Studying metacognitive processes at the single-neuron level

Paul G. Middlebrooks[1], Zachary M. Abzug[2], and Marc A. Sommer[2]

[1]Department of Psychology, Vanderbilt University, Nashville, TN 37240, USA
[2]Department of Biomedical Engineering, the Center for Cognitive Neuroscience, and the Duke Institute for Brain Research, Duke University, Durham, NC 27708, USA

Correspondence:
Paul G. Middlebrooks
Department of Psychology
301 Wilson Hall
Vanderbilt University
Nashville, TN 37240
USA
Phone: 615-322-2874
E-mail: paul.g.middlebrooks@vanderbilt.edu

Acknowledgments:
Supported by the NIH (NIMH Kirschstein NRSA F31 MH087094 to P.G.M. and NEI R01 EY017592 to M.A.S.) and the NSF (Graduate Research Fellowship to Z.M.A.).

Introduction

Over the past few decades, strides have been made toward understanding how higher-level cognitive processes are mediated by neuronal spiking activity. Neuronal correlates of functions such as attention, executive control, working memory, decision-making, and reward processing have all been elucidated, to an impressive level of detail, at the single cell and circuit levels. This explosion in neuroscience-based discovery has depended crucially on nonhuman animal (*animal*, hereafter) models of the behaviors and processes under question. Developing animal models becomes a greater challenge for cognitive functions that approach the complexity of those arguably unique to humans.

A prime example is metacognition. As reviewed in the *Foundations of metacognition* section of this volume, we know that humans engage in complex metacognitive behaviors. A metacognitive process is by definition *about* one of our own cognitive process, and is often referred to as "thinking about thinking". Hence it is not surprising that metacognition is often associated with our subjective or conscious sense of self (e.g. Nelson 1996). Beyond the human brain, evidence for metacognition is less clear. There is, as yet, no definitive evidence that animals experience a subjective awareness, or a continuity of mental experience, similar to our own. Consequently, many investigators conclude that animals must not possess metacognition as humans do. Recent behavioral evidence, however, makes a case for some degree of metacognitive capability in a variety of animal species.

Early attempts to test animals' metacognitive skills used paradigms that analyzed relatively simple metacognitive behaviors. Subsequent single neuron studies have followed suit by developing streamlined tasks that are quick in duration, austere in terms of sensory stimulation and motor response, and balanced as much as possible by control conditions. A subtle issue is that animals, primates in particular, are notorious for finding the simplest strategy for accomplishing a task, rather than the strategy desired by the experimenter. It is important to verify that subjects are not "cheating" at metacognitive tasks by using external cues (e.g. visual differences between conditions or motor differences between responses) instead of internal perceptions and memories (see Kornell 2013 for a review). Taking all of these considerations into account, investigators have designed a variety of tasks for evaluating the association between neuronal activity and metacognition. All of the tasks to date consist of a "cognitive" period followed immediately by a "metacognitive" period. Likewise, all of them use a

confidence response or surrogate thereof. Though the single neuron studies we describe in this chapter do not attempt to investigate the richness of metacognitive skills that we take for granted as humans, they serve as a starting point for what hopefully will continue to develop into a mechanistic neuronal account of metacognition in general.

We begin by describing the behavioral tasks used to test metacognition in animals, with a focus on those used in single neuron studies. Next we discuss a few possible ways neuronal firing rates might encode metacognitive processes. The bulk of the chapter is then devoted to describing and critiquing three studies that examined metacognitive processes at the single neuron level. Finally, we discuss the implications and limitations of these and future single neuron studies of metacognition in animals.

Streamlined metacognitive paradigms: opt-out and betting tasks

Before describing the tasks and experiments for studying the neuronal basis of metacognition, it is important to consider how the field arrived at this point. Before the term "metacognition" was used, experiments were performed in which subjects were asked to assess their own "feeling of knowing" whether an item was in their memory even though they could not presently recall it (Hart 1965). In the 1970s John Flavell coined the terms "metamemory" (Flavell 1971) and "metacognition" (Flavell 1976) in his studies of child development. Subsequently, Nelson and Narens (1990) developed a systematic framework for the study of metacognition that has been widely used since. In their framework, a distinction was made between two types of metacognitive processes. A 'monitoring' process *receives* information about ongoing cognitive operations. For example, a student might experience a sense of whether she is correctly recalling a list of memorized words. A 'control' process *provides* information to ongoing cognitive processes and allows a subject to strategically plan. For example, a student can estimate the effort it will take to memorize a list of words. Within these two main divisions, monitoring and control processes, Nelson and Narens' framework provides multiple sub-categories that classify metacognitive processes according to factors such as which facet of a cognitive process is interacting with the metacognitive process, the responses required by the subject, and whether the response occurs while a subject is learning or recalling material. A main goal for the neuroscientific study of metacognition is to embrace these psychological principles while

adapting the tasks for use in non-verbal subjects (animals) in settings that demand speed and efficiency (single neuron recordings).

Experiments to test animals' metacognitive abilities have almost all focused on monitoring processes. In the 1990s David Smith and colleagues tested whether animals (dolphins in particular) could monitor their own uncertainty (Smith et al 1995) during decision-making. Variously referred to as "uncertainty monitoring", "decline", or "escape" tasks, we will refer to this general class as "opt-out" tasks (Fig 1a). Animals are required to perform a primary decision

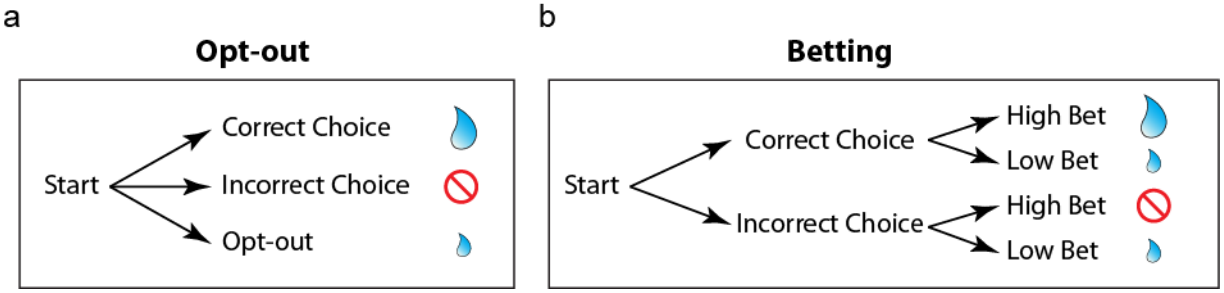

Figure 1: Metacognitive monitoring tasks. (a) Schematic of opt-out task paradigms. Opt-out tasks generally involve a two-choice perceptual discrimination. On some proportion of trials, a third opt-out target appears. Selection of the opt-out target results in a small but ensured reward. Participants utilizing a metacognitive strategy should select the opt-out target more often on more difficult trials, and make more accurate responses on trials the opt-out is offered. (b) Schematic of betting task paradigms. Betting tasks generally involve choice stage followed by a betting stage. Selection of the high bet target results in a large reward after a correct response, and no reward after an incorrect response. Selection of the low bet target results in a small but ensured reward. Participants using a metacognitive strategy should select the high bet target more often after correct decisions. Unlike opt-out tasks, betting tasks require a primary task decision on every trial.

task, such as making a two-choice perceptual discrimination. Reward is earned for correct responses. On some trials, an additional "opt-out" response choice is offered that, when selected, always delivers minimal reward (most studies have offered the opt-out concurrent with the primary task response targets, but see Hampton 2001 for an important innovation in which the opt-out is presented before the animal responds to the primary task). The animal thus can opt-out of the primary task, which will earn either a large reward if correct or no reward if incorrect, and instead receive an ensured small reward. The basic premise is that an animal capable of monitoring its own uncertainty will select the opt-out response more often during difficult trials. Likewise, when the animal does make a response to the primary task, accuracy will be higher on trials in which the opt-out response was offered than those when the animal was forced to perform the primary task. Multiple species have been shown to opt-out in a manner consistent with the ability to monitor their uncertainty, including dolphins (Smith et al 1995), rats (Foote & Crystal 2007; Kepecs et al 2008), rhesus macaques (Shields et al 1997; Smith et al 1998;

Hampton 2001; Beran et al 2006; Kiani & Shadlen 2009; Tanaka & Funahashi 2013), orangutans (Suda-King 2008), and gorillas (Suda-King 2013).

Another line of studies used what we refer to as "betting" tasks (Fig 1b). Like opt-out tasks, betting tasks require performance of a primary task, such as making a two-choice perceptual discrimination. However, a response to the primary task is required on every trial. Reward is not earned immediately. Instead, once a response is made, there is an option to make either a high bet or a low bet. High bets earn a large reward after correct decisions and no reward (or a timeout punishment) after incorrect decisions. Low bets earn a small but ensured reward following either correct or incorrect decisions. The premise is that animals able to monitor their own decisions will bet high more often after correct responses and bet low more often after incorrect responses. Rhesus macaques perform betting tasks in a manner that suggests they are able to monitor their ongoing cognitive operations (Shields et al 2005; Son & Kornell 2005; Kornell et al 2007; Middlebrooks & Sommer 2012).

Comparative studies of metacognition sometimes use other tasks as well. For example, a series of experiments showed that gorillas, chimpanzees, bonobos, orangutans, and rhesus monkeys will seek information when it is needed to perform better on a task (Call & Carpenter 2001; Hampton et al 2004; Call 2010; Beran & Smith 2011), a type of metacognitive control behavior. The field of comparative metacognition continues to grow and improve methodologically. But from a neuroscientific point of view, the relative simplicity of opt-out and betting tasks is attractive. Both tasks are rooted in one of the most successful fields of neuroscience: decision-making. It is not surprising, then, that the single neurons studies described below employed tasks adapted from previous behavioral opt-out and betting task studies.

Mechanisms of metacognition

Before we highlight the single neuron studies related to metacognition, it is worthwhile to consider some theoretical perspectives that propose what a metacognitive neuronal signal might look like. In what follows, we outline a few neuronal coding schemes that plausibly underlie metacognition. In keeping with the rest of this chapter, we frame our discussion within the context of opt-out and/or betting tasks, in which a metacognitive judgment is in temporal

proximity to its referent cognitive behavior (a decision). In principle, though, the mechanisms could apply to other metacognitive tasks with some modification.

Researchers have approached the study of metacognition from decision-making sciences. The framework of decision-making can be extended to include how information in the signals that encode decisions could be used and/or further processed to encode related behavior, i.e. a metacognitive signal. Much of what we understand about how decisions are made, and especially how perceptual decisions are made, is encapsulated by a family of cognitive models known as sequential sampling models (Usher & McClelland 2001; Gold & Shadlen 2007; Smith & Ratcliff 2009). Rooted in signal detection theory (Green & Sweats 1960), sequential sampling models posit that available perceptual evidence is repeatedly sampled until the amount of evidence reaches a criterion threshold. At that point an appropriate response is executed. Neuronal firing rates in many regions of the brain resemble what sequential sampling models predict. Neurons' firing rates increase stochastically, at a rate proportional to available sensory evidence, until a consistent threshold is reached and a response is made. These regions include superior colliculus (Krauzlis & Dill 2002; Ratcliff et al 2003; Ratcliff et al 2007), lateral intraparietal cortex (Shadlen & Newsome 2001; Roitman & Shadlen 2002), dorsolateral prefrontal cortex (Kim & Shadlen 1999), the caudate nucleus (Ding & Gold 2010) and the frontal eye field (Hanes & Schall 1996; Purcell et al 2010; Ding & Gold 2012).

Psychologists have long thought confidence may be encoded simultaneously with decisions (e.g. Peirce & Jastrow 1884; Vickers 1979; Heath 1984; Link 1992). After all, we usually experience a sense of how well we're performing some task while in the middle of performing it. Building on that theme, various proposals have been made by which confidence in a decision could be encoded using the same mechanisms underlying the decision, at the same time the decision process occurs. In that case, brain regions encoding cognitive decisions could concurrently encode metacognitive decisions.

One example proposes that a metacognitive signal is encoded by comparing the firing rates of neurons selective for the alternative responses in a decision task (Fig 2a). Consider two hypothetical neurons, each neuron selective for one of the two alternative responses. During any given trial, the firing rate of the neuron selective for the response that was chosen (D_c, solid black) will likely be different than (and usually exceed) the firing rate of the neuron selective for the un-chosen response (D_u, dashed black). As the signals develop over the course of the trial, a

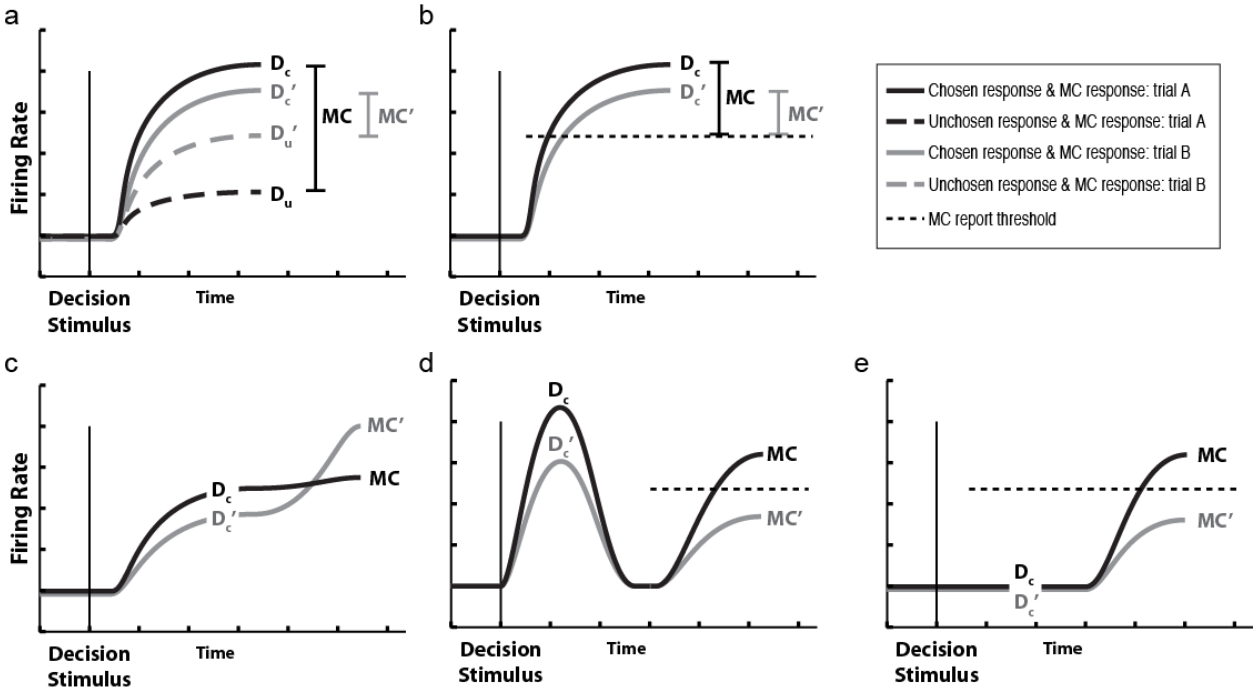

Figure 2: Possible mechanisms of metacognition. (a) Comparison of decision-related activity between response alternatives. In this model, confidence (MC) is a function of the difference between neural activity for the chosen (D_c) and unchosen (D_u) responses. (b) Comparison of decision-related activity to an independent threshold. In this model, confidence is a function of the difference between neural activity for the chosen response and the threshold. It is not dependent on activity for the unchosen response. (c) Sequential coding of decision and confidence. In this model, evidence continues to accumulate after the decision is made (D_c or D_c') to subsequently produce a confidence response (MC or MC', respectively). (d, e) Distinct coding of decision and confidence. In these models, confidence is encoded separately from the decision. Confidence can be encoded by the same neurons involved in the decision (d) or by distinct neurons (e).

metacognitive signal could be computed at any time by taking the difference between the two decision signals (MC). The magnitude of the difference could guide the metacognitive behavior to make. A large difference would correlate with a high level of confidence, for example. During a different, more difficult trial (grey lines), the firing rates of the two neurons may differ less and thus lead to a lower confidence rating (MC'). Models of this nature have accounted for human confidence ratings (e.g. van Zandt 2000; Merkle & van Zandt 2006; de Martino et al 2012).

A similar but alternate mechanism (Fig 2b) would compare firing rates of response selective neurons not with the firing rate of neurons selective for the other response, but with a threshold level representing the boundary between the alternative choices. During one trial (black line), a decision response (D_c) is made and the magnitude of the metacognitive signal (MC, black) is proportional to the difference between the neuron's firing rate and the threshold (black dashed line). During a more difficult trial (grey line), the same decision response (D_c') is

made but with less confidence (MC'). A model of this kind was used to describe rat behavior in one of the single neurons studies described below (Kepecs et al 2008).

The two mechanisms described assume that cognitive and metacognitive processes are encoded simultaneously within a brain area. An alternative proposal entails a sequence of processing stages within a single brain region, in which the metacognitive follows the cognitive process (Fig 2c). Like the previous models, this model exploits the sequential sampling framework. During a given trial (black line), evidence accumulates to a decision (D_c). The metacognitive process, however, depends on evidence continuing to accumulate until a metacognitive response is made (MC). During a different trial (grey line), the same decision may be made (D_c'), but further processing could lead to a higher confidence response (MC'). The stage processing mechanism accounts for human confidence responses (Pleskac & Busemeyer 2010) and for changes of mind after a decision has been made (van Zandt & Maldonado-Molina 2003). It should be noted that simultaneous and multi-stage models of metacognition are not mutually exclusive. A metacognitive process could be encoded in parallel with a cognitive process *and* after the cognitive process, and there is some evidence for such a scenario (Petrusic & Baranski 2003). In that study, humans' decision response times (RTs) increased when the task required confidence responses, suggesting the metacognitive process interacted with the cognitive process. In addition, confidence response RTs varied with the confidence level reported, suggesting some post-decisional processing took place as well.

So far we've considered extensions of the sequential sampling models that have enjoyed much success describing decision-making. The framework developed by Nelson and Narens (1990) suggests separate cognitive and metacognitive processes that interact via information flow, as described earlier. For opt-out and betting tasks, confidence in a given decision would be encoded separately from the decision. It is possible this could occur in one brain region, as illustrated in Fig 2d. The hypothetical neuronal firing rates encode the decision (D_c vs. D_c') and later the metacognitive signal (MC vs. MC'). Alternatively, perhaps most closely aligned with the Nelson and Narens framework, the metacognitive signal could be encoded in a separate brain region than the cognitive signal (Fig 2e). If that were the case, one might observe little or no decision-related activity (D_c vs. D_c'). Instead, information about the decision would arrive from an external source, as for example a corollary discharge from the brain region encoding the

decision (Crapse & Sommer 2008; Sommer & Wurtz 2008). This copy of the information could be used to encode the metacognitive signal (MC vs. MC').

The mechanisms discussed are by no means exhaustive. Perhaps the most obvious alternative is to posit that metacognitive signals are not encoded by firing rates, but by a different signal. For example, neural oscillations could be used as a coding principle, affecting the correlated timing of spikes within a brain region and/or within a brain circuits across regions (Buzsaki 2006). Another possibility is that metacognitive signals are encoded by reading out some function of the variance of decision-related spiking neurons during a task (Yeung et al 2012). Finally, the worst case scenario (or most interesting scenario, depending on one's viewpoint) is that metacognition is represented along multiple dimensions of neuronal activity, including one, more than one, or all of the possibilities listed in this section. This is one reason that single neuron studies are so important. Different neurons within a brain region or between brain regions may in fact be encoding similar cognitive attributes in different ways. Methods that sample aggregate activity (e.g. fMRI, EEG) are unable to tease apart such variegated strategies for neuronal encoding. While single neuron recordings suffer from their own limitations (e.g. small sample sizes), they are exquisitely appropriate for discovering the coding mechanisms exploited by the brain for sensory, motor, or cognitive functions (Wurtz & Sommer 2006).

Single neuron studies

Metacognition has been studied only recently at the neuronal level in animals. Though many studies allow for the possibility of metacognition within their design, only three thus far have tested metacognitive processes specifically. By a metacognitive task, we mean one in which the activity of single neurons are correlated with, and therefore could be used for, monitoring a cognitive process and acting with respect to that process. A related field of study in neuroscience is so called "performance monitoring", which correlates neuronal activity with trial outcomes and rewards (Stuphorn et al 2000). Performance monitoring signals have been shown to correlate with adjustments in performance, like changes in trial response times that depend on previous trial outcomes (e.g. Pouget et al 2011). But previous performance monitoring tasks were not designed to test whether information in the signals could be use to directly affect the outcome within a concurrent trial. Here we focus on studies in which animals were encouraged to use the monitoring information functionally.

Orbitofrontal cortex Kepecs and colleagues (2008) examined neuronal correlates of confidence in rat orbitofrontal cortex (OFC), an area associated with reward, risk, and uncertainty (e.g. Hsu et al 2005; Tobler et al 2007; O'Neill & Schultz 2010). The rats performed an odor discrimination task, the goal of which was to report the majority component odor within a mixture of two odors (Fig 3a). Decisions were reported by poking their nose into one of two ports, one port for each odor. Reward was delivered after a brief delay if the decision was correct. As expected, rats made more correct decisions on easy trials, i.e. when the proportion of one odor dominated the other (Fig 3b).

OFC neuronal activity was analyzed during a time when the rats would likely experience confidence in their decisions: after the decision had been made, while the rats waited for reward delivery. Firing rates varied as a function of trial difficulty (Fig 3c) and choice accuracy (Fig 3d). In the population of recorded neurons, many (21%, or 120/563) had higher firing rates during more difficult trials, like the example neuron in Fig 3c,d. Some had the opposite pattern, higher firing rates

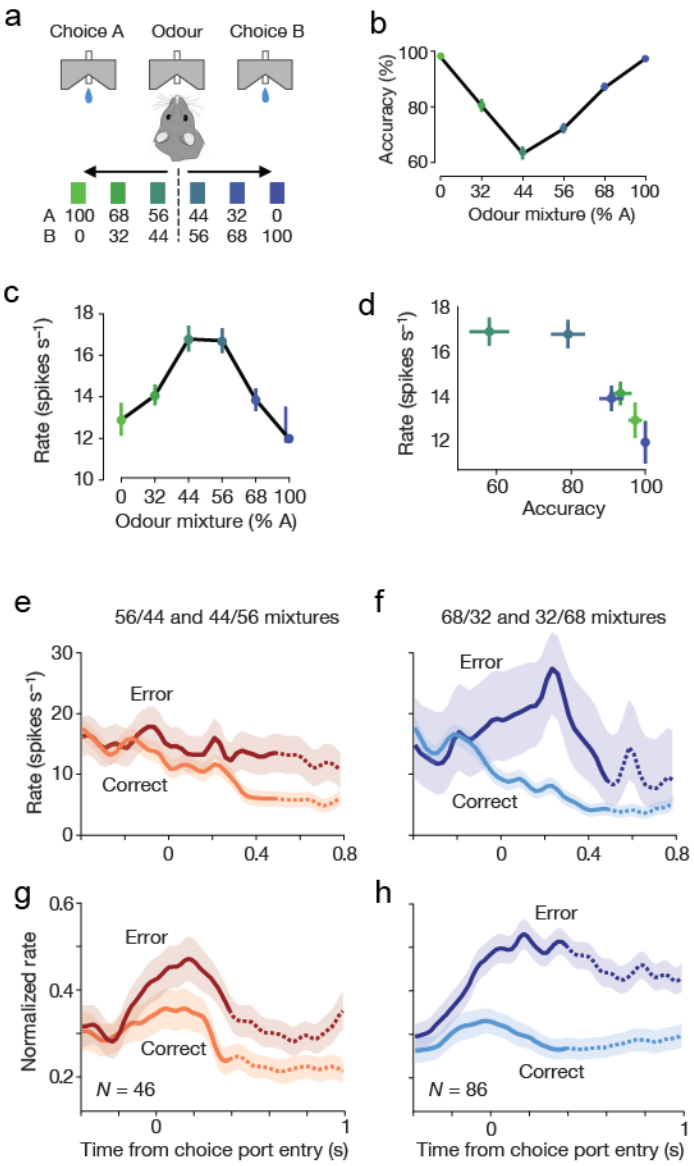

Figure 3: Confidence-related neural activity in rodent orbitofrontal cortex. (a) Rodents discriminated the majority odor component in a two-choice odor discrimination task. Decisions were reported by a nose-poke into one of two adjacent ports. (b) Rodents performed better when the one odor component dominated the other. (c) An example OFC neuron that had higher firing rates during more difficult trials. Firing rates were measured after the decision, while the rat waited for reward, (d) The same neuron had higher firing rates after less accurate decisions. (e, f) Activity of an example OFC neuron differentiated between correct decisions and errors. This difference was greater and appeared sooner on easier trials (f) than harder trials (e). (g, h) Population activity differentiated between correct decisions and errors. The patterns seen in (e, f) are conserved in a subpopulation of neurons (66/563). Adapted with permission, from Kepecs et al 2007.

during easier trials (12%, 66/563). Further, many neurons differentiated between correct and incorrect decisions within a single level of difficulty. Most had higher firing rates during incorrect than correct trials, like an example neuron during relatively difficult trials (Fig 3e) and during easier trials (Fig 3f). The neuron's firing rate began to distinguish correct from incorrect choices before the decision was made, and sustained the signal throughout reward anticipation and reward delivery. This pattern of activity was evident across the subpopulation of neurons with higher firing rates for incorrect choices (Fig 3g,h). Another subpopulation of neurons had the opposite pattern- higher firing rates during correct choices (not shown).

The OFC neuron signals could encode confidence in the decisions, and are consistent with mechanisms in Fig 2a-c. However, the rats were not required to behave in a metacognitive fashion. To assess the rats' confidence in their decisions, the authors added a manipulation to the experiment. Once a rat poked its nose into a port, a random delay was imposed before reward delivery after a correct decision (as before no reward was delivered after an incorrect decision). The rats could endure the wait and earn reward (or risk waiting longer for no reward), or they could abort the trial and immediately start the next trial. Thus the task was a hybrid between an opt-out and a betting task. The rats' behaved as if they experienced varying levels of confidence. They waited longer for reward after an easy correct trial than a difficult correct trial, and conversely after errors they aborted more often when the error was made on an easy trial than on a difficult trial. OFC neurons were not recorded during the delayed-reward trials, so we must cautiously assume the neuronal activity during the modified task was similar to that during the original task (which is not a fail-safe assumption; see Petrusic & Baranski 2003). Instead Kepecs et al (2008) offered two models, like Figures 2a and 2b, in which confidence was encoded simultaneously with the decision. The models correctly predicted the animals' behavior during the delayed-reward (opt-out version) trials and matched the pattern of OFC firing rates from recordings made during the initial discrimination task.

In sum, rat OFC neurons recorded during an odor discrimination task carried signals that could be used to make metacognitive judgments about the decisions. When the rats were subjected to a modified version of the task that encouraged metacognitive behavior, their performance was consistent with experiencing varying degrees of confidence. Models provided a link between neuronal activity during the discrimination task and performance during the metacognitive task.

Lateral intraparietal cortex Kiani and Shadlen (2009) used an opt-out task and recorded single neurons in rhesus macaque lateral intraparietal (LIP) cortex, an area implicated in visuo-spatial cognition, attention, and decision-making (Colby et al 1996; Gottlieb et al 1998; Shadlen & Newsome, 2001). Monkeys were trained to discriminate the motion direction of a visual display of randomly moving dots that had overall coherence in one direction (Fig 4a). During an initial fixation period, targets appeared in the periphery. A patch of moving dots appeared briefly then disappeared, followed by a delay, and then a cue to make a response. Trial difficulty varied with the overall motion strength of the moving dots and with the duration that the moving dots appeared. On half of the trials, a response to the stimulus was required, by making an eye movement in the same direction the dots appeared to be moving. (Fig 4a, lower panels). Reward was delivered after correct decisions. On the other half of trials, an opt-out response was offered after the moving dots stimulus disappeared, called the "sure target". If chosen it ensured a small reward (Fig 4a, upper panels).

When offered the sure target, the likelihood of choosing it increased with trial difficulty (Fig 4b). In addition, more accurate responses were made on trials when the sure target was offered but a motion stimulus target was chosen than on trials when the monkey was forced to choose a motion stimulus target (Fig 4c, closed circles are trials with sure target present, open circles are forced-choice trials). Thus the monkeys optimized reward by choosing the sure target when the probability of being correct was low, performance consistent with experiencing less confidence (more uncertainty) on those trials.

LIP activity varied as a function of choosing the sure target or one of the motion targets, illustrated by an example neuron (Fig. 4d). During forced-choice trials (left), while the monkey viewed the moving dots, the neuron's firing rate increased on trials when the target in the response field was the correct motion stimulus target (black line) relative to the incorrect target (grey line). These signals were maintained until a saccade was made to a target, similar to many previous reports of LIP neurons during the dots task (e.g. Roitman & Shadlen 2002). In trials when the sure target was offered, firing rates were again high or low when one of the motion targets was chosen (Fig 4d right panel, solid black and gray lines). When the sure target was chosen, however, the neuron's firing rates were intermediate (dashed black and gray lines). Thus varying levels of firing rates were suggested to correlate with the monkeys' confidence. This same pattern of activity was evident across the population of 70 LIP neurons (Fig 4e).

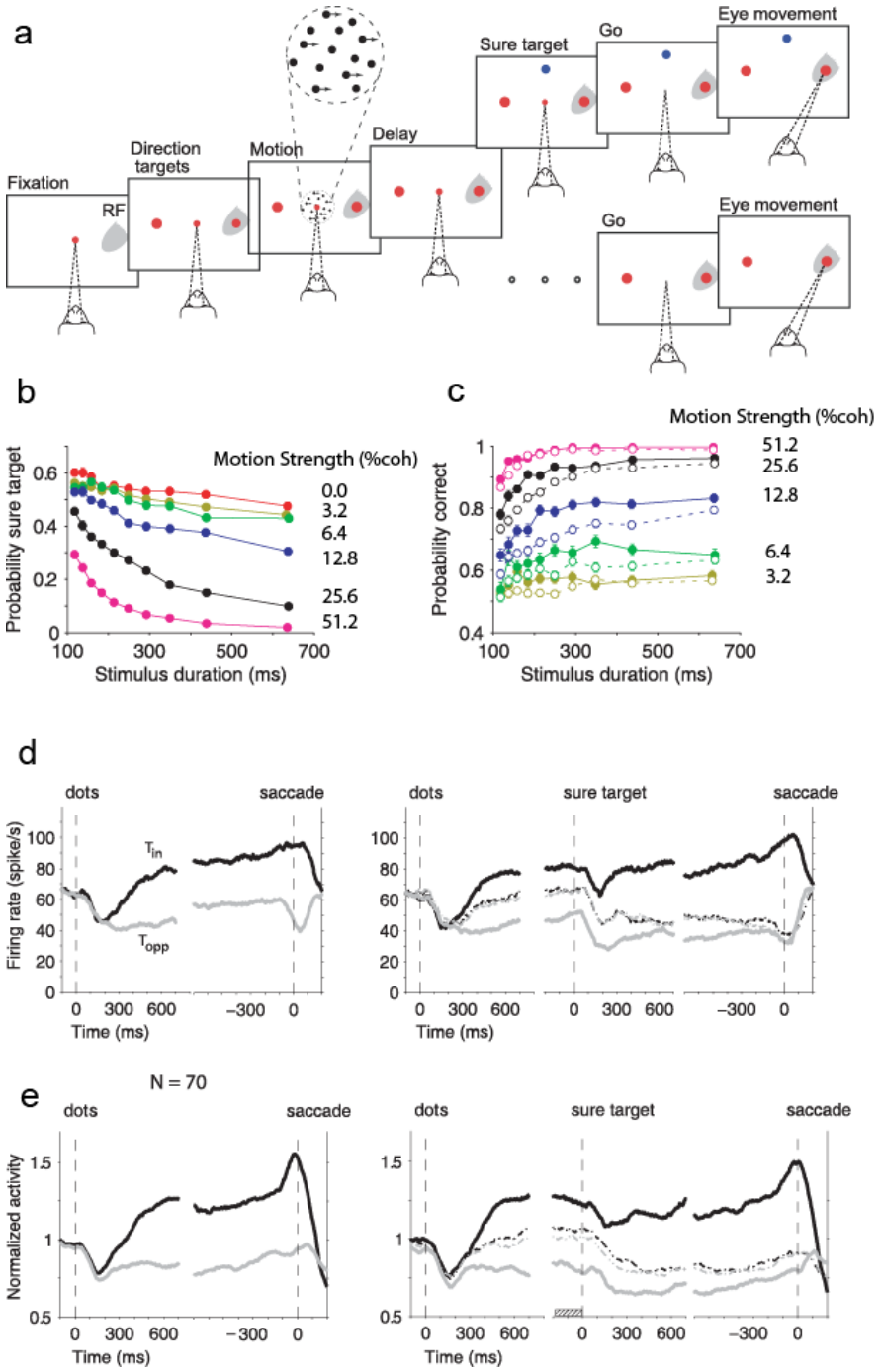

Figure 4: Confidence-related neural activity in macaque lateral intraparietal cortex (LIP). (a) Opt-out task schematic. Monkeys had to discriminate the motion direction of a random dot-motion stimulus. Decisions were reported by a saccade to one of two peripheral targets. On some trials, a third "sure target" appeared after the motion stimulus but before the animal was permitted to respond. (b) Subjects were more likely to select the sure target when stimulus presentation time was shorter and overall motion coherence was lower. (c) Subjects also performed better on trials in which the sure target was offered than on trials in which there was no sure target. (d) Activity of an example LIP neuron varied with the decision to choose one of the direction targets or the sure target. In trials without sure target (left), the cell was more active when the direction target in its receptive field was selected (black line) than when the alternate target was selected (grey line). In trials with sure target (right), activity corresponding to its selection was intermediate (dashed lines). (e) This same pattern of activity was found in a population of 70 LIP neurons. Adapted with permission from Kiani and Shadlen 2009.

Kiani and Shadlen (2009) concluded, as did Kepecs et al. (2008), that confidence was encoded along with the decision-related signal, manifested as graded levels of that signal. The authors likewise modeled their data using a sequential sampling framework. Response to the sure target depended on a dynamic threshold of neuronal activity throughout the trial, the level of which was set as a function of prior likelihoods of choosing the correct motion target. The LIP neuronal data and model therefore, like the rat OFC activity, are consistent with the mechanisms proposed in Fig 2a and 2b.

Frontal eye field, dorsolateral prefrontal cortex, and supplementary eye field Middlebrooks and Sommer (2012) carried out the most recent single neuron study of metacognition. A betting task (Fig 1b) was used, inspired by previous behavioral experiments that tested monkeys' metacognitive skills (Shields et al 2005; Kornell et al 2007).

Each trial consisted of a decision stage and a subsequent bet stage (Fig 5a). The goal of the decision stage was to detect the location of a red target square. The trial began by fixating a central spot. A red target appeared randomly at one of four possible locations, then after a varying delay white mask stimuli appeared at all four locations. A correct decision was reported by making a saccade to the location where the target appeared, and an incorrect decision was a saccade to one of the other locations. Difficulty varied as a function of the delay between the target and mask appearance, known as the stimulus onset asynchrony (SOA). Immediately after a decision was made, a new fixation spot appeared in the center of the screen to begin the bet stage. The goal of the bet stage was to make a bet regarding whether the decision was correct. Once the new fixation spot was obtained, two bet targets appeared in the periphery- a red high bet target and a green low bet target. The monkey placed a bet by making a saccade to one of the bet targets. Reward was earned based on the conjunction of decision responses and bets. A correct decision followed by a high bet (CH: correct-high) earned maximum reward, and an incorrect decision followed by a high bet (IH: incorrect-high) earned a brief timeout punishment. Low bets earned minimal juice rewards regardless of the decision (CL and IL: correct- and incorrect-low). Thus a metacognitive strategy would maximize reward: bet high after correct decisions and bet low after incorrect decisions.

There are a few noteworthy differences between the betting task and the tasks described above. First, each trial requires both a decision and a bet. The moving dots opt-out task (Fig 4a)

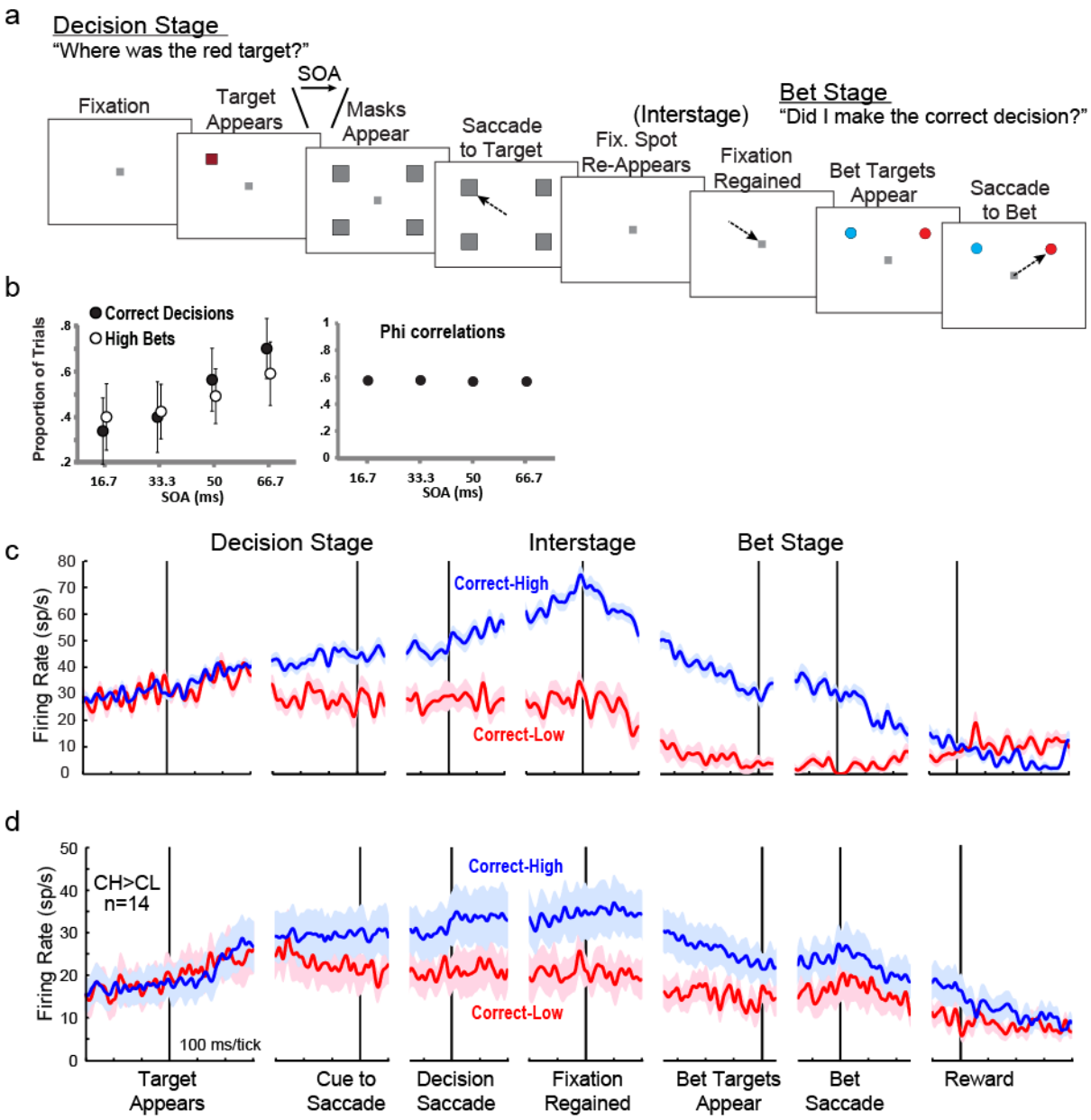

Figure 5: Confidence-related neural activity in macaque supplementary eye field (SEF). (a) Betting task schematic. During the decision stage of the task, the goal was to detect the location of a red target. The red target appeared at one of four locations, and white mask stimuli appeared at all four locations after a brief delay (stimulus onset asynchrony, SOA). Decisions were reported by a saccade to one of the four masks. After a decision, the animal regained fixation to begin the bet stage. During the bet stage a high bet target and a low bet target appeared in the periphery, and the monkey made a saccade to one of the bet targets. (b) Decision accuracy and the proportion of high bets increased as task difficulty decreased (left panel, greater SOA values correspond to easier target detection). On trial-by-trial basis, high bets were correlated with correct decisions regardless of trial difficulty (right panel, phi correlations greater than zero indicated correlated decisions and bets) (c) Activity in an example SEF neuron varied with the likelihood of choosing the high bet after correct decisions. (d) This pattern of activity was conserved in a population of 14 SEF neurons. Adapted with permission from Middlebrooks and Sommer 2012.

required a single response (a decision or an opt-out), so neuronal activity related to a decision is potentially complicated during trials in which the animal opted out. The odor discrimination task (Fig 3a) is more similar to the betting task by requiring a decision response each trial followed by either an action (abort trial to restart) or no action (wait for reward). However, by requiring a saccadic bet on each trial, trials can be compared in which identical behaviors can result in alternative outcomes, controlling for behavior as a possible explanation of neuronal activity.

Another difference between tasks involves the type of perceptual decisions required. The decision stage of the betting task required *detection* of a stimulus. In contrast, the odor discrimination and the moving dots tasks both required *discrimination* of a stimulus. The subtle difference may serve better to separate decision related signals from metacognitive signals. During discrimination tasks, perceptual evidence is thought to accumulate in neuronal activity over time until a decision is reached. This approach has provided rich contributions to understanding how decisions are made in the brain (Gold & Shadlen 2007; Smith & Ratcliff 2009). It is possible though that signals related to the evolving perceptual evidence overlap with signals related to metacognition. Although this complication is not an issue when analyzing neuronal activity well after a decision (like the OFC activity during the odor discrimination task), it may affect interpretation of signals early during the task (like the LIP activity of the moving dots task). Using a detection task, involving a brief pulse of sensory information on which to base a decision, provides separation between perceptual and metacognitive signals, thus untangling them.

The monkeys' performance during the betting task indicated they used a metacognitive strategy. During the decision stage, target detection varied as expected with trial difficulty- correct decisions increased as a function of SOAs (Fig 5b, left panel). Bets also varied with trial difficulty- high bets increase as a function of SOA. This overall pattern, the tendency to bet high on trials more likely to bet correct was expected if the animals monitored their decisions. But it also could result from a probabilistic betting strategy based solely on the difficulty of the decisions. If so, high bets on average would parallel correct decisions (and low bets would parallel incorrect decisions), but on a trial-by-trial basis high (low) bets might not follow correct (incorrect) decisions. To ensure the monkeys adopted a metacognitive betting strategy, a trial-by-trial analysis confirmed that high bets mostly followed correct decisions and low bets mostly

followed incorrect decisions, regardless of trial difficulty (Fig 5b, right panel). Thus, monkeys accurately monitored their decisions to make appropriate bets.

Neurons were recorded in three separate cortical regions: the frontal eye field (FEF), dorsolateral prefrontal cortex (PFC), and the supplementary eye field (SEF). The decision stage of the task was inspired by previous reverse masking tasks in which FEF neuron firing rates varied with monkeys' ability to detect the target (Thompson & Schall 1999, 2000). FEF is involved in oculomotor behavior (Bruce & Goldberg 1985), higher-level processes like attention (Moore & Fallah 2001), and is known to send copies of eye movement signals to other brain areas (Crapse & Sommer 2008, Sommer & Wurtz 2008). Middlebrooks and Sommer (2012) reasoned that FEF activity might also vary with monkeys' processing of the decision stage to guide a subsequent metacognitive bet. For similar reasons, neurons in PFC and SEF were recorded. PFC has been implicated in a range of high-level cognition, like working memory (Funahashi et al 1989), decision-making (Kim & Shadlen 1999), and goal-driven behavior (Tanji & Hoshi 2008). SEF, in addition to having activity related to visual processing and oculomotor behavior (Schlag & Schlag-Rey 1987; Schall 1991), has a known role in so-called performance monitoring- signals related to errors, response conflicts, and rewards (Stuphorn et al 2000). Performance monitoring signals produced during the decision stage could be used to encode an upcoming bet.

Neuronal firing rates were first analyzed with respect to decision outcomes during the task, regardless of subsequent bets. Early during the decision stage, when sensory evidence about target location might be encoded, all three cortical regions' neuronal firing rates were modulated with decision accuracy. During the planning, execution, and immediate aftermath of the saccadic response, only SEF firing rates were modulated. FEF and PFC neurons were active and task-related, but did not differentiate correct and incorrect decisions (Schall & Hanes, 1996, e.g.). In short, as expected, neuronal activity in each brain region varied with decision accuracy.

To test whether neurons in these regions were involved in monitoring decisions, neuronal firing rates were compared between the conjunctions of decision and bet outcomes- the metacognitive processes. If neurons encoded the accuracy of monitoring decisions, a prediction would be that their firing rates would be modulated between trials in which different bets were made after having made the same (correct or incorrect) decision. Trial outcomes were thus divided to compare CH vs. CL and to compare IH vs. IL trials.

Of the three cortical regions tested, SEF seemed most involved in metacognitive processing. There were neurons in SEF that differentiated CH and CL trials (15%, or 20/133) and neurons that differentiated IH and IL trials (8%, or 10/133). An example neuron that had higher firing rates for CH than CL trials is shown in Fig 5c. Firing rates for CH and CL outcomes are shown throughout the trial. The signals diverge quickly after the target appears (before the decision has been made), reach a peak difference between the decision stage and the bet stage, and maintain a difference through the betting stage. The example neuron was typical of the population that had CH firing rates greater than CL (Fig 5d). In general, SEF activity during the betting task provided more support that metacognitive processes could be encoded concomitant with and in the same brain region as cognitive processes.

Discussion

What do we know about the neuronal basis of metacognition? As attested by the studies described above, it is too early in this burgeoning field to make definitive claims about how neuronal activity translates into metacognitive behavior. Neurons in LIP, OFC, and SEF all had firing rates that varied with metacognitive behavior. There is no way to tell whether the neuronal activity was necessary for the metacognitive behavior, however, because none of the studies used causal manipulations. Microstimulation techniques and reversible inactivation or lesions of brain regions are needed to provide evidence that any region plays a causal role. A caveat to such approach however, is that it may be difficult to ascribe effects solely to metacognitive processing if the same brain regions are encoding the cognitive processes.

A major challenge facing single neuron metacognition research is the extent to which animal models of metacognition apply to human metacognition. It seems likely, based on the success of opt-out tasks, that many animals experience some measure of confidence along with the decisions they make. It also seems likely, based on betting tasks, that some animals keep track of the accuracy of their decisions, at least over short period of times. It is an open question whether these behaviors occur naturally in the environment or are a product of nurturing rudimentary metacognitive abilities by extensive laboratory training.

The relative simplicity and streamlined design of the tasks described above has advantages and disadvantages. Each task used a metacognitive component temporally yoked to the cognitive component. Notable advantages of this task design are the abilities to observe the

dynamics of neuronal activity within a single trial, and to interpret the signals within the context of the large body of knowledge in decision-making neuroscience. A disadvantage is that they do not capture the complexity we traditionally associate with human metacognitive processes, which can refer to events many years in the past or even potential events years in the future. It will be a challenge for future animal studies to tap into more complex forms of metacognition.

All three studies in this chapter reported neuronal signals consistent with an account of metacognition being encoded in near simultaneity and in the same brain region as the referent cognitive process (Fig2a-b). None reported signals that clearly support Nelson and Narens' (1990) framework, in which a metacognitive process is distinct from and monitors or controls a cognitive process. One explanation is that the limited scope of metacognitive behaviors tested, confidence and uncertainty in perceptual processes, falls short of complexity that would require distinct circuits. Though we generally refer to metacognition as if it were a single process, it is more likely to encompass multiple functions that require various brain circuits (e.g. Gigerenzer et al 1991; Juslin & Olson 1997), depending on the cognitive processes involved and the nature of the task. Thus there may be systems yet discovered that encode metacognitive processes in a way more compatible with Nelson and Narens' framework.

It should also be noted that even if metacognitive signals reported are directly available from the cognitive signals, they are not instantaneously available. Instead, most proposed mechanisms require *some* computation to read out the metacognitive signal, whether it's a comparison between two neurons' firing rates (Fig 2a), or a comparison between one neuron's firing rate and a signal representing a threshold from memory, etc. Therefore the results of the single neuron studies do not rule out separate cognitive and metacognitive systems. In Nelson and Narens' framework, information is proposed to flow between the metacognitive and cognitive processors. Information in the brain, in the form of actions potential patterns, flows at the millisecond time scale. Hence cognitive and metacognitive processes could easily overlap in time and location.

An important point to consider is that metacognitive judgments may dissociate from cognitive performance. In other words, the monitoring or control of cognitive information (metacognition) is likely based on reduced-fidelity versions of that information. This occurs in healthy individuals but is worsened in some neuropsychiatric disorders (see the chapters in the *Neuropsychiatric disorders of metacognition* section). One possible source for these

metacognitive “errors” is inaccurate transformations/computations during the readout of decision-related signals. This is consistent with an account of metacognitive judgments that depend primarily on accurate translation of cognitive signals. Another potential source of error is misinterpretation of external cues like familiarity with task stimuli, consistent with metacognitive judgments derived from sources outside the cognitive signals (Kornell 2013). These potential sources of error are not mutually exclusive, as metacognitive judgments could be affected by both factors.

A related issue is that studying high level processes at the single neuron level presents the inherent difficulty of interpreting what is actually represented in the neuronal signals. Because metacognition can involve so many other cognitive processes, one might expect multiplexed information in the neuronal firing rates. Each of the three described studies addressed this issue and ruled out some alternative accounts of the neuronal signals. Thus, cognitive functions like risk assessment and reward-related processing did not explain the neuronal activity overall. However, it is unknown how much these and other processes, like attention, might contribute from trial to trial. It is important to consider these issues moving forward.

As interesting as it is that some cortical regions were involved in metacognitive processes, it is also interesting that others were not. Specifically, neither FEF nor PFC neurons varied with metacognitive performance. The simplest interpretation is that these regions are not part of the circuit that mediates metacognition. Another possibility is that metacognition is implemented by some other coding scheme than firing rates. For example, variation in coherence of action potential timing among pools of neurons may contribute to metacognitive processes (e.g. Lisman & Jensen 2013; Nikolic et al 2013). Lastly, perhaps FEF and PFC do not contribute to the specific type of task used but may contribute when other facets of metacognition are tested.

In conclusion, the study of metacognition at the level of single neurons has been productive. With further refinement of animal-specific tasks and more detailed surveys of task-related signals across brain areas and species, single neuron data should continue to complement and inform the growing body of research on human metacognition.

REFERENCES

Beran MJ, Smith JD (2011) Information seeking by rhesus monkeys (Macaca mulatta) and capuchin monkeys (Cebus apella). Cognition 120:90-105.

Beran MJ, Smith JD, Redford JS et al (2006) Rhesus macaques (Macaca mulatta) monitor uncertainty during numerosity judgments. J Expt Pscyhol Anim Behav Proc 32:111-119.

Bruce CJ, Goldberg ME (1985) Primate frontal eye fields. I. Single neurons discharging before saccades. J Neurophysiol 53:603-635.

Buzsaki G (2006). Rhythms of the brain. Oxford, New York.

Call J (2010) Do apes know that they could be wrong? Anim Cogn 13:689-700.

Call J, Carpenter M (2001). Do chimpanzees and children know what they have seen? AnimCogn, 4:207-220.

Colby CL, Duhamel JR, Goldberg ME (1996) Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. J Neurophysiol76:2841-2852.

Crapse TB, Sommer MA (2008) Corollary discharge across the animal kingdom. Nat Rev Neurosci 9:587-600.

De Martino B, Fleming SM, Garrett N et al (2013) Confidence in value-based choice. Nat Neurosci 16:105-110.

Ding L, Gold J (2010) Caudate encodes multiple computations for perceptual decisions. J Neurosci 30:15747–15759.

Ding L, Gold J (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. Cereb Ctx, 22:1052-1067.

Flavell JH (1971) First discussant's comments: What is memory development the development of? Human development.

Flavell JH (1976) Metacognitive aspects of problem solving. In: Resnick LB (Ed) The nature of intelligence, Erlbaum, Hillsdale, NJ, p. 231-236.

Foote AL, Crystal JD (2007) Metacognition in the rat. Curr Biol 17:551-555.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331-349.

Gigerenzer G, Hoffrage U, Kleinbölting H (1991) Probabilistic mental models: A Brunswikian theory of confidence. Psychol Rev 98:506–528.

Gold JI, Shadlen MN (2007) The neural basis of decision making. Ann Rev Neurosci, 30:535–574.

Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481-484.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley, New York.

Hampton RR (2001) Rhesus monkeys know when they remember. PNAS 98:5359-5362.

Hampton RR, Zivin A, Murray EA (2004) Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. Anim Cogn 7:239-246.

Hanes DP, Schall JD (1996) Neural control of voluntary movement initiation. Science, 274:427-430.

Hart JT (1965) Memory and the feeling-of-knowing experience. J Educ Psychol 56:208-216.

Heath RA (1984) Random-walk and accumulator models of psychophysical discrimination—A critical evaluation. Perception, 13:57–65.

Hsu M, Bhatt M, Adolphs R et al (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

Juslin P, Olsson H (1997) Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. Psychological Review 104:344–366.

Kepecs A, Uchida N, Zariwala HA et al (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227-231.

Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324:759-764.

Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neuro 2:176-185.

Kornell N, Son LK, Terrace HS (2007) Transfer of metacognitive skills and hint seeking in monkeys. Psychol Sci 18:64-71.

Kornell N (in press) Where is the “meta” in metacognition? J Comp Psychol.

Krauzlis R, Dill N (2002) Neural correlates of target choice for pursuit and saccades in the primate superior colliculus. Neuron. 35:355-63.

Link SW (1992) The wave theory of difference and similarity. Erlbaum, Hillsdale, NJ.

Lisman JE, Jensen O (2013) The theta-gamma neural code. Neuron 77:1002-1016.

Merkle EC, van Zandt T (2006) An application of the Poisson race model to confidence calibration. J Exp Psychol Gen 135:391–408.

Middlebrooks PG, Sommer MA (2011) Metacognition in monkeys during an oculomotor task. J Exp Psychol Learn Mem Cogn 37:325-337.

Moore T, Fallah M (2001) Control of eye movements and spatial attention. PNAS 98:1273-1276.

Nelson TO (1996) Consciousness and metacognition. Am Psychol 51:102-116.

Nelson TO, Narens L (1990) Metamemory: A theoretical framework and new findings. Psychol Learn Motiv 26:125-141.

Nikolić D, Fries P, Singer W (2013) Gamma oscillations: precise temporal coordination without a metronome. Trends Cogn Sci 17:54-55.

O'Neill M, Schultz W (2010) Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. Neuron 68:789-800.

Petrusic WM, Baranski JV (2003) Judging confidence influences decision processing in comparative judgments. Psychon Bull Rev 10:177–183.

Pierce CS, Jastrow J (1884) On small differences in perception. Memoirs of the National Academy of Sciences 3:75-83.

Pleskac TJ, Busemeyer JR (2010) Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol Rev 116:864-901.

Pouget P, Logan GD, Palmeri TJ et al (2011) Neural basis of adaptive response time adjustment during saccade countermanding. J Neurosci 31:12604-12612.

Purcell BA, Heitz RP, Cohen JY et al (2010) Neurally constrained modeling of perceptual decision making. Psych Rev 117:1113–1143.

Ratcliff R, Cherian A, Segraves M (2003) A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. J Neurophysiol 90:1392– 1407.

Ratcliff R, Hasegawa YT, Hasegawa RP et al (2007) Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. J Neurophysiol 97:1756– 1774.

Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci, 22:9475–9489.

Schall JD (1991) Neuronal activity related to visually guided saccadic eye movements in the supplementary motor area of rhesus monkeys. J Neurophysiol 66:530-558.

Schlag J, Schlag-Rey M (1987) Evidence for a supplementary eye field. J Neurophysiol 57:179-200.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916-1936.

Shields WE, Smith JD, Washburn DA (1997) Uncertain responses by humans and rhesus monkeys (Macaca mulatta) in a psychophysical same-different task. J Exp Psychol Gen 126:147-164.

Shields WE, Smith JD, Guttmannova K et al (2005) Confidence judgments by humans and rhesus monkeys. J Gen Psychol 132:165-186.

Stuphorn V, Taylor TL, Schall JD (2000) Performance monitoring by the supplementary eye field. Nature 408:857-860.

Smith JD, Schull J, Strote J et al (1995) The uncertain response in the bottlenosed dolphin (Tursiops truncatus). J Exp Psychol Gen 124:391-408.

Smith JD, Shields WE, Allendoerfer KR et al (1998) Memory monitoring by animals and humans. J Exp Psychol Gen 127:227-250.

Smith PL, Ratcliff R (2009) An integrated theory of attention and decision making in visual signal detection. Psych Rev, 116:283–317.

Sommer MA, Wurtz RH (2008) Brain circuits for the internal monitoring of movements. Ann Rev Neurosci 31:317-338.

Son LK, Kornell N (2005) Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In Terrace HS, Metcalfe J (eds) The missing link in cognition: Origins of self-reflective consciousness, Oxford Press, New York, p. 296-320.

Suda-King C (2008) Do orangutans (Pongo pygmaeus) know when they do not remember? Anim Cogn 11:21-42.

Suda-King C, Bania AE, Stromberg EE et al (2013) Gorillas' use of the escape response in object choice memory tests. Anim Cogn 16:65-84.

Tanaka A, Funahashi S (2012) Macaque monkeys exhibit behavioral signs of metamemory in an oculomotor working memory task. Behav Brain Res 233:256-270.

Tanji J, Hoshi E (2008) Role of the lateral prefrontal cortex in executive behavioral control. Physiol Rev 88:37-57.

Thompson KG, Schall JD (1999) The detection of visual signals by macaque frontal eye field during masking. Nat Neurosci 2:283-288.

Thompson KG, Schall JD (2000) Antecedents and correlates of visual detection and awareness in macaque prefrontal cortex. Vision Res 40:1523-38.

Tobler PN, O'Doherty JP, Dolan RJ et al (2007) Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. J Neurophysiol 97:1621-1632.

Usher M, McClelland JL (2001) The time course of perceptual choice: The leaky, competing accumulator model. Psych Rev, 108:550–592.

van Zandt T (2000) ROC curves and confidence judgments in recognition memory. J Expt Psych Learn Mem Cogn 26:582–600.

van Zandt T, Maldonado-Molina MM (2004) Response reversals in recognition memory. J Expt Psych Learn Mem Cogn 30:1147–1166.

Vickers, D. (1979) Decision processes in visual perception. Academic Press, New York.

Wurtz RH, Sommer MA (2006) Single neurons and primate behavior. In Senior C, Russell T, Gazzaniga M (eds), Methods in Mind. MIT, Cambridge p. 123-139.

Yeung N, Summerfield C (2012) Metacognition in human decision-making: confidence and error monitoring. Philos Trans R Soc Lond B Biol Sci 367:1310-1321.

FIGURE LEGENDS

Figure 1: Metacognitive monitoring tasks. (a) Schematic of opt-out task paradigms. Opt-out tasks generally involve a two-choice perceptual discrimination. On some proportion of trials, a third opt-out target appears. Selection of the opt-out target results in a small but ensured reward. Participants utilizing a metacognitive strategy should select the opt-out target more often on more difficult trials, and make more accurate responses on trials the opt-out is offered. (b) Schematic of betting task paradigms. Betting tasks generally involve choice stage followed by a betting stage. Selection of the high bet target results in a large reward after a correct response, and no reward after an incorrect response. Selection of the low bet target results in a small but ensured reward. Participants using a metacognitive strategy should select the high bet target more often after correct decisions. Unlike opt-out tasks, betting tasks require a primary task decision on every trial.

Figure 2: Possible mechanisms of metacognition. (a) Comparison of decision-related activity between response alternatives. In this model, confidence (MC) is a function of the difference between neural activity for the chosen (D_c) and unchosen (D_u) responses. (b) Comparison of decision-related activity to an independent threshold. In this model, confidence is a function of the difference between neural activity for the chosen response and the threshold. It is not dependent on activity for the unchosen response. (c) Sequential coding of decision and confidence. In this model, evidence continues to accumulate after the decision is made (D_c or D_c') to subsequently produce a confidence response (MC or MC', respectively). (d, e) Distinct coding of decision and confidence. In these models, confidence is encoded separately from the decision. Confidence can be encoded by the same neurons involved in the decision (d) or by distinct neurons (e).

Figure 3: Confidence-related neural activity in rodent orbitofrontal cortex. (a) Rodents discriminated the majority odor component in a two-choice odor discrimination task. Decisions were reported by a nose-poke into one of two adjacent ports. (b) Rodents performed better when the one odor component dominated the other. (c) An example OFC neuron that had higher firing rates during more difficult trials. Firing rates were measured after the decision, while the rat waited for reward, (d) The same neuron had higher firing rates after less accurate decisions. (e, f)

Activity of an example OFC neuron differentiated between correct decisions and errors. This difference was greater and appeared sooner on easier trials (f) than harder trials (e). (g, h) Population activity differentiated between correct decisions and errors. The patterns seen in (e, f) are conserved in a subpopulation of neurons (66/563). Adapted with permission from Kepecs et al (2007).

Figure 4: Confidence-related neural activity in macaque lateral intraparietal cortex (LIP). (a) Opt-out task schematic. Monkeys had to discriminate the motion direction of a random dot-motion stimulus. Decisions were reported by a saccade to one of two peripheral targets. On some trials, a third "sure target" appeared after the motion stimulus but before the animal was permitted to respond. (b) Subjects were more likely to select the sure target when stimulus presentation time was shorter and overall motion coherence was lower. (c) Subjects also performed better on trials in which the sure target was offered than on trials in which there was no sure target. (d) Activity of an example LIP neuron varied with the decision to choose one of the direction targets or the sure target. In trials without sure target (left), the cell was more active when the direction target in its receptive field was selected (black line) than when the alternate target was selected (grey line). In trials with sure target (right), activity corresponding to its selection was intermediate (dashed lines). (e) This same pattern of activity was found in a population of 70 LIP neurons. Adapted with permission from Kiani and Shadlen (2009).

Figure 5: Confidence-related neural activity in macaque supplementary eye field (SEF). (a) Betting task schematic. During the decision stage of the task, the goal was to detect the location of a red target. The red target appeared at one of four locations, and white mask stimuli appeared at all four locations after a brief delay (stimulus onset asynchrony, SOA). Decisions were reported by a saccade to one of the four masks. After a decision, the animal regained fixation to begin the bet stage. During the bet stage a high bet target and a low bet target appeared in the periphery, and the monkey made a saccade to one of the bet targets. (b) Decision accuracy and the proportion of high bets increased as task difficulty decreased (left panel, greater SOA values correspond to easier target detection). On trial-by-trial basis, high bets were correlated with correct decisions regardless of trial difficulty (right panel, phi correlations greater than zero indicated correlated decisions and bets) (c) Activity in an example SEF neuron varied with the

likelihood of choosing the high bet after correct decisions. (d) This pattern of activity was conserved in a population of 14 SEF neurons. Adapted with permission from Middlebrooks and Sommer (2012).