



# HHS Public Access

Author manuscript

*Behav Neurosci*. Author manuscript; available in PMC 2022 April 01.

Published in final edited form as:

*Behav Neurosci*. 2021 April ; 135(2): 129–137. doi:10.1037/bne0000412.

## The ever-changing OFC landscape: what neural signals in OFC can tell us about inhibitory control

Adam T. Brockett<sup>1,2,\*</sup>, Matthew R. Roesch<sup>1,2</sup>

<sup>1</sup>Department of Psychology, University of Maryland, College Park, MD 20742

<sup>2</sup>Program in Neuroscience and Cognitive Science, University of Maryland, College Park, MD 20742

### Abstract

Despite decades of research on OFC function, the exact function(s) of OFC remain elusive. In recent years, one of the earliest hypotheses about OFC function, namely its involvement in inhibitory control, has drifted to the periphery of the functional OFC landscape in favor of theories suggesting a role for OFC in the representation of task or state space. The reasons for this drift are valid, owing in part to the development of more sensitive behavioral approaches, a clear emphasis on cross-species and cross-method comparisons, as well as the elegant integration of reinforcement learning theories. However, recent evidence recording from OFC during the performance of traditional inhibitory control tasks has found new evidence supporting a role for OFC in inhibitory control. While the extent to which these findings can be integrated into existing frameworks is in its infancy, this review seeks to highlight these findings with the goal of providing new insights into function of OFC.

### Keywords

orbitofrontal cortex; inhibitory control; cognitive control; stop-signal

### Introduction

Decision-making is a fundamental component of human behavior. Virtually, everything we do from the time at which we decide to wake up until the time at which we decide to go to bed is a decision. What to eat, what to wear, whether to go to work, what to do at work, who to talk to, are just a small sampling of the types of decisions we make on a near daily basis. Naturally, given the frequency with which we make decisions, research probing the neural basis of decision-making has emerged as, and remains, a central feature in the research landscape of modern cognitive neuroscience.

Within this landscape, much focus has been directed towards the frontal areas of the brain which are thought to provide the computational infrastructure to support, among other processes, effective decision-making and goal-directed behavior. Numerous broad-reaching,

---

\*Correspondence: Adam Brockett (brockett@umd.edu).

Conflict of Interest: The authors declare no biomedical financial interests or potential conflicts of interest.

multi-brain region encompassing theories, as well as mountains of single brain region specific theories, have been proposed with the hopes of outlining a cogent experimental trajectory to better elucidate how the frontal brain regions make decisions.

The orbitofrontal cortex (OFC), a frontal brain region that sits directly above the eyes, is in many ways a paragon for the type of academic theorizing, testing, and re-theorizing that research examining the function of frontal brain areas has undergone, and currently undergoes. From its humble beginnings as a brain region important for inhibitory control to its current, albeit debated, place as a brain area that represents a multidimensional map of task space -- that also sometimes moonlights as a center for emotional regulation -- the OFC personifies the difficulty in which pinpointing an exact function to a brain region creates. The academic difficulty of this undertaking is compounded by the fact that with each new reinterpretation of the functional OFC landscape, the now questionable landscaping decisions of the previous interpretation, clutter the new vision trying to emerge.

This review seeks to revisit some the earliest questionable landscaping decisions in OFC theory, namely inhibitory control, in the hopes of providing new insight into the involvement of OFC in inhibitory control processes. We also hope to forge a clearer path towards integrating theories of inhibitory control into more current interpretations of OFC function. Importantly, before we begin, we concede that the original theories on inhibitory control and OFC function were too simplistic, and we appreciate more modern interpretations of this literature, which have convincingly demonstrated how these theories were misguided. However, like many, we also believe that recent evidence has provided enough clues to envision a role for OFC in inhibitory control that can exist within the framework of more prominent theories, and that is what we seek to convey here.

## The History of the OFC and Inhibitory Control

Some of the earliest theories of OFC function were based on observations of patients that had experienced frontal lobe damage. The classic study of Phineas Gage was one of the earliest observations to contribute to OFC theory. As is well documented, a tamping rod blasted through Gage's skull cauterizing the wound as it passed, but also effectively severing or severely damaging connections between frontal brain areas, particularly the OFC, and the rest of his brain (Harlow, 1993; Horn et al., 2012). After recovering, Gage was described as subdued, although generally pleasant, but was prone to intense, seemingly uncontrollable fits of rage, a departure from descriptions of his pre-injury self, and a condition that would later be described as a kind of behavioral disinhibition (Harlow, 1993). Behavioral disinhibition, or intense shifts in behavior from jovial to angry, would go on to be observed as a relatively common side effect in patients that underwent frontal lobectomy and OFC leucotomies (Bechara et al., 2000; Blair, 2010; Davidson et al., 2000; Reitman, 1946; Saver & Damasio, 1991). In some cases changes in behavior were so severe that researchers began classifying symptoms as a form of 'acquired sociopathy' (Bechara et al., 2000; Blair, 2010; Davidson et al., 2000; Rudebeck & Rich, 2018; Saver & Damasio, 1991).

Similarly, as psychologists and neuroscientists alike began looking for signs of disinhibition associated with OFC damage in patient, non-human primate, and rodent models, a wealth of

data emerged suggesting that the OFC was critical for reversal learning (Bechara et al., 1997; Chudasama & Robbins, 2003; Dias et al., 1996; Fellows & Farah, 2003; Hornak et al., 2004; Izquierdo et al., 2004; Izquierdo & Jentsch, 2012; Jones & Mishkin, 1972; Kim & Ragozzino, 2005; McAlonan & Brown, 2003; Rolls et al., 1994; Geoffrey Schoenbaum et al., 2002; Geoffrey Schoenbaum, Setlow, Nugent, et al., 2003; Walton et al., 2010). At its core, reversal learning requires a subject to map a response (e.g., press the right lever) onto the acquisition of some kind of reward. After learning the initial association, subjects are then asked to inhibit the previously learned response in favor of responding in the opposite manner (i.e., press the left lever) in order to obtain reward. Subjects with OFC lesions were unable to perform the reversal, despite in most cases being able to learn the initial discrimination. This was seen as evidence for the functional role of OFC in inhibitory control, namely in the inhibition of inappropriate or maladaptive responses.

The ease with which reversal learning could be operationalized and tested across species in combination with the observed behavioral deficits and the clinical behavioral disinhibition literature, led many to propose a role for OFC in inhibitory control, and for a time, established OFC at the center of inhibitory control.

## The Argument Against OFC and Inhibitory Control

As research into the role of OFC in inhibitory control ramped up, several pivotal findings during this time ultimately led to this hypothesis' demise (Rudebeck & Rich, 2018). The first of these observations was that humans and animal models with OFC damage were always able to learn the initial contingency with relatively little, if any, difficulty. If OFC was critical for inhibiting a behavioral response, then presumably when subjects first learn a response, some degree of inhibition for the other, non-task relevant strategies or responses would be needed. Accordingly, it stands to reason that deficits in the acquisition of an association would also be evident in a subject with OFC damage. However, when the OFC of rats was lesioned in animals performing a classic inhibitory control task, the GO/ NOGO task, lesioned rats were able to acquire NOGO responses at the same rate as controls (Geoffrey Schoenbaum et al., 2002). The critical difference between controls and lesioned animals was only observed when response contingencies changed (Geoffrey Schoenbaum et al., 2002; Geoffrey Schoenbaum, Setlow, Nugent, et al., 2003). These findings were notably echoed by other work in rats (Riceberg & Shapiro, 2012) and non-human primates (Walton et al., 2010), and helped usher in a shift in landscape of OFC theory.

Data using a modified version of a reversal task where instead of subjects alternating between two associations, subjects were asked to choose between three possible associations further complicated this theory (Walton et al., 2010). By adding a third association, researchers observed that OFC lesions impaired monkeys from obtaining the highest valued reward, but not because of inflexibility/ perseveration (Walton et al., 2010). Instead, deficits in reward acquisition were explained by a failure to properly map reward value onto the appropriate response. In other words, monkeys were not fixated on the previously rewarded behavioral response in the way the inhibitory control hypothesis might predict, instead monkeys switched often but were unable to integrate the changing reward information with the appropriate response (Walton et al., 2010).

The so-called “nail in the coffin” for OFC’s role in inhibitory control came when famous OFC lesion studies in non-human primates were re-examined with different techniques (Kazama & Bachevalier, 2009; Rudebeck et al., 2013; Rudebeck & Murray, 2011). Previous work in non-human primates had established a role for OFC in reversal learning through the use of aspiration lesions, which while small, damaged both the OFC and the fibers of passage that passed through OFC. Results with excitotoxic lesions, suggested that this once classic effect of cognitive neuroscience may not be due to OFC damage per se, but rather due to damage of the fibers that pass through OFC. Indeed, in non-human primates it was shown that excitotoxic lesions produced minimal impairment of reversal learning, and that deficits in reversal learning could only be observed once a small anterior portion of OFC was aspirated (Rudebeck et al., 2013).

Collectively, these three streams of research helped to reshape the OFC landscape, shifting it from one dominated by a focus on the role of OFC in inhibitory control to one that focused more on OFC’s role in mapping value onto actions and stimuli. This new focus on value/economic decision making would usher in arguably the most productive and highly researched period in OFC function.

## Current Theories of OFC Function

### Outcome Expectancy

With the realization that impairment on reversal learning tasks were due, in part, to deficits in a kind of reward credit assignment, rather than a pure impairment of inhibitory control mechanisms, an aptly named, ‘new perspective’ on the role of the OFC in adaptive behavior began to emerge (Geoffrey Schoenbaum et al., 2009). While the history of this theory is beyond the scope of this review, this new configuring of the OFC landscape focused on reward expectancies instead of the inhibition of inappropriate responses. Specifically, it charged OFC with the role of mapping a unique value onto a specific set of features or expectancies an animal might have in a given environment (Schoenbaum et al., 2009).

This perspective was supported by evidence of strong anticipatory firing in OFC, that while not necessarily unique to OFC, was robust, and often occurred before being detected in other brain regions that showed similar response profiles (Gottfried et al., 2003; Hikosaka & Watanabe, 2004; G. Schoenbaum et al., 1998; Geoffrey Schoenbaum, Setlow, Saddoris, et al., 2003; Takahashi et al., 2011; Tremblay & Schultz, 1999; Wallis & Miller, 2003). Critically, it was also shown that features of prediction error encoding dopamine neurons was dependent on OFC (Takahashi et al., 2011). Specifically, researchers showed that without OFC, dopaminergic error signals failed to reflect internal information about the impending response that distinguished externally similar states leading to differently valued future reward (Takahashi et al., 2011). These results suggest that OFC maintains a kind of model-based representation or simulation of what the animal or subject should be expecting from performing a specific action or set of actions.

Behaviorally, data from Pavlovian reinforcer devaluation studies (Gallagher et al., 1999; Izquierdo et al., 2004; Machado & Bachevalier, 2007; Pickens et al., 2003, 2005; Geoffrey Schoenbaum, Setlow, Saddoris, et al., 2003), a Pavlovian Instrumental Transfer study

(Ostlund & Balleine, 2007), delayed-discounting studies (Kheramin et al., 2003; Mobini et al., 2002; Winstanley et al., 2004), as well as more recent chemogenetic work in mice (Baltz et al., 2018), and single neuron ablation work in the OFC of rats (Groman et al., 2019), impairment of OFC disrupted a subject's ability to adapt behavioral appropriately based on changing reward contingencies.

### Task/ State Space

This evidence supported a role for OFC as a kind of critic that evaluated actions and updated the value of said actions in a manner which optimized future decisions. To do so, OFC would need to maintain a kind of map to the subject's current position in the task in order to generate representations of what is expected (Stalnaker et al., 2015; Wilson et al., 2014). The theory that the OFC acts as a kind of cognitive map accounts for many of the previous visions of OFC function, including its role in reversal learning, delayed alternation, extinction, devaluation, post-extinction predictions, and prediction error (Wilson et al., 2014). Critically, this theory makes the prediction that the functional importance of OFC can be detected when changes in task contingencies are either covert or only partially signaled to the subject (Wilson et al., 2014). In rodents, medial OFC is only necessary when reversals are partially observable, but not when completely observable (Bradfield et al., 2015). Similarly in humans, covert changes in task space can be decoded from neural activity in OFC (Schuck et al., 2016). Moreover, this theoretical rendering of OFC function accounts for discrepancies in lesion studies, which collectively suggest that while slower, subjects without an OFC or with severe OFC damage exhibit only subtle differences in learning ability from those with healthy OFC. In this view, differences between healthy and OFC damaged subjects stem from the compromised ability of subjects to detect subtle differences between task spaces (Wilson et al., 2014), and fits with preexisting ideas concerning a convergence of reward and aversive information in OFC (Morrison & Salzman, 2009).

### Revisiting Inhibitory Control

With an expansive and experimentally grounded theory of OFC function in place, and the focus of the OFC landscape decidedly shifted away from initial ideas of inhibitory control, it may seem surprising to write a review about OFC and inhibitory control. However, several papers and theories have presented conflicting accounts of OFC function as they relate to stopping behaviors (Aron et al., 2014). For instance, OFC lesions in rats performing a 5-choice serial reaction time task have been shown to increase perseverative responding (Chudasama & Robbins, 2003) and work using a stop-signal task has shown that OFC lesions increase stop-signal reaction time (SSRT), a measure of the time needed to inhibit a prepotent response (Eagle et al., 2008). Moreover, administration of the attention-deficit hyperactivity drug (ADHD), atomoxetine, to OFC has been shown to increase inhibitory control, and stabilize SSRTs on a stop-signal task (Bari et al., 2011). Collectively, these findings do not provide conclusive evidence linking OFC to inhibitory control, but it is worth noting that they also do not provide evidence to the contrary. Instead, these findings hint to the possible utility of classical inhibitory control tasks, such as the stop-signal task, in the context of behavioral neurophysiology to further assess the role of OFC in inhibitory

control, as well as to help answer the question; what is OFC signaling during response inhibition?

Importantly, while there is good reason to suspect that the primary role of OFC is not inhibitory control from single neuron recording studies, we must also recognize that much of the work that supports this view was conducted using non-traditional inhibitory control tasks; that is, in tasks designed to test neural correlates related to other functions, such as reward processing or reversal of stimulus- or response-outcome contingencies. The primary goal of stop-signal tasks is to test how automatic or habitual responding are inhibited outside the context of reward manipulation (Verbruggen et al., 2019). The fact that studies have shown that OFC lesions and pharmacological manipulations of OFC changes SSRTs during stop-signal performance suggest that it must be signaling something important for task performance (Eagle et al., 2008). Despite this relatively straightforward prediction, there have only been a small number of behavioral recording studies examining single unit firing in the OFC of animals performing analogs of tasks that specifically probe inhibitory function.

This issue has recently been addressed in monkey performing a classic stop-signal, also described as countermanding (Balasubramani et al., 2020). In this task, monkeys were trained to fixate on a white circle for 300 ms before a peripheral cue, presented to either the right or left of the center point was presented (Fig. 1A). On 67% of trials (GO trials) monkeys need to make a saccade toward the peripheral cue in order to receive reward. On the remaining 33% of trials (STOP trials), a second cue, a grey square, was presented at the center point, after the presentation of the first cue. Presentation of the STOP cue was always delayed relative to the GO cue, and the delay was titrated for each animal in order to maintain accuracy at approximately 50%. This delay is commonly referred to as a stop signal delay (SSD) and the longer it is, the more difficult it is to inhibit responding. Using this task, researchers found that the activity of some OFC neurons correlated with successful task completion (Fig. 1B) and that ensemble decoding revealed that firing in OFC significantly distinguished firing on successful versus failed inhibition trials (Balasubramani et al., 2020). Interestingly, a separate analysis revealed that inhibition signals were orthogonal to value encoding, although signaled by the same neurons, suggesting that inhibition and value information are represented discretely (Balasubramani et al., 2020). Thus, the results demonstrate that OFC does contribute to inhibitory control via task-related signals that are outside the realm of reward or value encoding.

Consistent with these results, a study done in humans, also performing a stop-signal task, showed that patients with OFC lesions exhibited diminished N2 and error-related negativity suggesting abnormalities in action monitoring (Solbakk et al., 2014). Moreover, patients exhibited enhanced P3 error positivity and post-error beta response, which suggests a role for OFC in outcome evaluation (Solbakk et al., 2014). Thus, human work also suggests that OFC manages functions related to both reward and inhibitory control.

In addition to these studies, we have recently contributed to this literature by recording from OFC in rats performing a novel variant of stop-change tasks, which are similar to stop-signal tasks but require subjects to not only inhibit behavior but to also redirect it in the opposite



direction (Bryden & Roesch, 2015). During performance of this task, on 80% of trials (GO trials), rats were trained to respond quickly to a directional light cue that indicated one of two fluid wells that the rat would need to move to in order to get a small liquid sucrose reward (Bryden & Roesch, 2015). On the remaining 20% of the trials, STOP-change trials, rats received the same first directional cue, but within 100ms of receiving the first cue, a second cue in the opposite direction was illuminated instructing rats to inhibit their initial response, and redirect their response to the fluid well signaled by the second cue. In this task the need to resolve ‘conflict’ between two actions requires inhibition of one (i.e., the direction signaled by the first cue light) and promotion of the other (i.e., the direction signaled by the second cue light).

At the neural level we see the resolution of conflicted response signals in DMS emerge when rats correctly inhibit and redirect behavior on STOP-change trials (Bryden et al., 2012); specifically we find that action plans in DMS are slow to signal the correct direction and that the overall the strength of the directional signals are attenuated on STOP trials. These changes in neural firing correlate well with successful stopping and slower reaction times observed on STOP-change trials (Bryden et al., 2012). Remarkably, when we recorded in OFC in the same task, we found directional signals in OFC were not attenuated on STOP-change trials. Moreover, directional signals on STOP trials were strongest on trials where rats experienced the greatest difficulty inhibiting behavior (Bryden & Roesch, 2015). That is, OFC was the most engaged when the need to inhibit behavior was at its highest. Loss of this function would fit well with OFC lesion studies showing longer SSRTs in rats (Eagle et al., 2008), suggesting that movement signals in downstream areas, like DMS, are slower to resolve without the necessary boost in encoding from OFC.

In this view, OFC is not merely activated during inhibition, but contributes to the amplification of directional signals when the need for inhibition and adaptive behavior arises. This hypothesis is further supported by the observation that firing in OFC was strongly modulated during ‘conflict adaptation’ in rats performing our task. Conflict adaptation is the well-known behavioral phenomena whereby subjects perform better after difficult or high conflict responses by slowing down and being more engaged in the task at hand. In our task, rats respond more slowly to the first cue, and more quickly and easily resolve conflict on STOP trials that follow STOP trials or errant responses. It is during these trials, that the directional signals in OFC are the strongest. This is illustrated in Figure 2B and C, which plots the average firing rate of OFC neurons during GO trials and sS (STOP following STOP trials) trials. The strength of the directional signals (i.e., the difference between thick and thin lines) was the strongest during sS trials via a mechanism by which increases and decreases in firing were observed for movements to be made into and away from each neuron’s response field, effectively boosting or amplifying the differentiation of the two competing actions. Such proactive amplification of response direction encoding likely contributes to better adaptive behavior in the event the animal experiences conflict again (Bryden & Roesch, 2015). Further, we see that directional signals in response to the first cue are slower to emerge and do not persist as long, thus inhibiting the prepotent drive to respond to the first cue (Fig. 2D). In many ways these results support previous work in humans suggesting that OFC is involved in the preparation of enacting inhibitory control (Chikazoe et al., 2009).

Further support for OFC in this role comes from a lesion study in non-human primates performing a variant of the Wisconsin Card Sorting Task (Mansouri et al., 2014). Here, conflict was manipulated by whether the two rules (i.e., pay attention to the shape of the stimuli or the color) required the same or different response. Shifts in rewarded rule were unannounced to the monkey, and occurred after the monkey reach 85% accuracy with the current rule. Researchers found that OFC lesions increased the number of trials required to reach criterion, suggesting that monkeys were less likely to rule-shift. Moreover, conflict adaptation (i.e., the slowing of behavior following a high conflict trial to increase the likelihood of correctly responding on the next trial) was absent in lesioned monkeys. Further, recordings from OFC during task performance showed firing that was indeed sensitive to the level of conflict during decision-making periods.

It is important to note that none of the results presented here argues against the role of OFC in maintaining a cognitive map like representation of task space *per se*, however, these results provide clear evidence that neural activity in OFC is modulated by conflict, suggesting that these results merit further exploration. Specifically, these results provide more insight into the means by which inhibitory control may be implemented in OFC, while also supports much of the previous work implicating OFC in these and other processes (Chikazoe et al., 2009; Dias et al., 1996; Eagle et al., 2008; Majid et al., 2013; Roberts & Wallis, 2000).

## Inhibitory Control in Task Space

Given that neurons in OFC are modulated by conflict, or the need for inhibitory control, several possibilities for the utilization of this information emerge. With the perspective of the cognitive map theory in mind, it might be thought that inhibitory control signals in OFC may be more indicative of the subjective value of stopping during a stop-signal task. This idea merges current thoughts on OFC and its importance in value encoding with stopping behavior (Padoa-Schioppa, 2011), however, does not align, with recording data suggesting that inhibitory control signals are orthogonal to value signals (Balasubramani et al., 2020). Again, from this perspective, a more likely scenario might be that inhibitory control signals may be important for shifting a subject from one task space to the next or even shifting the location of the subject within a given task space. Given the observation that deficits in behavioral performance due to OFC lesions emerge when cues of task environment are only partially observable (Bradfield et al., 2015), this suggests that OFC is vital for discriminating. This discrimination is likely based on a variety of sensory and motivational factors, which help OFC to process which state a subject is in. In this view, in order for a subject to successfully navigate task space, inhibitory control mechanisms need to be in place to adapt behavior when task contingencies change. This fits with the assertion that OFC is needed to dissociate between two perceptually similar actions during conflict. Unfortunately, data testing this hypothesis directly has yet to emerge. Functional evidence linking OFC function to the failure to implement inhibitory control is needed to show that inhibitory control signals in OFC direct this kind of task space switching/ discrimination.



## A Behavioral Economic Perspective

Alternatively, a computationally distinct, but emerging hypothesis maintains that inhibitory control is just one part of a suite of signals that OFC processes as it transforms a stimulus into an action (Balasubramani et al., 2020; Yoo & Hayden, 2018). While not necessarily specific to OFC, the fact that value and inhibitory control information are orthogonal, yet multiplexed, by the same neurons suggests that this information can contribute to either helping define a task space or more directly to deciding whether to proceed with an intended action (Balasubramani et al., 2020).

Importantly, this hypothesis is not mutually exclusive with the task space hypothesis, and concedes that OFC may play a primary role in task space (Yoo & Hayden, 2018). However, unlike the cognitive map theory of OFC function, which sees inhibitory control as a potential mechanism to switch task spaces, this more dendriform view of brain functioning raises the possibility that these inhibitory control signals may be part of a larger control network tasked with making decisions about which action to perform. In other words, this more behavioral economics perspective, might suggest that OFC is just one part of a larger inhibitory control circuit, charged with biasing behavior in a particular direction, and that works in concert with several other frontal areas in parallel. This fits with the findings of Bryden and Roesch (Bryden & Roesch, 2015) as well, in that almost all firing in OFC was directionally specific, and once again, that greater OFC activity was observed on trials requiring the greatest control.

This perspective is also partially supported by a recent finding in rats performing a response preparation task which showed that optogenetic inhibition of OFC was associated with impairments in reactive responding, specifically with regards to motor planning and execution (Hardung et al., 2017). Moreover, the results of this study showed that guiding of motor planning occurred in parallel across several frontal regions, supporting the idea of dendriform processing of information (Hardung et al., 2017).

Lastly, it is important to note that this emerging idea is still very much just that, emerging, at least with respect to experimental evidence. However, in principle, this idea critically expands OFC's involvement in inhibitory control in a way that fits with recording data collected from stop-signal tasks, while also accounting for, ideas of value and expected outcomes (Padoa-Schioppa, 2011; Yoo & Hayden, 2018).

## Conclusions

Despite the constant reshaping of the OFC functional landscape, inhibitory control has managed to remain a constant, albeit at times unsightly, member of this landscape. The recent emergence of recording findings from OFC in animals performing classic inhibitory control tasks is starting to shed new light on the role of OFC in control. Future work must pay close attention to the ways in which OFC helps to form actions, and to dissociate whether its computations serve a more value-based or a motor-planning one. Particularly, work exploring how decision-making works with measures of self-control may usher in a new vision for OFC function that once again places control at the forefront (Balasubramani

et al., 2020; Yoo & Hayden, 2018). The functional OFC landscape is ever changing, but we believe understanding how inhibitory signals fit within this landscape is crucial to developing the most accurate understanding of OFC function.

## Acknowledgments:

This work was supported by the following grants: NIMH: MH1117836 to ATB and NIDA: DA031695 to MRR

## References

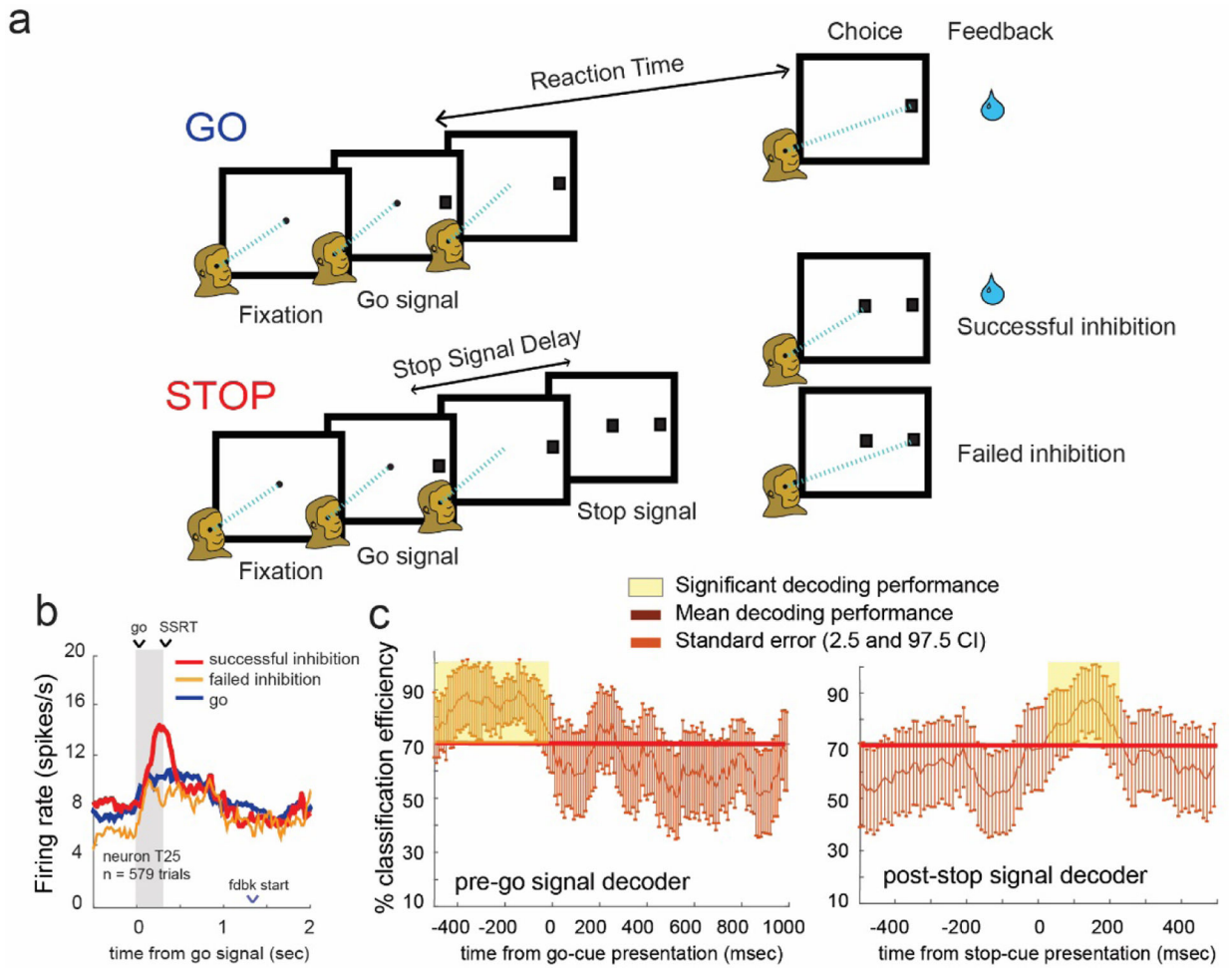
- Aron AR, Robbins TW, & Poldrack RA (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, 18(4), 177–185. 10.1016/j.tics.2013.12.003 [PubMed: 24440116]
- Balasubramani PP, Pesce MC, & Hayden BY (2020). Activity in orbitofrontal neuronal ensembles reflects inhibitory control. *European Journal of Neuroscience*, 51(10), 2033–2051. 10.1111/ejn.14638
- Baltz ET, Yalcinbas EA, Renteria R, & Gremel CM (2018). Orbital frontal cortex updates state-induced value change for decision-making. *ELife*, 7, e35988. 10.7554/eLife.35988 [PubMed: 29897332]
- Bari A, Mar AC, Theobald DE, Elands SA, Oganya KCNA, Eagle DM, & Robbins TW (2011). Prefrontal and Monoaminergic Contributions to Stop-Signal Task Performance in Rats. *Journal of Neuroscience*, 31(25), 9254–9263. 10.1523/JNEUROSCI.1543-11.2011 [PubMed: 21697375]
- Bechara A, Damasio H, & Damasio AR (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex (New York, N.Y.)*, 10(3), 295–307. 10.1093/cercor/10.3.295
- Bechara A, Damasio H, Tranel D, & Damasio AR (1997). Deciding advantageously before knowing the advantageous strategy. *Science (New York, N.Y.)*, 275(5304), 1293–1295. 10.1126/science.275.5304.1293
- Blair RJR (2010). Psychopathy, frustration, and reactive aggression: The role of ventromedial prefrontal cortex. *British Journal of Psychology*, 101(3), 383–399. 10.1348/000712609X418480 [PubMed: 19321035]
- Bradfield LA, Dezfouli A, van Holstein M, Chieng B, & Balleine BW (2015). Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable Task Situations. *Neuron*, 88(6), 1268–1280. 10.1016/j.neuron.2015.10.044 [PubMed: 26627312]
- Bryden DW, Burton AC, Kashtelyan V, Barnett BR, & Roesch MR (2012). Response inhibition signals and miscoding of direction in dorsomedial striatum. *Frontiers in Integrative Neuroscience*, 6, 69. 10.3389/fnint.2012.00069 [PubMed: 22973206]
- Bryden DW, & Roesch MR (2015). Executive control signals in orbitofrontal cortex during response inhibition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(9), 3903–3914. 10.1523/JNEUROSCI.3587-14.2015 [PubMed: 25740519]
- Chikazoe J, Jimura K, Hirose S, Yamashita K, Miyashita Y, & Konishi S (2009). Preparation to inhibit a response complements response inhibition during performance of a stop-signal task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(50), 15870–15877. 10.1523/JNEUROSCI.3645-09.2009 [PubMed: 20016103]
- Chudasama Y, & Robbins TW (2003). Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: Further evidence for the functional heterogeneity of the rodent frontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(25), 8771–8780. [PubMed: 14507977]
- Davidson RJ, Putnam KM, & Larson CL (2000). Dysfunction in the neural circuitry of emotion regulation—A possible prelude to violence. *Science (New York, N.Y.)*, 289(5479), 591–594. 10.1126/science.289.5479.591
- Dias R, Robbins TW, & Roberts AC (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, 380(6569), 69–72. 10.1038/380069a0 [PubMed: 8598908]

- Eagle DM, Baunez C, Hutcheson DM, Lehmann O, Shah AP, & Robbins TW (2008). Stop-signal reaction-time task performance: Role of prefrontal cortex and subthalamic nucleus. *Cerebral Cortex* (New York, N.Y.: 1991), 18(1), 178–188. 10.1093/cercor/bhm044
- Fellows LK, & Farah MJ (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain: A Journal of Neurology*, 126(Pt 8), 1830–1837. 10.1093/brain/awg180 [PubMed: 12821528]
- Gallagher M, McMahan RW, & Schoenbaum G (1999). Orbitofrontal cortex and representation of incentive value in associative learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 19(15), 6610–6614. [PubMed: 10414988]
- Gottfried JA, O'Doherty J, & Dolan RJ (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* (New York, N.Y.), 301(5636), 1104–1107. 10.1126/science.1087919
- Groman SM, Keistler C, Keip AJ, Hammarlund E, DiLeone RJ, Pittenger C, Lee D, & Taylor JR (2019). Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes. *Neuron*, 103(4), 734–746.e3. 10.1016/j.neuron.2019.05.042 [PubMed: 31253468]
- Hardung S, Epple R, Jäckel Z, Eriksson D, Uran C, Senn V, Gibor L, Yizhar O, & Diester I (2017). A Functional Gradient in the Rodent Prefrontal Cortex Supports Behavioral Inhibition. *Current Biology*, 27(4), 549–555. 10.1016/j.cub.2016.12.052 [PubMed: 28190729]
- Harlow JM (1993). Recovery from the passage of an iron bar through the head. *History of Psychiatry*, 4(14), 274–281. 10.1177/0957154X9300401407
- Hikosaka K, & Watanabe M (2004). Long- and short-range reward expectancy in the primate orbitofrontal cortex. *The European Journal of Neuroscience*, 19(4), 1046–1054. 10.1111/j.0953-816x.2004.03120.x [PubMed: 15009152]
- Horn JDV, Irimia A, Torgerson CM, Chambers MC, Kikinis R, & Toga AW (2012). Mapping Connectivity Damage in the Case of Phineas Gage. *PLOS ONE*, 7(5), e37454. 10.1371/journal.pone.0037454 [PubMed: 22616011]
- Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, & Polkey CE (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience*, 16(3), 463–478. 10.1162/089892904322926791 [PubMed: 15072681]
- Izquierdo A, & Jentsch JD (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology*, 219(2), 607–620. 10.1007/s00213-011-2579-7 [PubMed: 22134477]
- Izquierdo A, Suda RK, & Murray EA (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(34), 7540–7548. 10.1523/JNEUROSCI.1921-04.2004 [PubMed: 15329401]
- Jones B, & Mishkin M (1972). Limbic lesions and the problem of stimulus—Reinforcement associations. *Experimental Neurology*, 36(2), 362–377. 10.1016/0014-4886(72)90030-1 [PubMed: 4626489]
- Kazama A, & Bachevalier J (2009). Selective aspiration or neurotoxic lesions of orbital frontal areas 11 and 13 spared monkeys' performance on the object discrimination reversal task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(9), 2794–2804. 10.1523/JNEUROSCI.4655-08.2009 [PubMed: 19261875]
- Kheramin S, Body S, Ho M-Y, Velázquez-Martinez DN, Bradshaw CM, Szabadi E, Deakin JFW, & Anderson IM (2003). Role of the orbital prefrontal cortex in choice between delayed and uncertain reinforcers: A quantitative analysis. *Behavioural Processes*, 64(3), 239–250. 10.1016/s0376-6357(03)00142-6 [PubMed: 14580695]
- Kim J, & Ragozzino ME (2005). The involvement of the orbitofrontal cortex in learning under changing task contingencies. *Neurobiology of Learning and Memory*, 83(2), 125–133. 10.1016/j.nlm.2004.10.003 [PubMed: 15721796]
- Machado CJ, & Bachevalier J (2007). Measuring reward assessment in a semi-naturalistic context: The effects of selective amygdala, orbital frontal or hippocampal lesions. *Neuroscience*, 148(3), 599–611. 10.1016/j.neuroscience.2007.06.035 [PubMed: 17693034]

- Majid DSA, Cai W, Corey-Bloom J, & Aron AR (2013). Proactive selective response suppression is implemented via the basal ganglia. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 33(33), 13259–13269. 10.1523/JNEUROSCI.5651-12.2013 [PubMed: 23946385]
- Mansouri FA, Buckley MJ, & Tanaka K (2014). The Essential Role of Primate Orbitofrontal Cortex in Conflict-Induced Executive Control Adjustment. *Journal of Neuroscience*, 34(33), 11016–11031. 10.1523/JNEUROSCI.1637-14.2014 [PubMed: 25122901]
- McAlonan K, & Brown VJ (2003). Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behavioural Brain Research*, 146(1–2), 97–103. 10.1016/j.bbr.2003.09.019 [PubMed: 14643463]
- Mobini S, Body S, Ho M-Y, Bradshaw CM, Szabadi E, Deakin JFW, & Anderson IM (2002). Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology*, 160(3), 290–298. 10.1007/s00213-001-0983-0 [PubMed: 11889498]
- Morrison SE, & Salzman CD (2009). The Convergence of Information about Rewarding and Aversive Stimuli in Single Neurons. *Journal of Neuroscience*, 29(37), 11471–11483. 10.1523/JNEUROSCI.1815-09.2009 [PubMed: 19759296]
- Ostlund SB, & Balleine BW (2007). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(18), 4819–4825. 10.1523/JNEUROSCI.5443-06.2007 [PubMed: 17475789]
- Padoa-Schioppa C (2011). Neurobiology of economic choice: A good-based model. *Annual Review of Neuroscience*, 34, 333–359. 10.1146/annurev-neuro-061010-113648
- Pickens CL, Sadoris MP, Gallagher M, & Holland PC (2005). Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. *Behavioral Neuroscience*, 119(1), 317–322. 10.1037/0735-7044.119.1.317 [PubMed: 15727536]
- Pickens CL, Sadoris MP, Setlow B, Gallagher M, Holland PC, & Schoenbaum G (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(35), 11078–11084. [PubMed: 14657165]
- Reitman F (1946). Orbital cortex syndrome following leucotomy. *American Journal of Psychiatry*, 103(2), 238–241. 10.1176/ajp.103.2.238
- Riceberg JS, & Shapiro ML (2012). Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(46), 16402–16409. 10.1523/JNEUROSCI.0776-12.2012 [PubMed: 23152622]
- Roberts AC, & Wallis JD (2000). Inhibitory control and affective processing in the prefrontal cortex: Neuropsychological studies in the common marmoset. *Cerebral Cortex (New York, N.Y.: 1991)*, 10(3), 252–262. 10.1093/cercor/10.3.252
- Rolls ET, Hornak J, Wade D, & McGrath J (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery, and Psychiatry*, 57(12), 1518–1524. 10.1136/jnnp.57.12.1518
- Rudebeck PH, & Murray EA (2011). Balkanizing the primate orbitofrontal cortex: Distinct subregions for comparing and contrasting values. *Annals of the New York Academy of Sciences*, 1239, 1–13. 10.1111/j.1749-6632.2011.06267.x [PubMed: 22145870]
- Rudebeck PH, & Rich EL (2018). Orbitofrontal cortex. *Current Biology: CB*, 28(18), R1083–R1088. 10.1016/j.cub.2018.07.018 [PubMed: 30253144]
- Rudebeck PH, Saunders RC, Prescott AT, Chau LS, & Murray EA (2013). Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nature Neuroscience*, 16(8), 1140–1145. 10.1038/nn.3440 [PubMed: 23792944]
- Saver JL, & Damasio AR (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, 29(12), 1241–1249. 10.1016/0028-3932(91)90037-9 [PubMed: 1791934]
- Schoenbaum G, Chiba AA, & Gallagher M (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience*, 1(2), 155–159. 10.1038/407 [PubMed: 10195132]

- Schoenbaum Geoffrey, Nugent SL, Saddoris MP, & Setlow B (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport*, 13(6), 885–890. 10.1097/00001756-200205070-00030 [PubMed: 11997707]
- Schoenbaum Geoffrey, Roesch MR, Stalnaker TA, & Takahashi YK (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews. Neuroscience*, 10(12), 885–892. 10.1038/nrn2753 [PubMed: 19904278]
- Schoenbaum Geoffrey, Setlow B, Nugent SL, Saddoris MP, & Gallagher M (2003). Lesions of Orbitofrontal Cortex and Basolateral Amygdala Complex Disrupt Acquisition of Odor-Guided Discriminations and Reversals. *Learning & Memory*, 10(2), 129–140. 10.1101/lm.55203 [PubMed: 12663751]
- Schoenbaum Geoffrey, Setlow B, Saddoris MP, & Gallagher M (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron*, 39(5), 855–867. 10.1016/s0896-6273(03)00474-4 [PubMed: 12948451]
- Schuck NW, Cai MB, Wilson RC, & Niv Y (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, 91(6), 1402–1412. 10.1016/j.neuron.2016.08.019 [PubMed: 27657452]
- Solbakk A-K, Funderud I, Løvstad M, Endestad T, Meling T, Lindgren M, Knight RT, & Krämer UM (2014). Impact of orbitofrontal lesions on electrophysiological signals in a stop signal task. *Journal of Cognitive Neuroscience*, 26(7), 1528–1545. 10.1162/jocn\_a\_00561 [PubMed: 24392904]
- Stalnaker TA, Cooch NK, & Schoenbaum G (2015). What the orbitofrontal cortex does not do. *Nature Neuroscience*, 18(5), 620–627. 10.1038/nn.3982 [PubMed: 25919962]
- Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, & Schoenbaum G (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience*, 14(12), 1590–1597. 10.1038/nn.2957 [PubMed: 22037501]
- Tremblay L, & Schultz W (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, 398(6729), 704–708. 10.1038/19525 [PubMed: 10227292]
- Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, Brown JW, Chamberlain SR, Chambers CD, Colnius H, Colzato LS, Corneil BD, Coxon JP, Dupuis A, Eagle DM, Garavan H, Greenhouse I, Heathcote A, Huster RJ, ... Boehler CN (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife*, 8. 10.7554/eLife.46323
- Wallis JD, & Miller EK (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *The European Journal of Neuroscience*, 18(7), 2069–2081. 10.1046/j.1460-9568.2003.02922.x [PubMed: 14622240]
- Walton ME, Behrens TEJ, Buckley MJ, Rudebeck PH, & Rushworth MFS (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*, 65(6), 927–939. 10.1016/j.neuron.2010.02.027 [PubMed: 20346766]
- Wilson RC, Takahashi YK, Schoenbaum G, & Niv Y (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2), 267–279. 10.1016/j.neuron.2013.11.005 [PubMed: 24462094]
- Winstanley CA, Theobald DEH, Cardinal RN, & Robbins TW (2004). Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(20), 4718–4722. 10.1523/JNEUROSCI.5606-03.2004 [PubMed: 15152031]
- Yoo SBM, & Hayden BY (2018). Economic Choice as an Untangling of Options into Actions. *Neuron*, 99(3), 434–447. 10.1016/j.neuron.2018.06.038 [PubMed: 30092213]

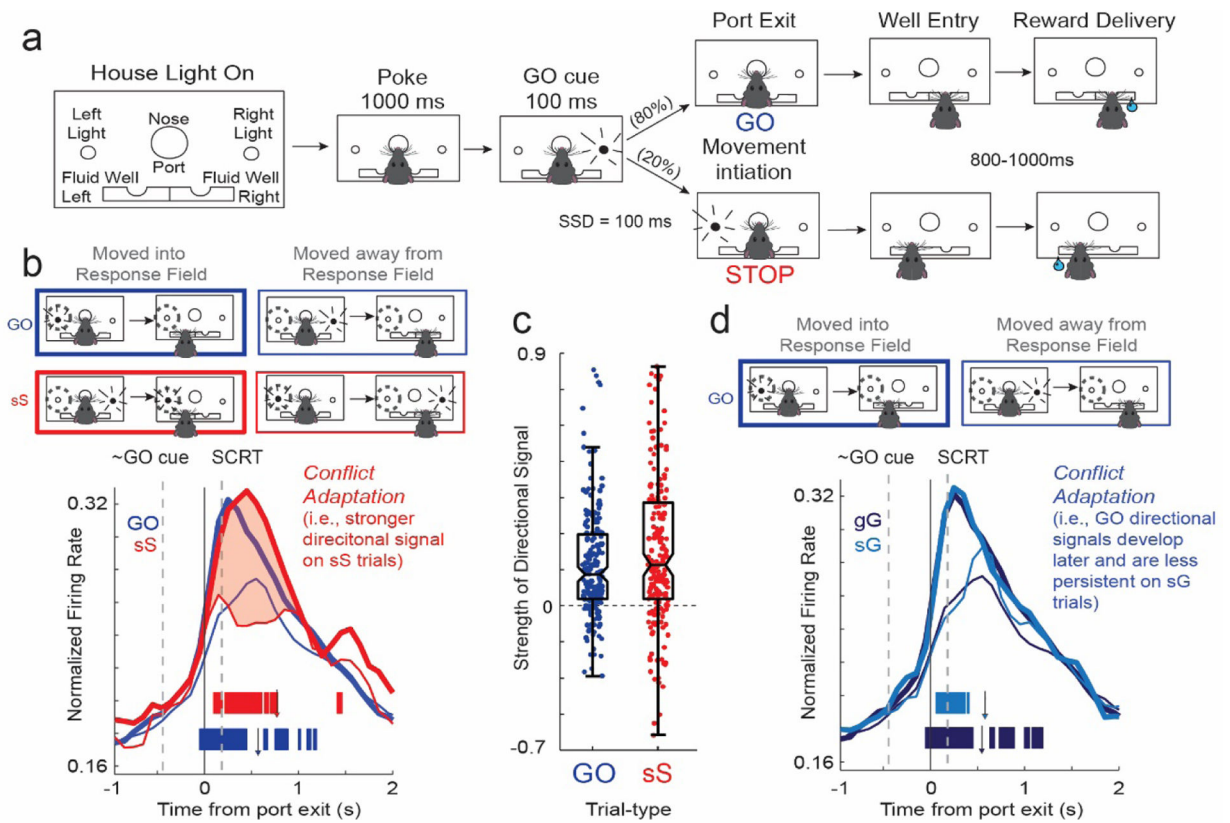




**Figure 1. Activity in OFC reflects inhibitory control.**

**a.** Illustration of task. Monkeys were trained to fixate on a center point before a GO cue was presented to either the left or right. On 67% of trials monkeys needed to make a saccade in the direction of the GO cue in order to receive reward. On the remaining 33% of trials a STOP cue (grey square) was displayed in the center after the GO cue was presented, directing the monkey to inhibit its initial response. **b.** Example firing from neuron T25 across 579 trials. Firing is presented for GO (blue), successful STOP trials (red) and failed STOP trials (orange). **c.** Ensemble analysis of pre-go signal data using post-stop trained decoders show accurate prediction of successful versus failed STOP trials. Time points in yellow denote start time of 100msec boxcars having percent accuracies of classification above 50%. Data above the red line indicates the 95 percentile value from permutation control tests. Graphs were provided by Benjamin Hayden and adapted from (Balasubramani et al., 2020).





**Figure 2. Executive control signals in rodent OFC during response inhibition.**

**a.** Illustration of the stop-change task used by Bryden and Roesch (Bryden & Roesch, 2015).

Rats were instructed to hold their nose in the center port for 1 s, at which point a GO cue would be presented to either the left or right of the rat. On 80% of trials the GO cue instructed the rat as to which of the two fluid wells to move to in order to receive reward. On the remaining 20% of trials, STOP trials, after the initial GO cue, a STOP cue was presented on the opposite side, instructing the rat to inhibit its initial response in the direction of the GO cue, in favor of responding in the direction of the STOP cue. **b.** Population histogram showing average normalized firing rates of increasing-type OFC neurons ( $n = 209$ ) on GO (blue) and sS (stop preceding a STOP trial; red) trials. Direction firing was strongest (filled region between thick and thin lines) on sS suggesting OFC is activated by previous instances of conflict/ need to engage inhibitory control mechanisms. Thick lines represent firing in the each cell's preferred direction or into the response field and thin lines represent firing in the non-preferred direction or away from response field. Tick marks represent 100 ms bins (slid by 10 ms) where there was a significant difference between response directions (Wilcoxon;  $p < 0.01$ ). Note that selectivity emerges before the stop change reaction time (i.e., SCRT; time needed to inhibit behavior) and is stronger on STOP trials. **c.** Plots the normalized difference firing during the response epoch between actions made into and away from the response field (i.e., into-away/ into+away). The strength of the directional signal was significantly stronger on sS trials compared to GO and gS (not shown) trials (Wilcoxon;  $p < 0.05$ ). These results suggest that OFC contributes to inhibitory control by strengthening directional signals during response conflict. **d.** Average firing during GO trials broken down

by preceding trial-type: sG = STOP preceding GO; gG = GO preceding GO. Tick marks represent 100 ms bins (slid by 10 ms) where there was a significant difference between response directions (Wilcoxon;  $p < 0.01$ ). Note that differences in firing between response directions became significant later and were less persistent on sG trials. Thus, after STOP trials the emergence of responding the first cue is proactively diminished or inhibited to improve performance if inhibition is to be required again. Downward arrows represent average time of fluid well entry (i.e., completion of behavioral response). All graphs were adapted from (Bryden & Roesch, 2015).