Systems/Circuits

Prior Cocaine Self-Administration Increases Response–Outcome Encoding That Is Divorced from Actions Selected in Dorsal Lateral Striatum

Amanda C. Burton,[1,2] Gregory B. Bissonette,[1] Adam C. Zhao,[1] Pooja K. Patel,[1] and Matthew R. Roesch[1,2]

[1]Department of Psychology and [2]Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland 20742

Dorsal lateral striatum (DLS) is a highly associative structure that encodes relationships among environmental stimuli, behavioral responses, and predicted outcomes. DLS is known to be disrupted after chronic drug abuse; however, it remains unclear what neural signals in DLS are altered. Current theory suggests that drug use enhances stimulus–response processing at the expense of response–outcome encoding, but this has mostly been tested in simple behavioral tasks. Here, we investigated what neural correlates in DLS are affected by previous cocaine exposure as rats performed a complex reward-guided decision-making task in which predicted reward value was independently manipulated by changing the delay to or size of reward associated with a response direction across a series of trial blocks. After cocaine self-administration, rats exhibited stronger biases toward higher-value reward and firing in DLS more strongly represented action–outcome contingencies independent from actions subsequently taken rather than outcomes predicted by selected actions (chosen-outcome contingencies) and associations between stimuli and actions (stimulus–response contingencies). These results suggest that cocaine self-administration strengthens action–outcome encoding in rats (as opposed to chosen-outcome or stimulus–response encoding), which abnormally biases behavior toward valued reward when there is a choice between two options during reward-guided decision-making.

Key words: action–value; cocaine; decision; dorsal striatum; rat; single neuron

Significance Statement

Current theories suggest that the impaired decision-making observed in individuals who chronically abuse drugs reflects a decrease in goal-directed behaviors and an increase in habitual behaviors governed by neural representations of response–outcome (R–O) and stimulus–response associations, respectively. We examined the impact that prior cocaine self-administration had on firing in dorsal lateral striatum (DLS), a brain area known to be involved in habit formation and affected by drugs of abuse, during performance of a complex reward-guided decision-making task. Surprisingly, we found that previous cocaine exposure enhanced R–O associations in DLS. This suggests that there may be more complex consequences of drug abuse than current theories have explored, especially when examining brain and behavior in the context of a complex two-choice decision-making task.

Introduction

Chronic drug use is thought to impair model-based goal-directed mechanisms governed by response–outcome (R–O) encoding while enhancing model-free stimulus-guided processing that controls habits via stimulus–response (S–R) encoding (Jentsch and Taylor, 1999; Robbins and Everitt, 1999; Everitt et al., 2001; Cardinal et al., 2002; Yin et al., 2004; Daw et al., 2005, 2011, Everitt and Robbins, 2005, 2016; Vanderschuren et al., 2005; Hyman et al., 2006; Belin and Everitt, 2008; Tricomi et al., 2009; Balleine and O'Doherty, 2010; Koob and Volkow, 2010; Lucantonio et al., 2012, 2014; Wied et al., 2013). Although these theories are well supported, few have actually recorded from the brains of cocaine-exposed animals to determine whether such correlates are altered. Further, the large majority of this theory is based on simple paradigms meant to isolate behaviors that are under the control of R–O and S–R associations in tasks in which there is a singular response or no real choice at all (Pavlovian) (Robbins and Everitt, 1999; Everitt et al., 2001; Everitt and Rob-

bins, 2005; Schoenbaum and Setlow, 2005; Vanderschuren et al., 2005; Nelson and Killcross, 2006, 2013; Nordquist et al., 2007; Ostlund and Balleine, 2008; Redish et al., 2008; Hogarth et al., 2013; LeBlanc et al., 2013; Corbit et al., 2014; Lucantonio et al., 2014; Schmitzer-Torbert et al., 2015; Everitt and Robbins, 2016). Although these studies are elegant in their design and have provided critical information to the field of drug abuse and neural control of behavior, work is necessary to understand how brain function changes in behaviors that offer choices between different rewards in situations in which there is not necessarily a simple and direct pairing among stimuli, responses, and outcomes. Such paradigms might better reflect everyday decision-making and are already known to evoke multiple neural representations outside of the classic domain of R–O and S–R correlates that might be affected by drugs of abuse.

Here, we examine firing in dorsal lateral striatum (DLS), an area known to be affected by chronic cocaine use (Vanderschuren et al., 2005; Takahashi et al., 2007; Belin and Everitt, 2008; Everitt and Robbins, 2013; Lucantonio et al., 2014). In dorsal striatum, at least two different types of R–O correlates have been described in animals performing two-choice paradigms, one that reflects the relationship between selected actions and the outcomes that those actions predict, referred to as "chosen value," and another that reflects the value of possible actions independent from the action that will ultimately be selected, referred to as "action value" (Lau and Glimcher, 2007, 2008; Nakamura et al., 2012). Action-value signals are thought to represent the relationship between potential actions and their predicted value so that they can be compared during decision-making. Others refer to similar signals as a correlate of response bias (Lauwereyns et al., 2002; Nakamura et al., 2012). Response-bias signals in striatum emerge before the instruction to move, reflecting the behavioral bias that the animal has for one direction over another.

Currently, it is unknown how these neural representations are affected by chronic cocaine use. However, we do know that, when faced with a choice between two options, rats that have been exposed previously to cocaine will more strongly bias their behavior toward more valuable options compared with controls (Roesch et al., 2007; Simon et al., 2007; Mendez et al., 2010). To better understand the neural correlates that give rise to this behavioral effect, we implemented a 2-week cocaine self-administration protocol with a 1-month withdrawal period and then recorded from single neurons in DLS while rats performed a reward-guided decision-making task.

Materials and Methods

Subjects. Male Long–Evans rats were obtained at 175–200 g from Charles River Laboratories. Rats were tested at the University of Maryland, College Park, in accordance with University and National Institutes of Health guidelines.

Reward-guided decision-making. Before surgery, all rats were trained on the reward-guided decision-making task (see Fig. 1A) for ~6 weeks. Rats were mildly water deprived to ensure motivation to complete a session of task performance. Rats were shaped to nose poke in a central odor panel that was controlled via computer to lengthen the amount of time spent in the odor port and the adjacent fluid wells to receive reward. Once proper responding was attained, rats were introduced to the different instructive odors. On each trial, nose poke into the odor port after house light illumination resulted in delivery of an odor cue to a hemicylinder located behind this opening. One of three different odors (2-octanol, pentyl acetate, or carvone) was delivered to the port on each trial. One odor instructed the rat to go to the left fluid well to receive reward (forced choice), a second odor instructed the rat to go to the right fluid well to receive reward (forced choice), and a third odor indicated

that the rat could obtain reward at either well (free choice). Odors were counterbalanced across rats. The meaning of each odor did not change across sessions. Odors were presented in a pseudorandom sequence such that the free-choice odor was presented on 7/20 trials and the left/right odors were presented in equal proportions.

During training and recording, one well was randomly designated as short (500 ms) and the other long (1–7 s) at the start of the session (see Fig. 1A, Block 1). In the second block of trials, these contingencies were switched (see Fig. 1A, Block 2). The length of the delay under long conditions abided by the following algorithm: the side designated as long started off as 1 s and increased by 1 s every time that side was chosen on a free-choice odor (up to a maximum of 7 s). If the rat chose the side designated as long <8 out of the previous 10 free-choice trials, then the delay was reduced by 1 s for each trial to a minimum of 3 s. The reward delay for long forced-choice trials was yoked to the delay in free-choice trials during these blocks. In later blocks, we held the delay preceding reward delivery constant (500 ms) while manipulating the size of the expected reward (see Fig. 1A, Blocks 3 and 4). The reward was a 0.05 ml bolus of 10% sucrose solution. For big reward, an additional bolus was delivered 500 ms after the first bolus for all big-reward trials in that block. At least 60 trials per block were collected for each neuron. Essentially, there were four basic trial types (short, long, big, or small) by two directions (left or right) by two stimulus types (free- or forced-choice odor).

Surgery. Seven rats were implanted with catheters for self-administration and electrodes for single-unit recordings in 1 survival surgery (cocaine group, n = 5 and control group, n = 2). Four additional control rats received electrodes only and were used as controls previously (Burton et al., 2014). Rats were catheterized in the jugular vein with SILASTIC tubing (0.02 × 0.037 inches; Dow Corning) with a modified 22 G 5-up cannula (Plastics One), which was then fed through the fascia layer over the shoulder and cemented next to the electrode implant site on the skull. In the same surgery, electrodes (drivable bundles of 10- to 25-μm-diameter FeNiCr wires) were implanted dorsal to DLS (1 mm anterior to bregma, ±3.2 mm laterally, and 3.5 mm ventral to brain surface). Electrodes were advanced daily (40–80 μm). See our previous study (Burton et al., 2014) for more detail on recording methods. We recorded 935 DLS neurons, 565 from six control rats (n = 62, 82, 83, 88, 119, and 131 neurons) and 370 from five cocaine-exposed rats (n = 19, 43, 81, 82, and 145 neurons).

Twelve-day self-administration protocol. After rats recovered from surgery, a 12 d self-administration protocol was implemented using Med Associates operant behavioral boxes. During days 1–6, the cocaine group (n = 5) self-administered a 1 mg/kg dosage of cocaine via lever press with a maximum of 30 infusions or 3 h time limit. During days 7–12, a 0.5 mg/kg dosage of cocaine was self-administered with a maximum of 60 infusions or a 3 h time limit. Sessions began with illumination of house lights, extension of the lever, and an initial illumination of a cue light above the lever for 2.3 seconds. Active lever presses were paired with the cue light above the lever, which stayed on for the duration of the cocaine infusion (2.3 s per active infusion). Active lever presses for drug infusion could only take place 20 s after the previous active lever press. Lever presses during the inactive period resulted in no cue-light illumination or reward delivery. The active lever was extended for the duration of the session. Rats were taken out of the operant boxes after the maximum infusion or maximum time limit was reached as described above.

There were two control groups. One group received electrodes and catheters as did the cocaine group. This group (n = 2) followed the same protocol as the cocaine group except rats received sucrose pellets (test diet, 45 mg; 1–6 d = 2 pellets; 7–12 d = 1 pellet). All aspects remained the same as per the cocaine group regarding session start, active lever presses, cue light, and duration of cue light. The other control group (n = 4) only received electrodes and did not self-administer sucrose pellets. The latter group's data were published previously (Burton et al., 2014).

Behavioral analysis. Behavior during self-administration was evaluated by computing the average number of lever presses daily across rats in each group (cocaine and control) and computing the average number of lever presses across days 1–6 and days 7–12 across rats in each group. Behavior during performance of the task was evaluated by computing the percentage choice of value conditions (short, long, big, or small) on free-choice trials as well as percentage correct and reaction time (odor

offset to odor port exit) on forced-choice trials for each value condition (short, long, big, or small) for the first 10 and last 10 trials of each block of trials (total of four blocks per session). Percentage choice was computed by determining how often the rat chose the more- and less-valued option on free-choice trials. Percentage correct was computed by determining whether the direction the animal chose was the same that was instructed by the forced-choice odor. Response bias for each session and each rat in each group was computed for free- and forced-choice trials by subtracting the percentage of low-value choice from high-value choice (divided by the sum) for each session and by subtracting percentage correct on low-value trials from percentage correct on high-value trials (divided by the sum) for each session and then averaging the two.

Neural analysis. Recorded waveforms were extracted from active channels during recording sessions and recorded to disk by an associated workstation with event time stamps from the behavioral computer. Extracted single units were then sorted in Offline Sorter using template matching (Plexon) and exported to NeuroExplorer to determine time-stamped events related to spike activity. All further analysis was done using MATLAB (The MathWorks). Analysis epochs were computed by taking the total number of spikes and dividing by time. That analysis epoch was taken from odor onset to odor port exit. The baseline epoch was 1 s before odor onset. Increasing- and decreasing-type neurons were categorized by whether they increased or decreased firing significantly compared with baseline, respectively ($p < 0.05$). A multifactor ANOVA ($p < 0.05$) was performed for each increasing- and decreasing-type neuron to determine whether activity was modulated by stimulus (free- vs forced-choice odors), response direction (contralateral vs ipsilateral), and expected outcome (short, long, big, or small). Chi-square tests ($p < 0.05$) were performed to assess differences in the counts of neurons showing significant modulation across control and cocaine-exposed rats.

Results

Self-administration

All rats were trained on the reward-guided decision-making task (Fig. 1*A*) before implantation of electrodes in DLS (Fig. 1*G,H*) and catheters for cocaine self-administration (see Materials and Methods for more detail). During performance of the reward-guided decision-making task, on each trial, rats responded to one of two adjacent wells after sampling an odor at a central port (Fig. 1*A*). Rats were trained to respond to three different odor cues: one odor that signaled reward in the right well (forced choice), a second odor that signaled reward in the left well (forced choice), and a third odor that signaled reward at either well (free choice). Across blocks of trials in each recording session, we manipulated either the length of the delay preceding reward delivery (Fig. 1*A*, Blocks 1–2; ~60 trial/block) or the size of the reward (Fig. 1*A*, Blocks 3–4; ~60 trial/block).

Two weeks after surgery, rats self-administered sucrose pellets or cocaine over the course of 12 d. During days 1–6 (1 mg/kg cocaine or 2 sucrose pellets), the average number of active lever presses was 21.9 (± 8.6 SD) and 28.5 (± 5.2 SD) for cocaine and sucrose, respectively. During days 7–12 (0.5 mg/kg cocaine or 1 sucrose pellet), the average number active lever presses were 46.5 (± 14.3 SD) and 60 (± 0 SD) for cocaine and sucrose, respectively. After a month-long withdrawal period, rats were placed back in behavioral boxes interfaced with Plexon recording systems and recording commenced for ~2 months.

### Cocaine biased behavior in the direction of high-value rewards

After a month-long withdrawal period, activity in DLS was recorded during performance of the task. During these sessions, rats that had self-administered cocaine exhibited stronger response biases toward higher-value reward. Control and cocaine groups tracked value across trials blocks choosing short delay and

large reward more often than long delay and small reward on free-choice trials, respectively (Fig. 1*B*). In an ANOVA with group (sucrose control, nonsucrose control, and cocaine), value (high and low), value manipulation (size and delay), and phase of learning within each block (early: first 10 trials; late: last 10 trials) as factors, there was a significant main effect of value ($F_{(1,7456)} = 3443$, $p < 0.05$) and no interaction between value and value manipulation ($F_{(2,7456)} = 0.06$, $p = 0.94$). There was a significant interaction between group and value ($F_{(2,7456)} = 60.8$, $p < 0.05$), with rats in the cocaine group choosing the high-value reward more often than controls in the last 10 free-choice trials during both size (Fig. 1*B*; $t_{(933)} = 10.51$, $p < 0.01$) and delay (Fig. 1*B*; $t_{(933)} = 8.95$, $p < 0.01$) manipulations.

Consistent with the bias on free-choice trials, rats in the cocaine group were more strongly drawn to the high-value reward on forced-choice trials. In the ANOVA with percentage correct as the dependent variable, there was a significant main effect of value ($F_{(1,7456)} = 376.49$, $p < 0.05$) and value manipulation ($F_{(1,7456)} = 258.49$, $p < 0.05$), as well as a significant interaction among group, value, and value manipulation ($F_{(2,7456)} = 807.4$, $p < 0.05$). During the last 10 forced-choice trial types per block, rats that had previously self-administered cocaine were significantly more biased toward the side that produced better reward, as evidenced by higher percentage correct scores on short-delay (Fig. 1*C*; $t_{(933)} = 3.26$, $p < 0.01$) and large-reward (Fig. 1*C*, $t_{(933)} = 3.10$, $p < 0.01$) forced-choice trials and lower percentage correct scores on small-reward forced-choice trials (Fig. 1*C*; $t_{(933)} = 4.00$, $p < 0.01$).

Exaggerated response bias observed during both free and forced choice did not emerge as a result of faster block switching because there were no differences between groups during the first 10 trials of delay blocks (Fig. 1*B*; short: $t_{(933)} = 0.99$, $p = 0.32$; long: $t_{(933)} = 0.99$, $p = 0.32$; Figure 1*C*; short: $t_{(933)} = 0.65$, $p = 0.51$; long: $t_{(933)} = 1.47$, $p = 0.14$). Further, cocaine-exposed rats were actually slower to reverse contingencies early in size blocks, choosing small reward more often than large reward in the first 10 free-choice trials (Fig. 1*B*; $t_{(933)} = 5.44$, $p < 0.01$) and trending toward diminished accuracy during the first 10 large-reward forced-choice trials (Fig. 1*C*; $t_{(933)} = 1.85$, $p = 0.06$).

Overall, cocaine self-administration decreased reaction times (odor offset to nose poke exit); rats in the cocaine group responded to the odor significantly faster than controls on forced-choice trials in delay blocks and on free-choice trials during both delay and size blocks. In the ANOVA with reaction time as the dependent variable, there was a significant main effect of group ($F_{(2,7456)} = 1223.36$, $p < 0.05$) and no interaction between group and value ($F_{(2,7456)} = 0.11$, $p = 0.89$). There was also a significant group-by-value manipulation interaction ($F_{(2,7456)} = 124.54$, $p < 0.05$) and a significant value-by-value manipulation interaction ($F_{(2,74560)} = 11.3$, $p < 0.05$); cocaine rats were faster on forced-choice trials during delay blocks during the first 10 trials (Fig. 1*D*; short: $t_{(933)} = 3.13$, $p < 0.01$; long: $t_{(933)} = 2.93$, $p < 0.01$) and the last 10 trials (Fig. 1*D*; short: $t_{(933)} = 2.96$, $p < 0.01$; long: $t_{(933)} = 2.58$, $p < 0.05$). Cocaine-exposed rats did not exhibit different reaction times on forced-choice trials during performance of size blocks during the first 10 trials (Fig. 1*D*; big: $t_{(933)} = 0.07$, $p = 0.95$; small: $t_{(933)} = 0.22$, $p = 0.830$) or the last 10 trials (Fig. 1*D*; big: $t_{(933)} = 0.46$, $p = 0.65$; small: $t_{(933)} = 0.87$, $p = 0.38$). Finally, over all free-choice trial-types, cocaine exposed rats were faster relative to controls (Fig. 1*E*; short: $t_{(933)} = 3.87$, $p < 0.01$; long: $t_{(933)} = 3.61$, $p < 0.1$; big: $t_{(933)} = 2.27$, $p < 0.05$; small: $t_{(933)} = 2.06$, $p < 0.05$).
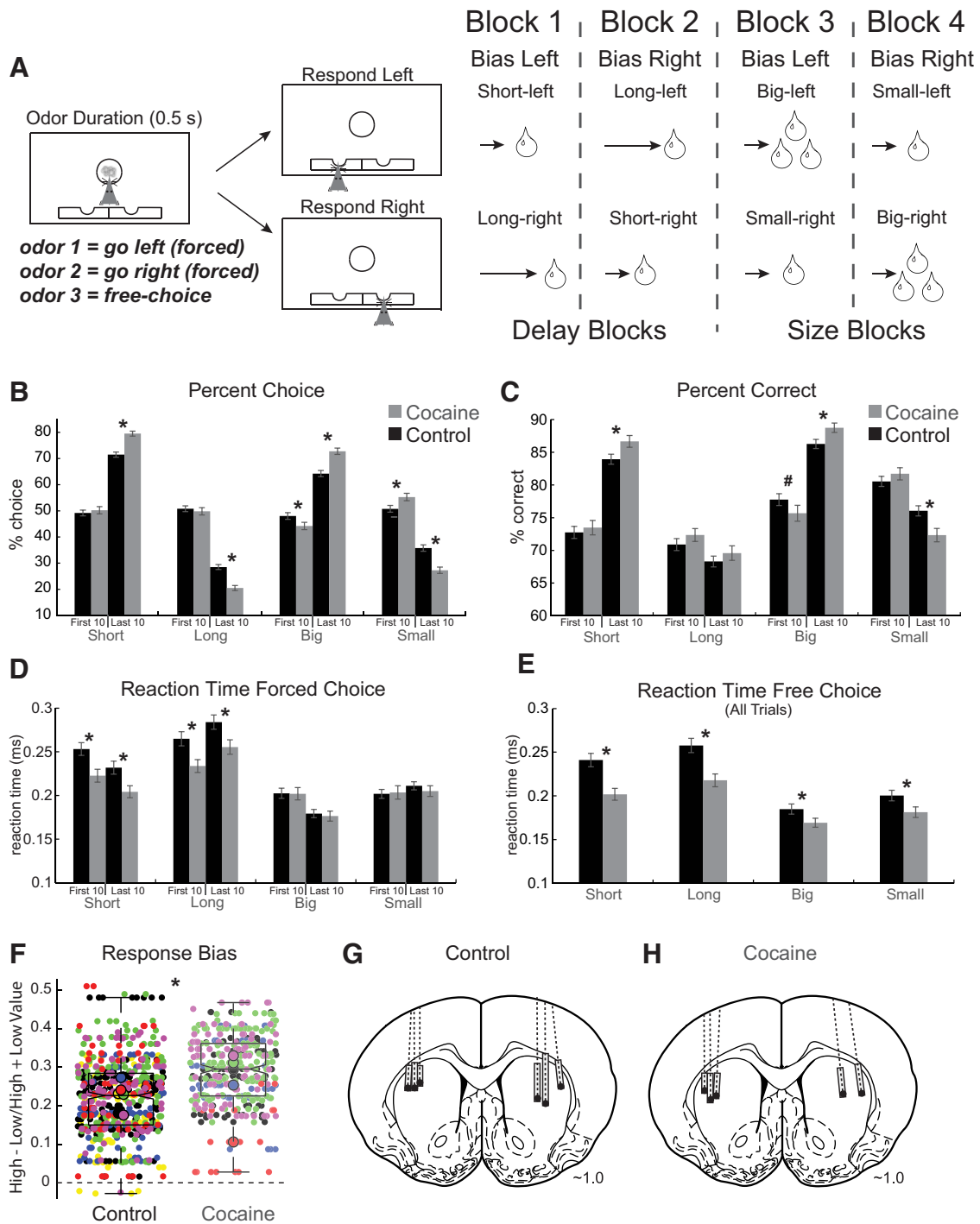
Figure 1. *A*, Task schematic showing sequence of events in one trial (left) and the sequence of blocks in a session (right). Rats were required to nose poke in the odor port for 0.5 s before the odor turned on for 0.5 s instructing them to respond to the adjacent fluid wells below where they would receive liquid sucrose reward after 500 – 7000 ms. For each recording session, one fluid well was arbitrarily designated as short (a short 500 ms delay before reward) and the other designated as long (a relatively long 1–7 s delay before reward) (Block 1). After the first block of trials (~60 trials), contingencies unexpectedly reversed (Block 2). With the transition to Block 3, the delays to reward were held constant across wells (500 ms), but the size of the reward was manipulated. The well designated as long during the previous block now offered two to three fluid boli, whereas the opposite well offered one bolus. The reward stipulations again reversed in Block 4. *B*, Percentage choice on free-choice trials in each value manipulation over the first 10 and last 10 trials of each block averaged across animals and sessions (controls, black bars; cocaine, gray bars). *C*, Percentage correct on forced-choice trials in the same manner as *B*. For these analyses, the sample was behavior associated with each unit to better represent changes in behavioral output as it relates to the firing rates of single neurons that contribute to the single-unit and population analysis described below. *D*, Reaction time (odor port exit minus odor offset) on forced-choice trials in the same manner as *B* and *C*. *E*, Reaction time (odor port exit minus odor offset) on all free-choice trials for each value manipulation. *F*, Response bias for each session and each rat in each group computed for free- and forced-choice trials by subtracting the percentage of low-value choice from high-value choice (divided by the sum) for each session and subtracting the percentage correct on low-value trials from the percentage correct on high-value trials (divided by the sum) for each session and then averaging over the two. Each color represents a single rat; small dots represent each session and large dots represent the median overall all sessions within one rat. *p < 0.05, #p = 0.06, t test. Error bars indicate SEM. *G*, *H*, Location of recording sites (Paxinos and Watson). Gray boxes mark the extent of the recording locations.

Overall, rats exposed to cocaine were biased toward higher value reward locations on both free- and forced-choice trials. This is further illustrated in Figure 1F, which plots a combined response bias measure for each session and each rat in each group. The response bias index was computed for the last 10 free- and forced-choice trials by subtracting the percentage of low-value choice from high-value choice (divided by the sum) for each session and by subtracting the percentage correct on low-value trials from the percentage correct on high-value trials (divided by the sum) for each session and then averaging over the two. Small dots represent each session and large dots represent the average over all sessions within one rat. Across sessions, the bias index was significantly larger in rats that had self-administered cocaine, consistent with the analysis described above ($t_{(933)} = 15.1, p < 0.01$). Only one rat (Fig. 1F, pale red) from the cocaine group fell below the median control response bias. Also, note that within the control group all rat averages fell between the lower and upper quartile, demonstrating that sucrose self-administration (Fig. 1F, sucrose controls, black and blue dots) did not affect response biases observed during task performance.

DLS firing reflects action–outcome contingencies divorced from the action selected after cocaine self-administration

In control and cocaine-exposed rats, 126 (22%) and 100 (27%) neurons increased and 262 (47%) and 124 (34%) decreased firing during odor sampling (odor onset to port exit) compared with baseline (1 s before odor onset; Wilcoxon; $p < 0.05$), respectively. The frequency of neurons that increased responding did not differ between groups ($\chi^2 = 1.5$; $p = 0.23$); however, there were significantly fewer that decreased responding in rats that had self-administered cocaine ($\chi^2 = 6.1$; $p < 0.05$). To determine how the firing of these neurons were modulated during task performance, we performed an ANOVA ($p < 0.05$) with outcome (short, long, big, or small), stimulus type (free or forced odor), and response direction (contralateral or ipsilateral) as factors on firing during the time between odor onset and odor port exit on trials in which reward was delivered (i.e., correct trials only). Initially, we hypothesized that chronic cocaine use would amplify correlates related to S–R processing by increasing the counts of neurons exhibiting significant main effects of stimulus or response direction or interactions between stimulus and response direction. Instead, we found an increase in neurons that exhibited an interaction between response direction and expected outcome (Fig. 2A). The only significant group difference was that more neurons exhibited a significant interaction between response and outcome in cocaine compared with control rats (Fig. 2A, $\chi^2 = 4.08$, $p < 0.05$). This occurred only in neurons that increased firing during the odor epoch compared with baseline. Although neurons that decreased firing were modulated selectively by task parameters, the counts of neurons showing significant main or interaction effects did not differ between control and cocaine-exposed rats (Fig. 2B).

As described in the introduction, in animals performing a two-choice reward paradigm, two different R–O correlates emerge, one that reflects the association between the action selected and the predicted outcome and the other that reflects the contingency between reward and response direction independent of the action selected. An example of the former is illustrated by the firing of the single neuron plotted in Figure 3A. Neurons such as this one encode the relationship between the selected action (contralateral or ipsilateral) and the outcome predicted (short, long, big, or small). This particular neuron had a response field contralateral to the recording electrode (response field illus-
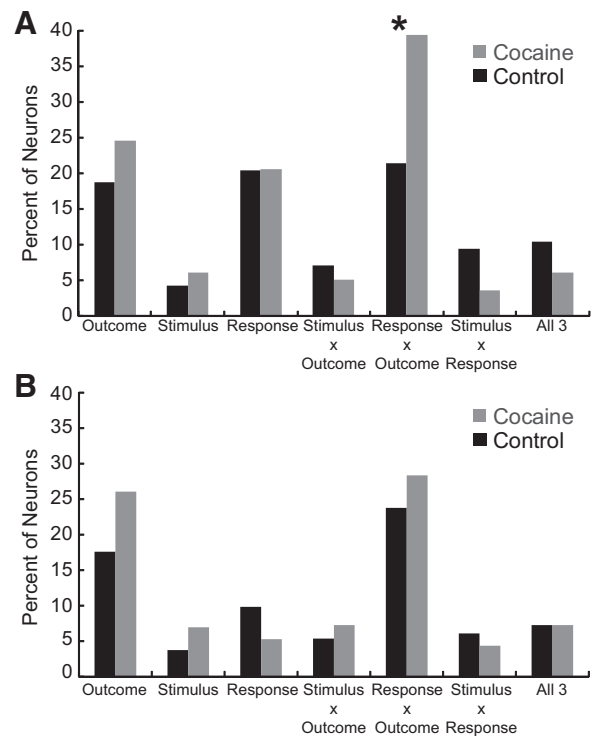

Figure 2. Percentage of increasing-type neurons (**A**) (control, $n = 126$, 22%; cocaine, $n = 100$, 27%) and decreasing-type neurons (**B**) (control, $n = 262$, 47%; cocaine, $n = 123$, 34%) with activity that was modulated significantly by outcome, stimulus, response, and interactions between these three factors in the ANOVA. *$p < 0.05$, χ^2 test.

trated by dashed circles), firing strongly for actions made in the contralateral direction (Fig. 3A, top). In addition, it was outcome selective, firing the most when the selected contralateral action would result in the delivery of reward after a short delay (Fig. 3A, top left). Therefore, this neuron conveyed information about the direction the rat was to select and the outcome that was predicted by that selection. We will refer to this as "chosen-outcome" encoding.

Other neurons in DLS that exhibited a significant interaction between response and outcome did not encode the outcome predicted by the action selected, but instead the outcome predicted if the rat was to move in a particular direction independent of whether the animal actually chose to move in that direction. Take, for example, the firing of the neuron shown in Figure 3B: this neuron fired strongly in blocks of trials in which the short delay is on the contralateral side, both when the rat will make a response toward the contralateral fluid well to obtain a reward after a short delay and when the rat will make a response to the ipsilateral fluid well to obtain reward after a long delay (Fig. 3B, blue). Therefore, activity of this neuron conveyed information that the short delay (a high-value reward) is in the contralateral direction independent of the action that was subsequently chosen. We will refer to this correlate as an "action–outcome" signal. In the following paragraphs, we will show that action–outcome correlates, as defined here, are overemphasized in the DLS after cocaine self-administration both at the population level and in the counts of single neurons.

The average population firing for control and cocaine-exposed rats is illustrated in Figure 4. In these plots, each neuron's preferred and nonpreferred outcome and direction was determined by the neuronal response that elicited the most activity (spikes/s) during cue sampling (odor onset to odor port exit). In
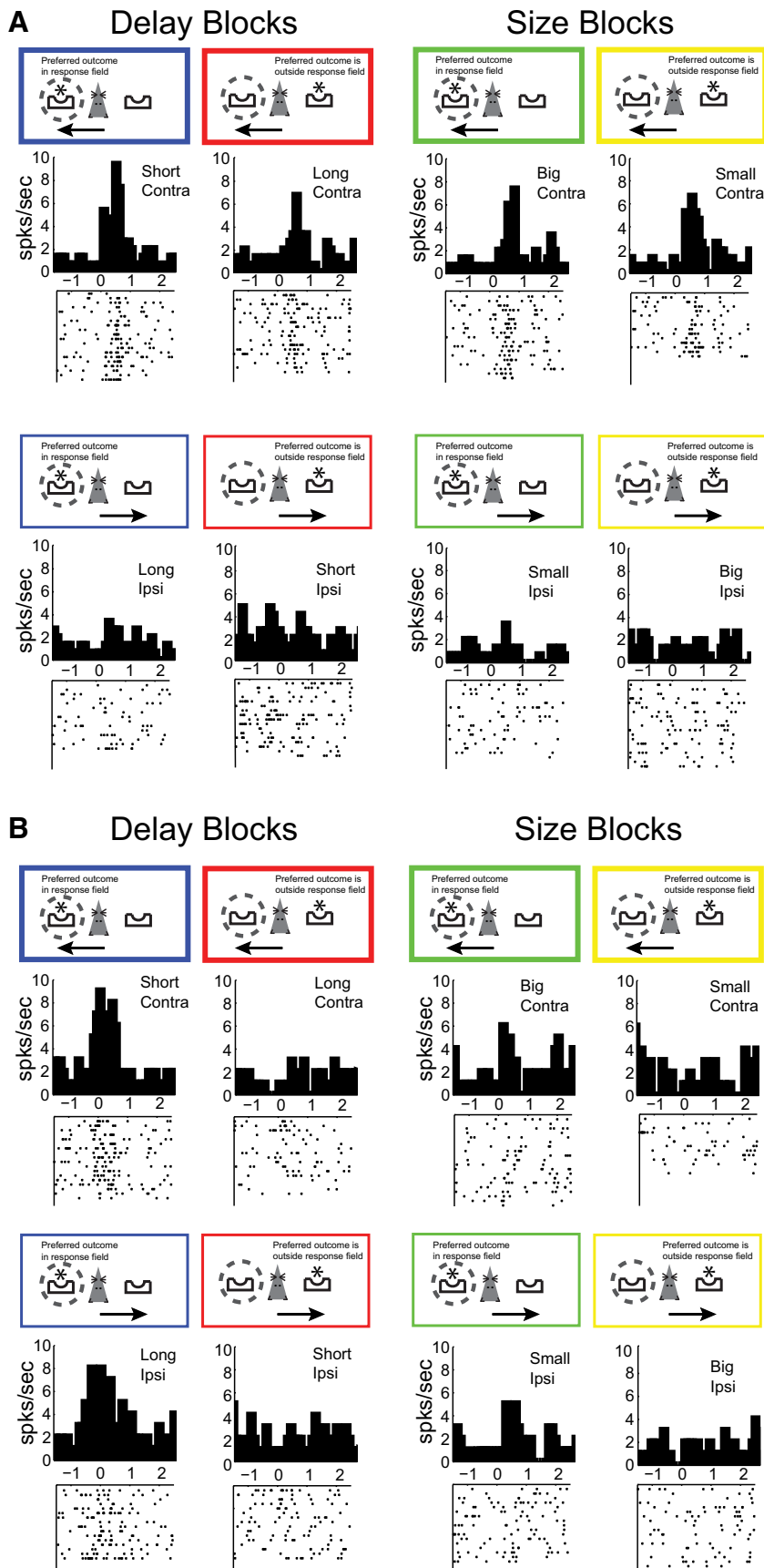
this figure, blue reflects activity in the neuron's preferred block of trials (i.e., when the preferred outcome was in the response field as represented by the asterisk in the dashed circle) and red reflects blocks of trials when the preferred outcome was outside of the response field (i.e., the non-preferred outcome). Green and yellow represent blocks of trials when the outcome of the same or opposite value (compared with blue and red, respectively) was in or outside the neuron's response field (e.g., if the preferred outcome was short, then the same and opposite value would be large and small, respectively). Individual trial types are color coded in Figure 3 as an example of how each neuron's response patterns fit into this color scheme.

Like the firing of the single-cell example shown in Figure 3A, population firing in the DLS of controls was highly selective. In controls, firing was significantly different during odor sampling for responses made into, but not away from, the response field for the preferred outcome (Fig. 4A,B; significance illustrated by SEM ribbons and running t test; $p <$ 0.01), reflecting the relationship between the selected action and the predicted outcome. This was not true after cocaine self-administration. Like controls, activity was also higher for responses made into the response field for the cell's preferred outcome (Fig. 4C); however, activity was also significantly higher for behavioral responses made away from the neuron's response field in the same block of trials (Fig. 4D). Therefore, the population activity in cocaine-exposed rats reflected the location of the preferred outcome in a particular context and did not reflect the outcome selected as in controls. That is, activity was high when the preferred outcome was in the cell's response field independent of the direction that the rat would eventually move, similar to the single-unit example in Figure 3B. Notably, selectivity in cocaine-exposed rats emerged before cue onset (black tick marks before zero), consistent with the idea that activity in DLS reflected the contingencies between actions and outcomes in a block of trials, which are known to the rat before cue onset.

The average population histogram suggests that there is an overabundance of
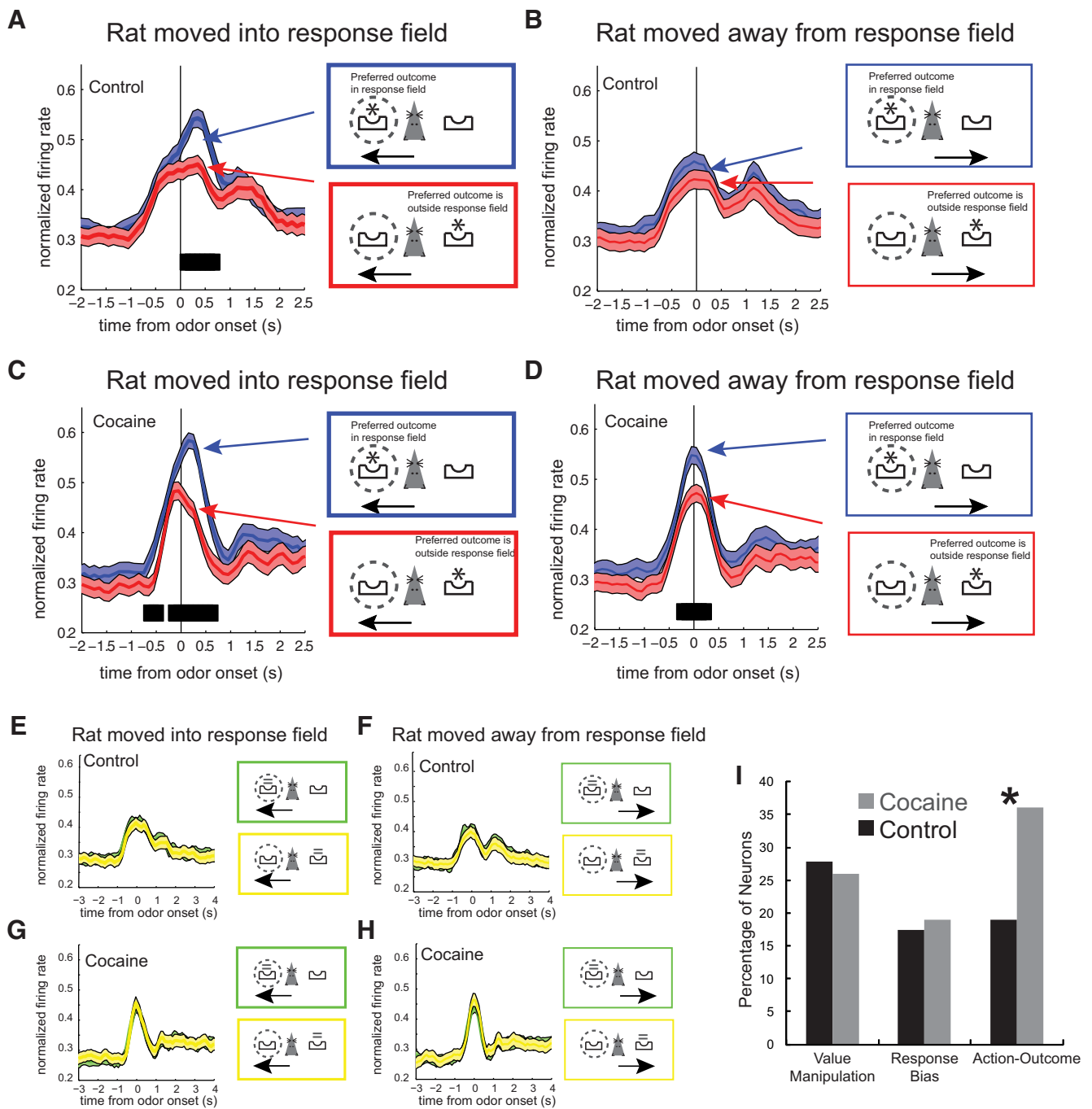
←

Figure 3. Single-neuron examples of chosen-outcome (**A**) and action–outcome (**B**) encoding. Dashed circles around fluid wells in the colored boxes indicate the neuron's response field. Asterisks in the colored boxes refer to the location of the preferred outcome (inside or outside response field). Arrows within the colored boxes refer to the direction in which the animal made a response (into or away from response field). Thick and thin lines represent movements into and away from the response field, respectively. Activity is aligned to odor onset (0 s) and binned at 100 ms. Each tick mark equals one action potential.

Figure 4. Action–outcome encoding is enhanced at the population level in cocaine-exposed rats. **A**, **B**, Population average for control rats ($n = 126$). **A**, Trials in which the rat moved into the neuron's response field (preferred direction; thick lines). **B**, Trials in which the rat moved away from the neurons response field (nonpreferred direction; thin lines). Blue and red represent when the preferred and nonpreferred outcome was in the response field, respectively. Dashed circles around fluid wells in the colored boxes indicate the neuron's response field. Asterisks in the colored boxes refer to the location of the preferred outcome (inside or outside of the response field). Activity is aligned to odor onset (0 s). Significance is illustrated by SEM ribbons and black tick marks ($p < 0.01$, running t test of two consecutive 100 ms bins) under each histogram. **C**, **D**, Average firing over time for cocaine-exposed rats ($n = 100$) for response made into (**C**) and away (**D**) from the response field. Firing was high when the preferred outcome was in the response field (blue) regardless of whether the rat moved into (**A**) or away (**B**) from the response field. Neural activity is plotted in the same manner as in **A** and **B**. **E–H**, Histograms of neural population activity in DLS of control rats (**E**, **F**) and cocaine-exposed rats (**G**, **H**) for the same- and opposite-value outcomes (green and yellow colored lines and boxes, respectively) with respect to **A–D**. Neural activity is plotted in the same manner as **A–D**. Histograms are normalized by dividing the maximum within each neuron. **I**, Proportion of neurons showing main effects of value manipulation (size or delay), response bias (contralateral response associated with high-value reward independent of value manipulation), or an interaction between these two factors (action–outcome) in ANOVA ($p < 0.05$). Cocaine-exposed rats had an increased number of neurons that exhibited action–outcome encoding compared with control rats. *$p < 0.05$, χ^2 test.

action–outcome encoding neurons after cocaine self-administration. To determine whether this was true, we performed a two-factor ANOVA with value manipulation (size or delay) and response bias (contralateral response associated with high-value reward vs

ipsilateral response associated with high-value reward regardless of value manipulation) as factors ($p < 0.05$). Neurons that show a main effect of response bias without an interaction with value manipulation would reflect stronger firing whenever high-value

outcomes (short delay or large reward) were in the response field (i.e., response bias or action value). Neurons that exhibit an interaction would reflect action–outcome contingencies, not response bias or action value, in that firing would only be higher for high-value outcomes within one value manipulation (size or delay).

The results of this analysis are shown in Figure 4I. The proportions of neurons that showed main effects of value manipulation and response bias exceeded chance levels in both groups ($\chi^2 > 40.97$, $p < 0.05$), but the frequency of main effects between cocaine-exposed and control rats did not differ significantly (value manipulation: cocaine = 26%; control = 28%; $\chi^2 = 0.007$, $p = 0.94$; response bias: cocaine = 19%; control = 17%; $\chi^2 = 0.006$, $p = 0.94$). However, as predicted by the population histograms (Fig. 4A–H) and the single-unit analysis described in Figure 3B, the counts of neurons that exhibited a significant interaction between value manipulation and response bias in cocaine animals (i.e., action–outcome) outnumbered those in controls significantly (Fig. 4I; cocaine = 36%; control = 19%; $\chi^2 = 4.11$, $p < 0.05$).

Action–outcome signals emerged earlier and were not stimulus specific after cocaine self-administration

From the above analysis, it is clear that prior cocaine self-administration induces stronger action–outcome selectivity. To determine whether this selectivity emerged earlier in a trial block in one group compared with the other, we plotted trial-by-trial neural activity during odor sampling when the preferred outcome was in the response field versus when the nonpreferred outcome was in the response field. The analysis was performed separately for rewarded free- and forced-choice trials and the trial at which activity was deemed selective was determined via a sliding t test ($p < 0.01$; Fig. 5, gray ticks).

In controls, selectivity developed at trials 8 and 16 for free- and forced-choice trials, respectively. Notably, on free-choice trials, activity reflected the previous block's contingencies before reflecting the new action–outcome pairing during free-choice (Fig. 5A, open ticks) in controls only. In cocaine-exposed rats, significant selectivity on free-choice trials emerged earlier than controls, arguably as early as trial 3, but convincingly by trial 7 (Fig. 5B). Under forced-choice trials, significant differences were evident by trial 6 in cocaine-exposed rats (Fig. 5D). Therefore, overall, we conclude that action–outcome encoding emerged earlier in cocaine-exposed rats.
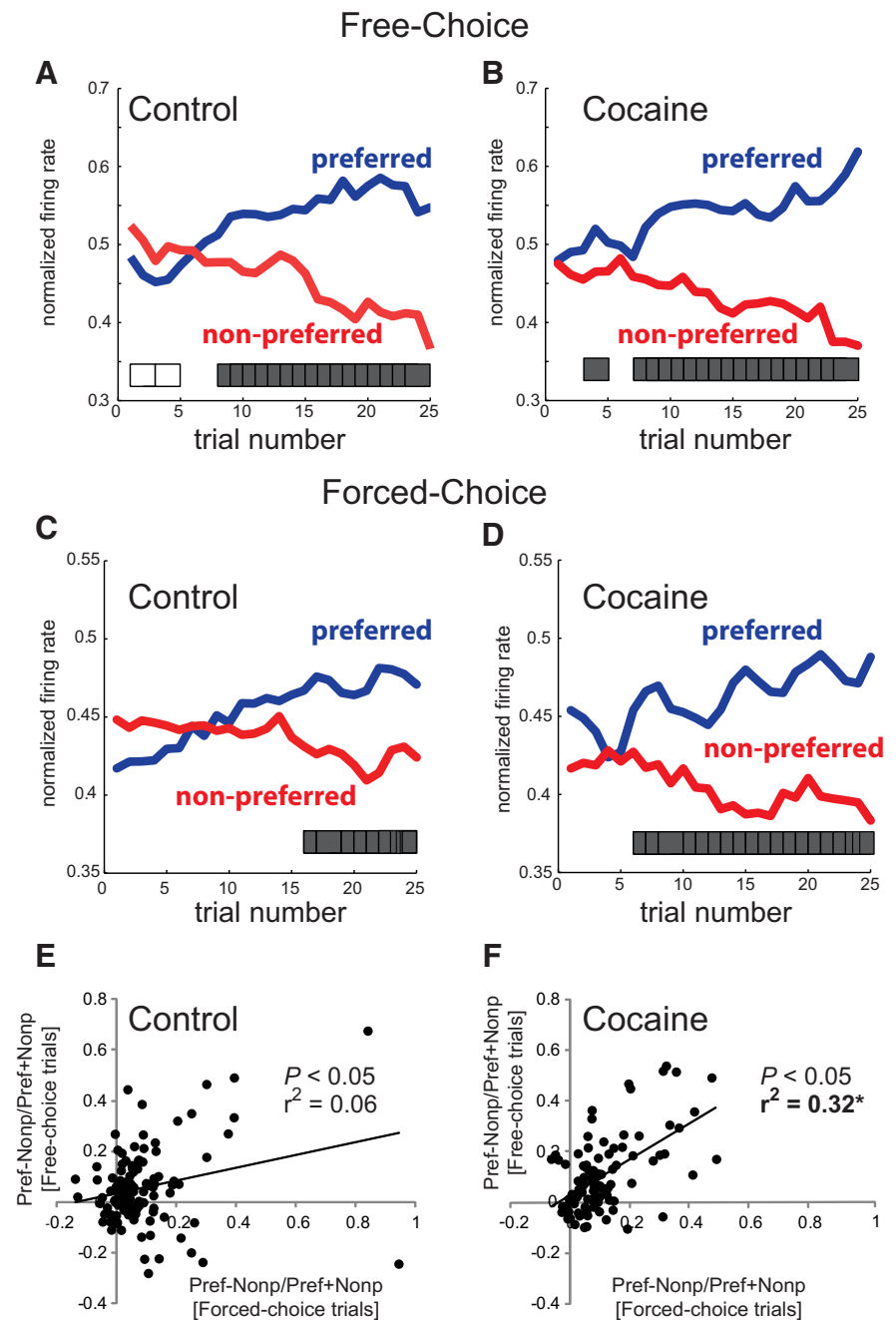
Figure 5. Neural selectivity emerges early in cocaine-exposed rats and was stimulus independent. ***A***–***D***, Trial-by-trial normalized neural activity during odor sampling on correct free- and forced-choice trials for control rats (***A***, ***C***) and cocaine-exposed rats (***B***, ***D***) when the preferred outcome was in the response field versus when the nonpreferred outcome was in the response field. Tick marks underneath plots indicate two consecutive trials in which there was a significant difference between preferred and nonpreferred action–outcome contingencies ($p < 0.01$, t test). ***E***, ***F***, Correlation between action–outcome selectivity (preferred minus nonpreferred/preferred plus nonpreferred) during free- and forced-choice trials. The neuron's preferred context was defined as the average spike activity of both free- and forced-choice trials that elicited the most activity during odor sampling. A positive correlation indicates that selectivity at the single-neuron level was similar on free- and forced-choice trials.

In the above analysis, preferred direction was defined separately for free- and forced-choice analysis. This was necessary because the firing of individual neurons might exhibit R–O selectivity for one but not the other trial type (Stalnaker et al., 2010). From that analysis, it cannot be determined whether single neurons that exhibit action–outcome signals did so regardless of whether it was a free- or forced-choice trial-type. If these signals genuinely reflect action–outcome contingencies within a block

of trials, then one would predict that the activity of these neurons might reflect this relationship independent from the stimulus presented at the beginning of the trial. Certainly, the observation that selectivity emerged before stimulus onset (Fig. 4) suggests that the effect is stimulus independent.

To further address this issue, we performed an analysis in which we defined the cell's preferred context based on the average of both free- and forced-choice trials. We then plotted the contingency (preferred minus nonpreferred/preferred plus nonpreferred) separately for forced- and free-choice trials (Fig. 5 E, F). Both groups exhibited a significant positive correlation (Fig. 5E, control: $r^2 = 0.06$, $p < 0.05$; Fig. 5F, cocaine: $r^2 = 0.32$, $p < 0.05$); however, the correspondence between free- and forced-choice trials was significantly stronger in the cocaine group compared with controls (Fig. 5F, $z = 2.87$, $p < 0.05$). This demonstrates that selective firing observed within a particular block of trials occurs regardless of whether the cue was a free- or forced-choice odor.

Discussion

Here, we show that rats exposed previously to cocaine exhibit stronger behavioral response biases toward higher-value reward and an overrepresentation of action–outcome signals at both the single-neuron and population levels. In cocaine-exposed rats, population firing in DLS did not reflect the outcome of the action that was about to occur (chosen-outcome encoding), but instead represented the location of the preferred outcome within the context of a particular block of trials, which we are referring to as action–outcome encoding. Absent from this dataset was any evidence that S–R associations were enhanced after chronic cocaine self-administration during performance of our reward-guided decision-making task.

Our results support the hypothesis that cocaine self-administration alters R–O and model-based mechanisms, but demonstrates that correlates are altered and not eliminated during decision-making. We show that population firing in DLS in cocaine-exposed rats failed to modify predictions based on the action selected, but instead showed an increase in neural activity reflecting the location of the preferred outcome regardless of subsequent instructions or movements in a particular block of trials. Further, action–outcome encoding in cocaine exposed rats was less stimulus dependent, failing to take into account states within each trial that informed predictions about reward availability. In fact, action–outcome selectivity in cocaine exposed rats emerged before the onset of the odor. These altered neural correlates likely bias behavior successfully on free-choice trials while diminishing performance on forced-choice trials. Overall, these results are consistent with a dysfunctional model-based system in that activity fails to represent the value of selected actions as governed by states in the task and overrepresent contingencies between potential outcomes and actions within the context of the block before and independent from the presented odor.

It might be argued that overrepresented action–outcome signals observed in cocaine-exposed rats reflect some sort of exaggerated S–R encoding. This does not seem to be the case for several reasons. First, firing rate selectivity observed in DLS differed between size and delay blocks even though S–R relationships were identical between them (i.e., smell this stimulus and make this response). Second, in the ANOVA performed on single neurons, we did not observe any differences between control and cocaine-exposed animals in neurons showing selectivity for stimulus type or response direction or an interaction between them (Fig. 2). Further, after cocaine self-administration, we found that these correlates were actually less stimulus bound, not more (Fig.

5F), meaning that neurons fired selectively within a particular context regardless of whether the cue was a free- or forced-choice odor. Therefore, this correlate does not appear to be modulated by stimuli and is outcome dependent, suggesting that it is not a form of S–R encoding.

It is also clear that other S–R correlates were not affected by cocaine self-administration in that, when we investigated how many neurons were selective for stimulus type or response direction (or an interaction between them), selectivity was similar between cocaine-exposed and control rats. This is true even though forced-choice contingencies (i.e., odor 1 = left, odor 2 = right) never changed over several months of training and recording. The fact that cocaine exposure did not affect stimulus and response processing was somewhat of a surprise to us because, in a previous study, we showed that lesions to nucleus accumbens (NAc) increased the counts of neurons that were selective for odors and responses in DLS and enhanced the overall strength of the signal at the population level (Burton et al., 2014). Considering that NAc is one of the first areas to be affected by drug use (Koob and Volkow, 2010; Steinberg et al., 2014; Keiflin and Janak, 2015; Lüscher, 2016), we expected a similar result after cocaine self-administration.

The lack of altered S–R correlates in our study likely reflects the complex nature of our two-choice paradigm in which there is no simple mapping among stimuli, responses, and outcomes, rather than being due to differences in doses or withdrawal times. Not only does this paradigm require constant tracking of reward across two responses, but also an awareness of what rewards are available on each trial as conveyed by odor identity. In this situation, behavior is governed by multiple action–outcome contingencies, which is known to encourage goal-directed behavior (Colwill and Rescorla, 1985; Dickinson et al., 2000; Colwill and Triola, 2002; Holland, 2004; Kosaki and Dickinson, 2010). Even though we did not classically devalue outcomes in our task and cannot prove definitively that rats were using representations of action–outcome contingencies, others have shown under similar circumstances that exposure to drugs of abuse does indeed leave goal-directed mechanisms intact and, in some cases, enhanced (Phillips and Vugler, 2011; Son et al., 2011; Halbout et al., 2016). In particular, Halbout and colleagues showed that, when rats are trained on two different action–outcome contingencies, those exposed to cocaine stop responding to the one that was devalued and, after contingency degradation, rats exposed to cocaine actually alter behavior more quickly than controls. Our findings are consistent with these results in that we did not see enhanced S–R encoding, did see increased action–outcome correlates, and rats exposed to cocaine made faster adjustments in behavior when contingencies change.

We conclude that prior cocaine exposure increases action–outcome processing without affecting the chosen-outcome and S–R encoding in our decision-making task. Here, we define action–outcome correlates as the relationship between actions and outcomes independent from that actual action that will be taken. In this way, it could be viewed that cocaine self-administration increases outcome encoding that is divorced from the action that will be selected. This correlate is more of a reflection of the contingencies available during decision-making as opposed to a representation of what will ultimately be selected. Such correlates can only be examined in the context of neural recording in a paradigm with at least two choices. Notably, representations of response- or action–outcome associations as more classically defined in the learning theory literature (i.e., selected action–outcome or chosen-outcome) were not altered by cocaine self-

administration. The increase in action–outcome correlates that we describe here are more similar to response bias or action-value signals described in the caudate of primates (Lauwereyns et al., 2002; Lau and Glimcher, 2007, 2008; Nakamura et al., 2012), which are thought to represent the association between actions and outcomes to push the motor system toward decisions that lead to high-value outcomes independent from instructed/selected movement. As in these reports, this signal emerged before instructional cues reflecting the context in which rewards were distributed before the decision period. Here, we show that similar signals, albeit outcome dependent (i.e., specific to size or delay), are amplified after chronic cocaine use at the expense of correlates that inform behavior via computations of predicted outcomes based on upcoming decisions.

References

Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35:48–69. CrossRef Medline

Belin D, Everitt BJ (2008) Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. Neuron 57:432–441. CrossRef Medline

Burton AC, Bissonette GB, Lichtenberg NT, Kashtelyan V, Roesch MR (2014) Ventral striatum lesions enhance stimulus and response encoding in dorsal striatum. Biol Psychiatry 75:132–139. CrossRef Medline

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352. CrossRef Medline

Colwill RM, Rescorla RA (1985) Instrumental responding remains sensitive to reinforcer devaluation after extensive training. J Exp Psychol Anim Behav Process 11:520–536. CrossRef

Colwill RM, Triola SM (2002) Instrumental responding remains under the control of the consequent outcome after extended training. Behav Processes 57:51–64. CrossRef Medline

Corbit LH, Chieng BC, Balleine BW (2014) Effects of repeated cocaine exposure on habit learning and reversal by N-acetylcysteine. Neuropsychopharmacology 39:1893–1901. CrossRef Medline

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711. CrossRef Medline

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215. CrossRef Medline

Dickinson A, Smith J, Mirenowicz J (2000) Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci 114:468–483. CrossRef Medline

Everitt BJ, Robbins TW (2005) Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 8:1481–1489. CrossRef Medline

Everitt BJ, Robbins TW (2013) From the ventral to the dorsal striatum: devolving views of their roles in drug addiction. Neurosci Biobehav Rev 37:1946–1954. CrossRef Medline

Everitt BJ, Robbins TW (2016) Drug addiction: updating actions to habits to compulsions ten years on. Annu Rev Psychol 67:23–50. CrossRef Medline

Everitt BJ, Dickinson A, Robbins TW (2001) The neuropsychological basis of addictive behaviour. Brain Res Rev 36:129–138. CrossRef Medline

Halbout B, Liu AT, Ostlund SB (2016) A closer look at the effects of repeated cocaine exposure on adaptive decision making under conditions that promote goal-directed control. Front Psychiatry 7.

Hogarth L, Balleine BW, Corbit LH, Killcross S (2013) Associative learning mechanisms underpinning the transition from recreational drug use to addiction. Ann N Y Acad Sci 1282:12–24. CrossRef Medline

Holland PC (2004) Relations between Pavlovian-instrumental transfer and reinforcer devaluation. J Exp Psychol Anim Behav Process 30:104–117. CrossRef Medline

Hyman SE, Malenka RC, Nestler EJ (2006) Neural mechanisms of addiction: the role of reward-related learning and memory. Annu Rev Neurosci 29:565–598. CrossRef Medline

Jentsch JD, Taylor JR (1999) Impulsivity resulting from frontostriatal dysfunction in drug abuse: Implications for the control of behavior by reward-related stimuli. Psychopharmacology (Berl) 146:373–390. CrossRef Medline

Keiflin R, Janak PH (2015) Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. Neuron 88:247–263. CrossRef Medline

Koob GF, Volkow ND (2010) Neurocircuitry of addiction. Neuropsychopharmacology 35:217–238. CrossRef Medline

Kosaki Y, Dickinson A (2010) Choice and contingency in the development of behavioral autonomy during instrumental conditioning. J Exp Psychol Anim Behav Process 36:334–342. CrossRef Medline

Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. J Neurosci 27:14502–14514. CrossRef Medline

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463. CrossRef Medline

Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. Nature 418:413–417. CrossRef Medline

LeBlanc KH, Maidment NT, Ostlund SB (2013) Repeated cocaine exposure facilitates the expression of incentive motivation and induces habitual control in rats. PLoS One 8.

Lucantonio F, Stalnaker TA, Shaham Y, Niv Y, Schoenbaum G (2012) The impact of orbitofrontal dysfunction on cocaine addiction. Nat Neurosci 15:358–366. CrossRef Medline

Lucantonio F, Caprioli D, Schoenbaum G (2014) Transition from "model-based" to "model-free" behavioral control in addiction: Involvement of the orbitofrontal cortex and dorsolateral striatum. Neuropharmacology 76:407–415. CrossRef Medline

Lüscher C (2016) The emergence of a circuit model for addiction. Annu Rev Neurosci 39:257–276. CrossRef Medline

Mendez IA, Simon NW, Hart N, Mitchell MR, Nation JR, Wellman PJ, Setlow B (2010) Self-administered cocaine causes long-lasting increases in impulsive choice in a delay discounting task. Behav Neurosci 124:470–477. CrossRef Medline

Nakamura K, Santos GS, Matsuzaki R, Nakahara H (2012) Differential reward coding in the subdivisions of the primate caudate during an oculomotor task. J Neurosci 32:15963–15982. CrossRef Medline

Nelson A, Killcross S (2006) Amphetamine exposure enhances habit formation. J Neurosci 26:3805–3812. CrossRef Medline

Nelson AJ, Killcross S (2013) Accelerated habit formation following amphetamine exposure is reversed by D1, but enhanced by D2, receptor antagonists. Front Neurosci 7:76. CrossRef Medline

Nordquist RE, Voorn P, de Mooij-van Malsen JG, Joosten RN, Pennartz CM, Vanderschuren LJ (2007) Augmented reinforcer value and accelerated habit formation after repeated amphetamine treatment. Eur Neuropsychopharmacol 17:532–540. CrossRef Medline

Ostlund SB, Balleine BW (2008) On habits and addiction: an associative analysis of compulsive drug seeking. Drug Discov Today Dis Models 5:235–245. CrossRef Medline

Phillips GD, Vugler A (2011) Effects of sensitization on the detection of an instrumental contingency. Pharmacol Biochem Behav 100:48–58. CrossRef Medline

Redish AD, Jensen S, Johnson A (2008) A unified framework for addiction: vulnerabilities in the decision process. Behav Brain Sci 31:415–487. CrossRef Medline

Robbins TW, Everitt BJ (1999) Drug addiction: bad habits add up. Nature 398:567–570. CrossRef Medline

Roesch MR, Takahashi Y, Gugsa N, Bissonette GB, Schoenbaum G (2007) Previous cocaine exposure makes rats hypersensitive to both delay and reward magnitude. J Neurosci 27:245–250. CrossRef Medline

Schmitzer-Torbert N, Apostolidis S, Amoa R, O'Rear C, Kaster M, Stowers J, Ritz R (2015) Post-training cocaine administration facilitates habit learning and requires the infralimbic cortex and dorsolateral striatum. Neurobiol Learn Mem 118:105–112. CrossRef Medline

Schoenbaum G, Setlow B (2005) Cocaine makes actions insensitive to outcomes but not extinction: Implications for altered orbitofrontal-amygdalar function. Cereb Cortex 15:1162–1169. Medline

Simon NW, Mendez IA, Setlow B (2007) Cocaine exposure causes long-term increases in impulsive choice. Behav Neurosci 121:543–549. CrossRef Medline

Son JH, Latimer C, Keefe KA (2011) Impaired formation of stimulus–

response, but not action–outcome, associations in rats with methamphetamine-induced neurotoxicity. Neuropsychopharmacology 36:2441–2451. CrossRef Medline

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2010) Neural correlates of stimulus–response and response–outcome associations in dorsolateral versus dorsomedial striatum. Front Integr Neurosci 4:12. CrossRef Medline

Steinberg EE, Boivin JR, Saunders BT, Witten IB, Deisseroth K, Janak PH (2014) Positive reinforcement mediated by midbrain dopamine neurons requires D1 and D2 receptor activation in the nucleus accumbens. PLoS One 9.

Takahashi Y, Roesch MR, Stalnaker TA, Schoenbaum G (2007) Cocaine exposure shifts the balance of associative encoding from ventral to dorsolateral striatum. Front Integr Neurosci 1:11. Medline

Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232. CrossRef Medline

Vanderschuren LJ, Di Ciano P, Everitt BJ (2005) Involvement of the dorsal striatum in cue-controlled cocaine seeking. J Neurosci 25:8665–8670. CrossRef Medline

Wied HM, Jones JL, Cooch NK, Berg BA, Schoenbaum G (2013) Disruption of model-based behavior and learning by cocaine self-administration in rats. Psychopharmacology (Berl) 229:493–501. CrossRef Medline

Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 19:181–189. CrossRef Medline