon, M.I. and Shadlen, M.N. (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24, 415–425

11 Kobayashi, S. *et al.* (2002) Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J. Neurophysiol.* 87, 1488–1498

12 Ikeda, T. and Hikosaka, H. (2003) Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* 39, 693–700

13 Sato, M. and Hikosaka, H. (2002) Role of primate substantia nigra pars reticulata in reward-oriented saccadic eye movement. *J. Neurosci.* 22, 2363–2373

14 Watanabe, K. *et al.* (2003) Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. *J. Neurosci.* 23, 10052–10057

15 Kawagoe, R. *et al.* (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1, 411–416

16 Lauwereyns, J. *et al.* (2002) A neural correlate of response bias in monkey caudate nucleus. *Nature* 418, 413–417

17 Hollerman, J.R. *et al.* (1998) Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* 80, 947–963

18 Olson, C.R. (2001) Object-based vision and attention in primates. *Curr. Opin. Neurobiol.* 11, 171–179

19 Treue, S. (2001) Neural correlates of attention in primate visual cortex. *Trends Neurosci.* 24, 295–300

20 Itti, L. and Koch, C. (2001) Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203

21 Yantis, S. and Serences, J.T. (2003) Cortical mechanisms of space-based and object-based attentional control. *Curr. Opin. Neurobiol.* 13, 187–193

22 Maunsell, J.H.R. and Cook, E.P. (2002) The role of attention in visual processing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1063–1072

23 Chelazzi, L. *et al.* (1998) Responses of neurons in inferotemporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940

24 Treue, S. and Martínez Trujillo, J.C. (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579

25 McAdams, C.J. and Maunsell, J.H.R. (2000) Attention to both space and feature modulates neuronal responses in macaque area V4. *J. Neurophysiol.* 83, 1751–1755

26 Fuster, J.M. and Jervey, J.P. (1982) Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J. Neurosci.* 2, 361–375

27 Luck, S.J. *et al.* (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42

28 Eskandar, E.N. and Assad, J.A. (1999) Dissociation of visual motor and predictive signals in parietal cortex during visual guidance. *Nat. Neurosci.* 2, 88–93

29 Spitzer, H. *et al.* (1988) Increased attention enhances both behavioral and neuronal performance. *Science* 240, 338–340

30 Boudreau, C.E. and Maunsell, J.H.R. (2001) Is spatial attention a limited resource for V4 neurons? *Soc. Neurosci. Abstr.* 27, 5743

31 Posner, M.I. (1980) Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25

32 Ramnani, N. and Miall, R.C. (2003) Instructed delay activity in the human prefrontal cortex is modulated by monetary reward expectation. *Cereb. Cortex* 13, 318–327

33 Lauwereyns, J. *et al.* (2002) Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron* 33, 463–473

34 Spitzer, H. and Richmond, B.J. (1991) Task difficulty: ignoring, attending to, and discriminating a visual stimulus yield progressively more activity in inferior temporal neurons. *Exp. Brain Res.* 83, 340–348

35 Platt, M.L. and Glimcher, P.W. (1997) Responses of intraparietal neurons to saccadic targets and visual distractors. *J. Neurophysiol.* 78, 1574–1589

36 Powell, K.D. and Goldberg, M.E. (2000) Response of neurons in the lateral intraparietal area to a distractor flashed during the delay period of a memory-guided saccade. *J. Neurophysiol.* 84, 301–310

37 O'Doherty, J.P. *et al.* (2002) Neural responses during anticipation of a primary taste reward. *Neuron* 33, 815–826

38 Basso, M.A. and Wurtz, R.H. (1997) Modulation of neuronal activity by target uncertainty. *Nature* 389, 66–69

39 Basso, M.A. and Wurtz, R.H. (1998) Modulation of neuronal activity in superior colliculus by changes in target probability. *J. Neurosci.* 18, 7519–7534

40 Dorris, M.C. and Munoz, D. (1998) Saccadic probability influences motor preparation signals and time to saccadic initiation. *J. Neurosci.* 18, 7015–7026

41 Sparks, D.L. (1999) Conceptual issues related to the role of the superior colliculus in the control of gaze. *Curr. Opin. Neurobiol.* 9, 698–707