SUPPLEMENTARY INFORMATION

METHODS

Apparatus

Monkeys sat in a wheeled transport cage in front of a joystick that could be moved in two directions (lift upwards or turned to the right) for Experiments 1 and 2. Reward food pellets (Noyes formula L/I, Research Diets Inc.) were delivered to a food-well placed directly above the joystick. The reward pellets were delivered according to the different schedules used in each experiment. Once delivered the reward pellets were immediately available to the animals. A food-box was also placed to the left of the food-well. At the end of the session, after the last trial had been completed, the food-box containing the bulk of the animal’s daily food rations, including proprietary monkey food and fruit, was opened. Testing was performed with the room lighting on unless otherwise specified. The joystick apparatus contained solenoids that could prevent a joystick movement in one of the directions or lock the joystick in the central position once a movement was made. Experiment 3 was conducted in a similar manner, except the monkeys were placed in front of a touchscreen monitor and food rewards were delivered to a food well to the right of the screen. All tasks were controlled by a computer to which the joystick apparatus and overhead lamp were connected.
Experiment 1: Error- and reward-guided action selection

Three distinct procedures of Experiment 1 (Experiments 1A, B, and C) were used both before and after surgery because it has been suggested that the ACC's importance may depend on whether other sensory cues are also present and because in some previous experiments, errors have been signaled by reward value reduction rather than by complete omission of reward\textsuperscript{1-3}. In the Standard Procedure (Experiment 1A) an error was signaled by the simple omission of reward; rewarded trials were immediately followed by a food pellet but no food pellet was given on error trials. Each trial began with an 800 Hz, 300 ms “start” tone. A rewarded movement was followed by a 500 Hz, 400 ms “correct” tone (conditioned stimulus), delivery of a single food pellet, and then a 600 ms inter-trial interval (ITI) before the next start tone. Incorrect responses were followed only by a 600 ms ITI. No cues were given to indicate an incorrect movement. Between trials, subjects were free to move within the ITI and movements within the ITI were recorded (ITI errors).

In the Salient Error Procedure (Experiment 1B) errors were followed by switching off the cubicle illumination for the duration of the ITI whenever an action led to an error. In the Delay-Error Procedure (Experiment 1C) the food pellet was not omitted when errors were made on task-imposed switch trials, but instead the value of the reward was gradually reduced by delaying food and accompanying auditory tone by 500 ms (delay-imposed reward). The delay increased by 500 ms for each subsequent error. The joystick was locked in the central position during the ITI in Experiment 1C; animals were therefore not able to make further responses during the ITI even when it was
lengthened. Once the subject switched to the correct response each correct response was rewarded without delay until the next switch trial and any further errors were simply followed by omission of the food pellet. The program was designed so that subjects could not prematurely change responses at the switch trial without encountering at least one delay-imposed reward. As in Experiment 1B the ITI was set at 1 s on other trials.

Pre-operatively, subjects were tested on each of procedures, 1A, 1B, 1C on different weeks. In each case they were required to maintain a performance criterion of 85% average correct across five successive days during the pre-operative session (750 rewarded trials). Lesions were then made in the ACCs group and three weeks were given for recovery in which no testing occurred for any of the subjects. After recovery, the animals were tested for five days (750 rewarded trials) on each task (regardless of performance level attained), starting first with Experiment 1A, followed by Experiment 1B and 1C, respectively. Animals were also tested on pre- and post-operative sessions of Experiment 3.

**Experiment 2: Dynamic foraging probabilistic matching task**

On every trial the availability of reward after each movement was determined by two independent probabilistic algorithms\textsuperscript{4,6}. On each trial, one, both, or neither response option could be allocated a reward. If, for example, a reward was allocated to response $X$ on trial $n$ and response $Y$ was selected on trial $n$, a reward automatically remained allocated to response $X$ on trial $n+1$ and remained allocated to response $X$ until response $X$ was selected on a later trial and the reward was collected by the monkey (i.e., the reward
probability for response $X$ on trial $n+1$ became 1.0 and stayed as such until response $X$ was chosen). Once response $X$ was selected, reward allocation for response $X$ reverted back to a probabilistic basis. This resembles the standard procedure used in matching experiments\textsuperscript{4} and in recent discrete trial matching paradigms\textsuperscript{5,6}.

Keeping a reward allocated to a response until that response was chosen discouraged subjects from ignoring the low probability response ($L$). If not selected for some time then the probability of receiving a reward for response $L$ could come to be greater than for the normally high probability response ($H$). This feature of the task also meant that the value of an action on any trial was dynamically dependent on which actions had recently been made.

The monkey’s objective was to work out which response was the more profitable response on any given testing day. Four action-reward pair probability ratios were used: 0.4:0.1, 0.5:0.2, 0.75:0.25, 1:0, where the numbers reflect the probability of reward for response $H$ and $L$ respectively. The four reward ratios were counterbalanced for both the lift and turn movements yielding eight conditions. All subjects performed the eight conditions in the same order (0.5:0.2, 0.1:0.4, 0.4:0.1, 0.2:0.5, 0.25:0.75, 1:0, 0:1, 0.75:0.25). By averaging across the two conditions for each reward ratio the experimental design counterbalanced for inter-individual variations in response bias. Occasionally an animal finished a day’s testing without acquiring a strong bias for $H$. On such occasions it was therefore necessary to give the animal an additional day of “remedial” testing to ensure that it acquired a bias for $H$ before moving on to the next test. This ensured, first,
that animals did not acquire a “failure set” that would have prevented them from attempting to learn which action was more profitable. Second, it ensured that animals that performed well on one day’s test were not at a disadvantage on a following day’s test where the opposite response may have been more profitable.

A cardinal feature of Experiment 2 is that as the expected reward from each decision is contingent on the rate that the monkey makes that decision, the fraction of responses of a given type tend to “match” the fraction of total rewards that can be earned by making that action, which optimizes overall reward intake. A decision’s value will vary with the time since it was last made because the expected reward is contingent on the rate of making the decision. Although this makes it difficult to calculate trial by trial variations in decision-making it is possible to calculate an average optimum ratio of response and the time it takes animals to reach this level (see below).

Rather than counting the number of trials until the optimum ratio was reached exactly, we examined how many trials it took each animal to obtain rewards at 97% of the optimal rate. There were a few cases in which the animals failed to reach the criterion. This was handled in a simple and conservative way: simply counting the total number of trials that were taken to obtain the total 150 food pellets that were available on each day’s testing session. Such a procedure treats an animal that failed to reach the criterion as if it had reached the criterion on the very last trial of the session. The procedure is conservative in that it may have led to some reduction in the ACCS group’s scores because this group was more likely to fail to reach the criterion. Nevertheless the procedure is not insensitive; if an animal were
very close to the criterion then it would have been performing nearly optimally, it would have received rewards at a high frequency, and it would have obtained all the food rewards available and completed the day’s test in a small number of trials. Importantly result patterns and significant statistical effects were similar to those shown in Figure 5a when 95% and 99% of optimal rate were employed as the threshold criteria instead of 97%. It can, however, be seen in Supplementary Figure 5 that when both responses are associated with low reward probabilities (for example, ratio 0.4:0.1) that many different response ratios still yield rewards at 97% of the optimum ratio (shown graphically by the large area of the yellow box in Supplementary Fig. 5 online). Thus, not surprisingly, at the extreme, many different response ratios would exceed a low criterion and it would not be possible to distinguish the performance of the two groups of animals. We therefore examined how many trials it took for each animal to get close to this optimum ratio in bins of fifty trials and then compared the number of trials taken by the control and ACCS animals.

**Experiment 3: Work-based cost-benefit decision making**

When animals were first encountering a given stimulus and learning its reward and effort associations, it was paired with another stimulus that was associated with no reward outcome. The non-rewarded visual stimulus was not used again subsequently but instead it was changed each time a new work/reward stimulus had to be learned.

Once they had learned all of the effort/reward stimuli (50 correct trials in a discrimination against a single response, unrewarded stimulus), the
animals then started pre-operative testing, involving choosing between a pair of rewarded stimuli that differed in their associated work demand and reward size. The stimulus with the greater reward size, or, if the reward sizes were equal, the stimulus requiring fewer responses, was termed the high reward stimulus \( (HR) \), the other the low reward stimulus \( (LR) \). Before being given any choice trials with a particular pair of stimuli, monkeys had to perform 40 consecutive forced trials, which alternated between trials on which only the \( HR \) was presented and those on which only the \( LR \) appeared. The side of presentation was varied pseudo-randomly. Sessions of choice trials also contained forced trials, always commencing with 4 successive forced (\( HR \) left-side of screen, \( LR \) left-side, \( HR \) right-side, \( LR \) right-side) and having 2 forced trials (1 \( HR \), 1 \( LR \), presented pseudo-randomly left or right) after every 4 choice ones. This ensured that animals repeatedly had to sample the effort-outcome association for each stimulus and prevented them from getting into a stereotyped pattern of responding. Subjects were tested with a particular pair of stimuli for 44 choice trials until behavior stabilized (meaning that performance was within \( \pm 1 \) of the previous day’s score) or, if their pattern of responses was variable, until they had completed 99 choice trials.

ANALYSES

General

Parametric ANOVAs were used because there are widely accepted parametric procedures for performing mixed model multi-factorial analyses and there are no generally accepted non-parametric alternatives. Parametric
ANOVA procedures are robust even when there are differences in variance. In addition the Huynh-Feldt correction procedure for dealing with violation of assumptions concerning variance was used when testing the significance of repeated measure factors. The results of Experiments 1 and 3 were analyzed with three factor ANOVAs using a factor of GROUP and testing SESSION and TRIAL TYPE. TRIAL TYPE in Experiment 3 corresponded to the three different work/reward problems tested. The results of Experiment 2 were analyzed with two factor ANOVAs using a factor of GROUP and TRIAL TYPE. TRIAL TYPE corresponded to the four different action-reward ratios tested.

**Experiment 1**

First, the total numbers of errors made in the pre- and post-operative sessions were compared. A between-subjects factor of GROUP (ACCS and CON groups) and a within-subjects factor of testing SESSION (pre-operative testing session or post-operative testing session) were used (Fig. 2a). A second analysis compared the probability of making the correct response after an error (“Error+1” trial) with the probability of making a correct response after a previously correct response (“Correct+1” trial) and so an additional factor of TRIAL TYPE (Error+1 or Correct+1) was also used (Fig. 2b). A third analysis compared performance on the ten trials before and ten trials after a switch trial. A variant of TRIAL TYPE (ten trials before or after switch trial) was also used (Fig. 2c). A fourth analysis examined the number of trials before the response was switched and the correct movement made after the
imposed switch trial (Supplementary Fig. 1 online) using factors of GROUP and SESSION.

**Experiment 1 – Sustaining rewarded behavior (EC curves)**

The EC curve analysis examined the impact of each correct (“C”) response (reward) on overall performance subsequent to each error (“E”). The first trial type of the EC analysis is E+1, the percentage correct on trials following an error. The measure corresponds to the Error +1 trial in Figure 2b. The second trial type in the EC analysis (EC+1) is the percentage correct on the trial following the first correct trial following an error. The third trial type (EC2+1) is the percentage correct on the trial following two successive correct trials following an error, and so on (Fig. 3, top). Pre-operatively both groups were poor at immediately correcting an error (E+1), but reached peak performance after two successive rewarded trials following an error (EC2+1). The bottom of Figure 3 displays the mean number of trials comprising each of the trial types of the EC analysis. The length of the EC analysis was limited to ensure every subject had at least 15 data points at the final trial type, which extended the analysis to EC8+1. Again analyses used factors of GROUP, SESSION, and a factor of TRIAL TYPE that now corresponded to a point on the EC curve.

**Experiment 1 - Influence of reward history on current choice (β analysis)**

An alternative way to examine the influence of previous reward history on subsequent choices is to use a multiple regression analysis. Such an
analysis can be used to obtain separate estimates of the weight of influence of each previous trial outcome in the reward history on the choice made on any given trial\[^6,8,9\]. A related approach has been used to estimate the influence of previous trial outcomes on the activity of a neuron in a subsequent trial\[^10\]. In the present study a multiple logistic regression was carried out to determine the combination of previous trials that best predicted the choice, \( Y \), on any given trial, \( i \). Current choices were included as the dependent categorical variable, and were represented as binary categories, 0 and 1 corresponding to pull and turn movements, respectively. The independent predictive variables were the outcomes of previous trials (\( X \)), which were represented as -1 and 1 for rewarded pulls and rewarded turns, respectively. Regressions generated a set of weights or “\( \beta \) values” reflecting the strength of influence of the outcome of the previous trial (\( \beta_1 X_{i-1} \)), the strength of the influence of the previous trial but one (\( \beta_2 X_{i-2} \)), and so on up to and including the influence of the outcome of the trial performed eight trials earlier, in the following form:

\[
\log(p(Y_i = 1)/\log(p(Y_i = 0))) = \beta_1 X_{i-1} + \beta_2 X_{i-2} + \beta_3 X_{i-3} + \ldots + \beta_8 X_{i-8} \quad \text{(Equation 1)}
\]

where \( p(Y_i=0) \) and \( p(Y_i=1) \) represent the probability, or relative frequency, of a pull and turn, respectively. Regressions were performed separately for each day’s data. For each animal, the \( \beta \) values for each point in the reward history were then averaged across days in the pre- and post-operative data sessions. The \( \beta \) values were then plotted separately for the control and ACC\(_S\) lesion group in both the pre- and post-operative sessions in Figure 4 and compared.
with a 3-way ANOVA implemented in a General Linear Model (GLM). The analyses of the reward history $\beta$ values used factors of GROUP, SESSION and TRIAL TYPE.

**Experiment 1 - Reaction Time Analysis**

*Supplementary Figure 2* shows the inter-response times of control (white bars) and ACC$_S$ monkeys (grey bars) in the pre- and post-operative sessions in Experiment 1A. Inter-response times (iRTs, the time between two successive movements) were used rather than reaction times (RT, the time between start cue and movement onset) to incorporate responses made during the ITI (“ITI errors”) in the assessment of behavioral response to reinforcement; an immediate behavioral response (e.g., an error that is quickly corrected in the ITI (“ITI error”)) can be indexed by the time between the two responses (iRT) whereas RTs in such an example do not reflect the behavioral response to the reinforcement of the previous trial but rather reflects the behavioral response to the start tone only after further ITI errors are withheld. When ITI errors occurred, the movement direction of the ITI error was investigated to identify whether the movement was in the correct direction based on the reinforcement of the previous movement. This examination of ITI error direction did not reveal any significant changes to the results presented in the main text.

The data in *Supplementary Figure 2* are divided into trials that follow previous correct (left) and previous error trials (right). The data are further subdivided according to whether or not the current trial is correct or incorrect. It can be seen that in general animals are quicker to respond after an error
than after a correct response. This is likely because they are engaged in consuming the food reward after a correct trial. One animal with an ACCS lesion was quicker to respond on the second test but i) performance was still within the control range, ii) the speeding of response was generalized across all conditions. Similar results were found upon analyzing the data from Experiments 1B and 1C. ACCS lesions had no effect on ITI error commission rates. Such analyses suggest that basic motivational factors, which might have influenced response speed, did not selectively affect error trials after the ACCS lesion.

**Experiment 2: Number of trials taken to reach optimal reward income rate**

The overall reward income rate is a function of both the response ratio (the ratio of response $H$ to response $L$) and the probability of reward associated with each action (which varied for each of the four action-reward ratios). Because of the dynamic nature of the task the reward probability of each action was dependent on the rate at which each action was made. The discrete trial matching task that we used, which involved no “change over delay”, is a particularly simple version of the matching task and the optimal response allocation ratio is the one at which both actions yield rewards at the same rate or “match”. This can be represented as:

$$p(R \mid H, r) = p(R \mid L, r) \quad (1)$$
\( p(R \mid H, r) \) indicates the conditional probability of a reward \( R \) given that response \( H \) is made and \( r \) is the average rate at which the monkey makes response \( H \). \( p(R \mid L, r) \) is the conditional probability of the reward \( R \) given that response \( L \) is made (again \( r \) remains the rate at which the monkey makes response \( H \)). It can be shown that:

\[
P(R \mid H, r) = 1 - \frac{r - rp}{p + r - rp}
\]

(2)

and that:

\[
P(R \mid L, r) = 1 - \frac{1 - r - q + rq}{1 - r + rq}
\]

(3)

Where \( p \) and \( q \) indicate the probability of a new reward being assigned to response \( H \) and \( L \) respectively (given that no untaken reward from a previous trial is already present). The two actions thus yield rewards at the same rate when the following equation is true:

\[
\therefore 1 - \frac{r - rp}{p + r - rp} = 1 - \frac{1 - r - q + rq}{1 - r + rq}
\]

(4)

Which can be reduced to the following equation where \( r_{opt} \) is the optimal rate for making response \( H \):

\[
r_{opt} = \frac{p - pq}{p + q - 2pq}
\]

(5)
The conditional probability of reward $R$ given each rate, $r$, of response $H$, $p(R \mid r)$, can be determined by using the following function for each reward ratio used (0.4:0.1, 0.5:0.2, 0.75:0.25, 1:0) where $p$ and $q$ again represent the probability of reward associated with each action $H$ and $L$.

$$p(R \mid r) = r \left[ 1 - \frac{p - rp}{p + r - rp} \right] + (1 - r) \left[ 1 - \frac{1 - r - q + rq}{1 - r + rq} \right]$$ (6)

**Supplementary Figure 4** shows the relative contributions of the $H$ and $L$ response in determining $r_{opt}$ for the case of the 0.4:0.1 action-reward ratio pairing. The point of intersection of the two curves denoting average reward income for each response is at $r_{opt}$. The function is shown further graphically for each action-reward ratio used (**Supplementary Fig. 5** online). The maximum of each curve is equal to the value of $r_{opt}$ calculated using equation 5. For each curve a yellow box indicates the region where rewards are being earned at more than 97% of the maximum possible rate.

It can be seen in **Supplementary Figure 5** that when both responses are associated with low reward probabilities (for example, ratio 0.4:0.1) that many different response ratios still yield rewards at 97% of the optimum ratio (shown graphically by the large area of the yellow box in **Supplementary Fig. 5** online). However, even in the condition with the lowest reward probability (0.4:0.1), a pattern of results and significant statistical effects to those shown in **Figure 5a** were found when 95% and 99% of optimal rate thresholds were employed. This can be visualized in **Supplementary Figure 6**, which shows the performance of the control and ACCS groups through the progression of the 0.4:0.1 condition. Each data point represents the ratio of $H$ to $L$.
responses during consecutive running average bins of 50 trials. The first data point is centered on the 26th trial and so it is possible by that time for some animals to have already learned to move away from the less profitable response and to make more of the day’s more profitable response. The threshold levels associated with 95%, 97%, and 99% of the optimum reward rate are shown (dashed yellow lines) and it can be seen that controls reached each threshold in many fewer trials compared to the ACCS animals. Similar differences between the controls and the ACCS group were seen for each action-reward ratio. We therefore examined how many trials it took for each animal to get to 97% of the optimum ratio in bins of fifty trials and then compared the number of trials taken by the control and ACCS animals (Fig. 5a).

SURGERY AND HISTOLOGY

The lesions included both banks of the ACCS from its rostral inception to a caudal position level with the midpoint of the precentral dimple that was visible during surgery (Figs. 1 and 8). This ensured that the lesion included the cingulate motor areas (CMAs) and the sulcal tissue lying rostral to this region. These regions are densely interconnected. While the CMAs project to the primary motor cortex and spinal cord11-13; it is the more rostral ACCs that is most interconnected with limbic and prefrontal regions such as the orbitofrontal and dorsolateral prefrontal cortices and amygdala and ventral striatum14-16. The most rostral CMA (CMAr) contains cells which respond when a reduction of reward is followed by a change in behavior1 but reinforcement related responses are also recorded in the ACCS rostral to
CMAr^{17-19}. This region of the macaque ACC is thought to be homologous with the human ACC region in which BOLD signal changes occur when human subjects perform error and conflict monitoring tasks\textsuperscript{2,20-24}. The ACC lesion therefore included the ACC region where reinforcement is most likely to influence action selection (Fig. 8). Post-mortem histology confirmed that the lesions were placed as intended but the lesion did not extend as far caudally in one animal (ACC\textsubscript{3}). It was noted that ACC\textsubscript{3} was the least impaired on the reversal tasks used in Experiment 1 but was sometimes the most impaired in the matching task in Experiment 2.

All surgery was carried out under sterile conditions with the aid of a binocular microscope. The lesion was made by aspiration with a fine gauge sucker. The cortex within the dorsal and ventral banks of the anterior cingulate sulcus (areas 24c, 24c') was removed, with the caudal limit of the lesion in the cingulate sulcus being an imaginary line drawn through the midpoint of the precentral dimple. The lesion extended rostrally for the full extent off the cingulate sulcus. Caudally strips of supporting tissue were spared underneath the ascending branches of the anterior cerebral artery using the method of Parker and Gaffan\textsuperscript{25}. This ensured the blood supply to the tissue dorsal and lateral to the lesion.

When the animals had completed their testing they were anesthetized with sodium pentobarbitone and perfused with 90% saline and 10% formalin. The brains were then removed and placed in 10% sucrose formalin until they sank. The brains were blocked in the coronal plane at the level of the most medial part of the central sulcus. Each brain was cut in 50 µm coronal sections. Every tenth section was retained for analysis and stained with
cresyl violet. Coronal sections through the frontal lobes, including the lesions, are shown in Figure 8. As can be observed, the lesions were made as intended. Both banks of the anterior cingulate sulcus were removed bilaterally but the anterior cingulate gyrus lying immediately ventral to the sulcus and the dorsal superior frontal gyrus remained intact. In one animal, ACCs3, the lesion did not extend as far caudally (Fig. 8).

REFERENCES


