

SCIENTIFIC REPORTS

**OPEN**

Performance in a GO/NOGO perceptual task reflects a balance between impulsive and instrumental components of behaviour

Received: 27 August 2015

Accepted: 17 May 2016

Published: 07 June 2016

A. Berditchevskaia, R. D. Cazé & S. R. Schultz

In recent years, simple GO/NOGO behavioural tasks have become popular due to the relative ease with which they can be combined with technologies such as *in vivo* multiphoton imaging. To date, it has been assumed that behavioural performance can be captured by the average performance across a session, however this neglects the effect of motivation on behaviour within individual sessions. We investigated the effect of motivation on mice performing a GO/NOGO visual discrimination task. Performance within a session tended to follow a stereotypical trajectory on a Receiver Operating Characteristic (ROC) chart, beginning with an over-motivated state with many false positives, and transitioning through a more or less optimal regime to end with a low hit rate after satiation. Our observations are reproduced by a new model, the Motivated Actor-Critic, introduced here. Our results suggest that standard measures of discriminability, obtained by averaging across a session, may significantly underestimate behavioural performance.

What is the impact of motivation on behaviour? Reinforcement learning theory assumes that an animal optimises behaviour according to the value placed on the goal of an action under different levels of deprivation^{1,2}. Beyond the value placed on the goal (*directional effect*), motivational state is also critically important in regulating the overall effort and rate of activity (*activating effect*) an animal engages in^{3,4}. The combination of behavioural training with sophisticated electrophysiological and imaging techniques is beginning to provide unprecedented insight into the functioning of neural circuits underpinning perceptually-driven decision making^{5,6}. A commonly employed paradigm for studying perceptual decisions is the two-category GO/NOGO task, where the animal performs a response to obtain a reward during a 'Go' stimulus, and needs to withhold the response for the 'NoGo' stimulus. This paradigm has been commonly used in recent years to draw inferences about sensory computations^{7–12}. Mice are readily able to perform such tasks, however natural biases¹³ can interfere with performance. Motivational influences on behaviour have also been documented both in rodents^{3,14} and other species¹⁵. These factors increase the difficulty of devising appropriate training protocols and can impact on the interpretation of results⁶.

The typical timeline for observation during simple decision-making tasks in modern neuroscience extends to hundreds of trials, thus capturing a range of motivational levels throughout a single behaviour session. While there have been mentions of changes in motivation within individual sessions previously^{16,17}, these effects are often ignored^{3,13} or factored out of analyses¹⁸. Here, we have analysed the effect of motivation on a GO/NOGO visual discrimination task at the single session level. To our knowledge, this is the first detailed exploration of these factors within the context of the GO/NOGO paradigm, although previously published work¹⁴ can be re-examined to support our conclusions. Our approach enables us to evaluate the role of motivation on classification performance, which has important implications for perceptual computation and decision-making.

Department of Bioengineering and Centre for Neurotechnology, Imperial College London, Exhibition Road, South Kensington, London SW7 2AZ, UK. Correspondence and requests for materials should be addressed to S.R.S. (email: s.schultz@imperial.ac.uk)

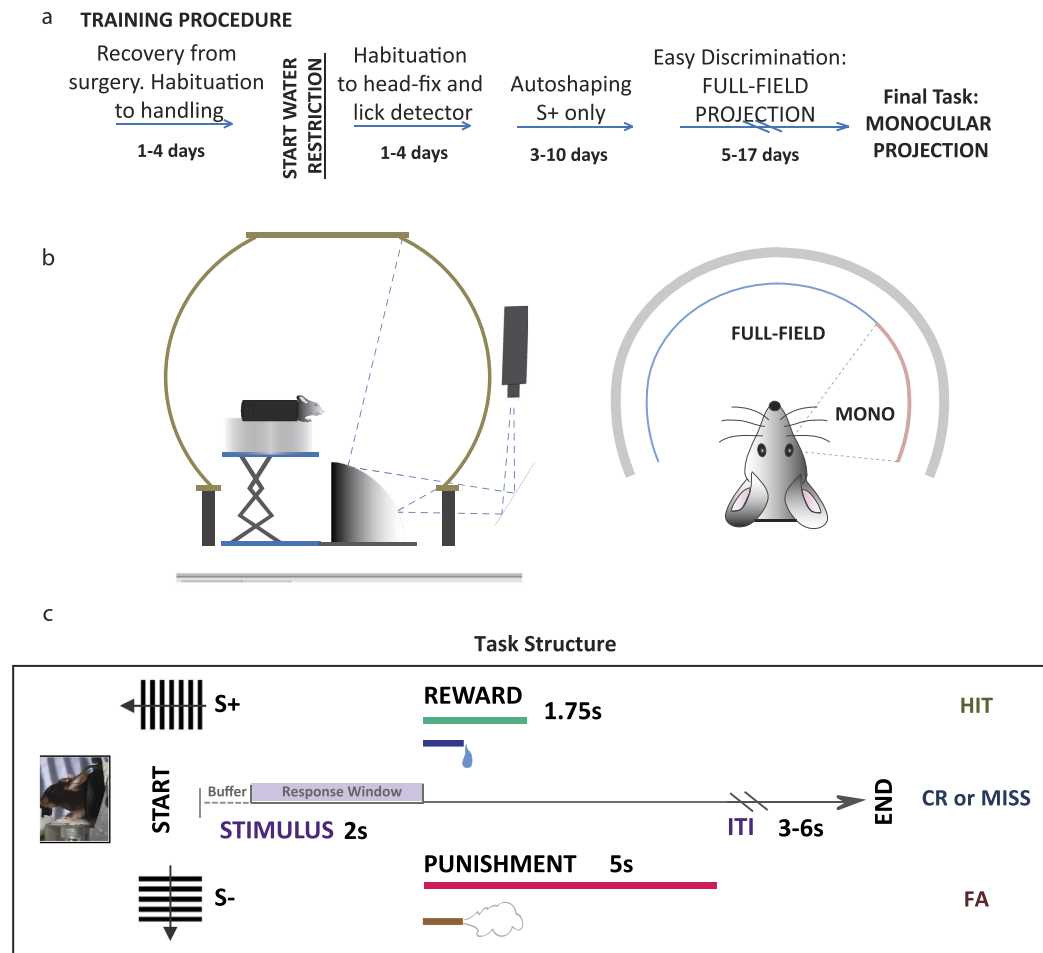


Figure 1. GO/NOGO visual discrimination task and training procedure. (a) Full behavioural training and experimental timeline. (b) All experiments were carried out under head-fixed conditions in an immersive dome environment. During initial training, stimuli were projected onto the entire inner surface of the dome (FULL-FIELD), before restriction to a monocular (MONO) projection region for the final task. (c) Task sequence showing all possible trial and outcome types with timing. Small rectangles (ie, orthogonal gratings, black, grey) show image projected onto screen during each trial state. A lick during the Stimulus (up to 2 s) state triggered immediate transition into a Reward (1.75 s) or Punishment (5 s) state, followed by a variable length (3–6 s) ITI state. S+ = Go stimulus, positively reinforced with water; S– = NoGo stimulus, negatively reinforced with airpuff; ITI = inter trial interval; buffer = 0.5 s at start of trial, licks during this period do not trigger a state change to Reward/Punishment state.

Our results demonstrate that motivational state changes throughout the course of a session in a quantifiable and consistent manner, as measured by response vigour (lick frequency). We show that an overly high level of motivation interferes with behaviour, causing poor performance until initial thirst is satisfied. Lick latency analysis allows further decomposition of the response into different behavioural components, aiding understanding of performance on the task. We conclude that while learning is likely to be reliant on maintenance of water restriction, it is more effective to keep animals in a more satiated state thereafter to gain accurate insights into behaviour.

Results

Overmotivation masks true sensitivity at the beginning of a session. We trained mice ($n = 15$) under water restriction (see Methods) to perform a visual discrimination task, using vertical (S+) and horizontal (S–) square-wave gratings initially projected to the entire visual field (full-field) and then the monocular region of the visual field (Fig. 1). After reaching criterion and showing a high level of discrimination during 3 consecutive sessions of the monocular behavioural task, we recorded the performance of the mice during a test session. Importantly, at this stage all animals were at a comparable level of performance. Figure 2a shows a typical example of a lick report from a post-criterion behavioural session. We define a single behaviour session as the 30–60 minute period where mice actively participated in the discrimination task. Despite slight differences in the absolute values of Hit and False Alarm (FA) rates, all mice showed the same pattern of changes to their licking behaviour over the course of a session. To analyse these changes, we split each session into three equal segments (Initial, Middle, Final). Our aim was to examine the impact of motivation on the expression of a learned instrumental

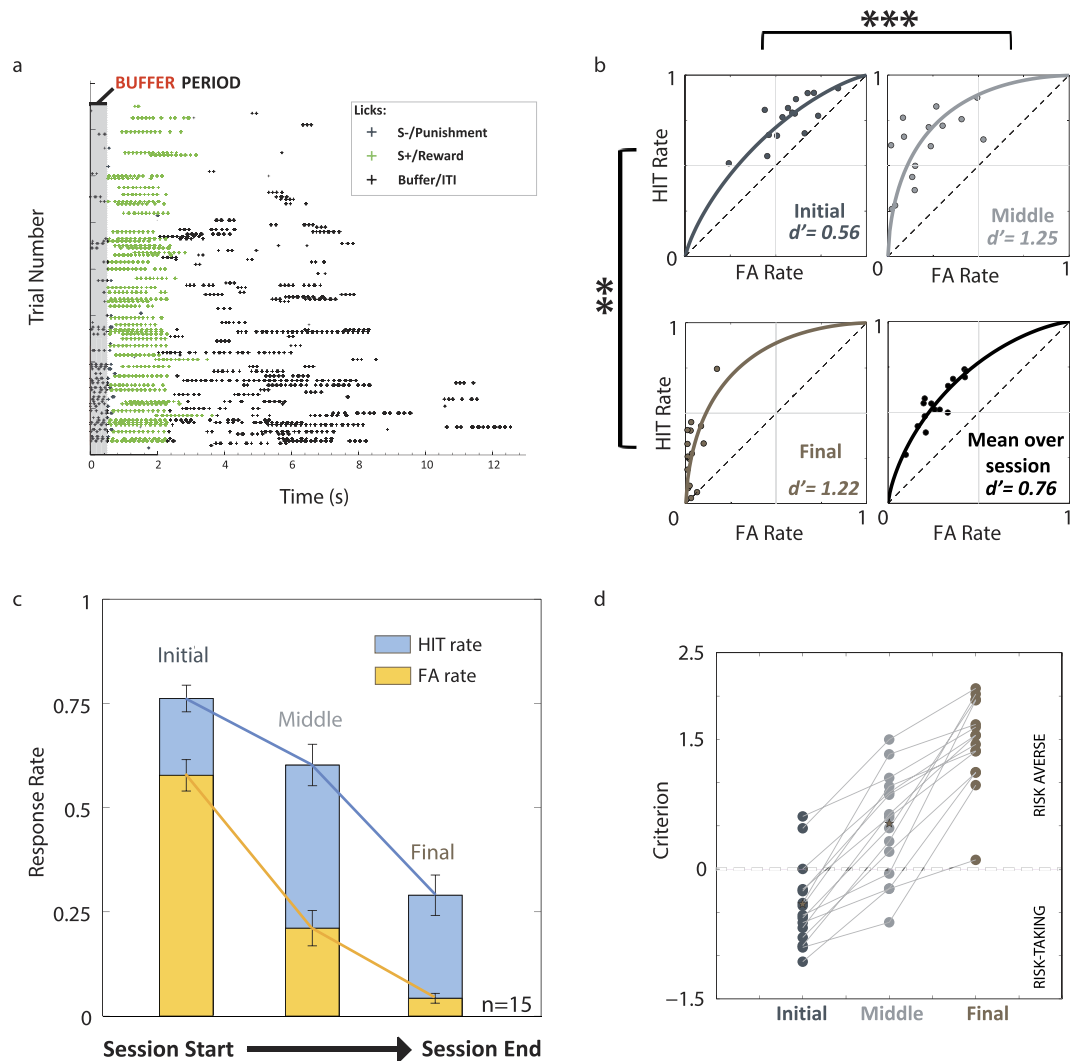


Figure 2. Discrimination sensitivity starts well below its final value, gradually improving as the level of deprivation decreases over the course of a session. (a) The behavioural readout from a typical session once criterion had been reached on the final version of the task. Time increases from bottom to top for the session and left to right for single trials. Licks per trial are tracked horizontally on each line, with each coloured cross representing a lick. Vertical orange lines provide a visual reference for subsequent lick latency ($LLat$) results. (b) Separate panels show group isosensitivity curves for the Initial, Middle and Final stages of the session. For each section of the session, we calculated the d' for each of the Hit/FA pairs within that section, and used the mean of these values to generate the curve. The bottom right panel shows the d' obtained from averaging over the whole session ($n = 15$, data represent mean \pm s.e.m. $^{*}p < 0.005$, $^{***}p < 0.0003$, Student's t -test). (c) Hit and FA rates over different stages of the testing session on the final task ($n = 15$).

behaviour on the timescale of a single session (within-session changes). We normalised the length of the three segments to the total number of trials, for each mouse, and used standard signal detection theory (SDT) measures of Hit and FA rates in each segment (see Methods).

At the start of the session mice performed with high values for both Hit and FA rates (Fig. 2c: $Hit_{Initial} = 0.76 \pm 0.03$, $FA_{Initial} = 0.58 (\pm 0.04)$). In the Middle segment, Hit rates remained high while FA rates dropped (Fig. 2c: $Hit_{Middle} = 0.60 \pm 0.05$, $FA_{Middle} = 0.21 \pm 0.04$). During the Final segment, both response rates decreased (Fig. 2c: $Hit_{Final} = 0.29 \pm 0.05$, $FA_{Final} = 0.04 \pm 0.01$). In other words, we observed that the Hit and FA rates of water-restricted mice showed dramatic variation over the course of a session, with corresponding changes in the SDT discriminability measure, d' . In Fig. 2, we show these changes in d' during the Initial, Middle and Final segments as performance curves in Receiver Operating Characteristic (ROC) space (see Methods). The sensitivity of the mice in the early part of the session was less than half of the value seen in the later segments of the session ($d'_{Initial} = 0.56 \pm 0.07$, $d'_{Middle} = 1.25 \pm 0.14$, $d'_{Final} = 1.22 \pm 0.16$). A comparison of the group values for each of the 3 segments showed that the sensitivity of the mice during the Initial segment was significantly different to the d' values that emerged in the later stages of the session (Student's t -test, $p < 0.0003$). The Final and Initial segment d' distributions were also significantly different (Student's t -test, $p < 0.005$) but no difference was found between

sensitivity values for the Middle and Final segments. This highlights that, despite differences in overall participation levels through the Middle and Final stages of the session, mice performed with consistent sensitivity. The bottom right panel of Fig. 2b (purple) demonstrates that mean d' across the session, the most commonly used behavioural sensitivity measure^{9,10,16}, substantially underestimates the true discrimination ability of the mice, with a d' of 0.76 ± 0.04 for the group. The use of means to assess both the within-session performance of individuals and group progression of learning can distort individual acquisition trends and can cause inaccurate estimates of performance¹⁹.

We conclude from this that stable discrimination sensitivity only emerges after the initial part of a behaviour session in water-restricted mice. The performance of the mice during this initial segment may skew the estimate of discrimination ability when mean discriminability is taken over the whole session. We hypothesise that the changes we observed in mouse behavioural performance were due to fluctuating motivational state. In the remainder of this paper we examine this hypothesis in more detail.

Behavioural performance can be explained by a motivation-augmented Actor-Critic model.

We first used the SDT response bias measure, criterion value, to assess motivation. In a learnt task modelled using SDT, the criterion designates the threshold at which an animal places the decision to “Go”. The negative or positive signage of the criterion reflects the direction of the bias away from a neutral baseline where the animal is equally likely to “Go” or not. Criterion value can thus be understood as reflecting the “directional” aspects of motivation. Figure 2d shows that the criterion value steadily increased over the course of the behaviour session ($C_{\text{initial}} = -0.42 (\pm 0.12)$, $C_{\text{middle}} = 0.51 (\pm 0.16)$, $C_{\text{final}} = 1.46 (\pm 0.13)$), changing from a low criterion (risk-taking) approach to a higher and more conservative (risk-averse) decision policy.

We next explored a modelling approach to test whether a “motivation” variable could account for the stereotyped behaviour we saw during the task. Actor-Critic models are commonly used to describe reinforcement learning in simple two-alternative and operant behavioural tasks^{2,20}. We compared the performance predicted by two models. The first (AC) was based on the classic Actor-Critic approach with Q-learning (see Methods for details). The second model (MAC) shown in Fig. 3a maintains the classic Q-learning framework and initialisation parameters, but also takes motivational state into account. Motivational state is incorporated through a “thirst” variable, which adjusts the Q-values (see Methods) of the Actor-Critic over the course of the behaviour session. The top panels in Fig. 3a show the proportion of Hits/FAs and ROC performance predicted by 15 iterations of the models. They predict the performance trajectory that would be seen in animals that have been trained on a GO/NOGO task. We used the baseline assumption that mice had prior knowledge of the task contingencies during the test session, so that S+ stimuli were valued more highly than the S- stimuli (ie, that mice had attained criterion performance) and initialised the starting Q-values accordingly. We estimated the goodness of the predictions made by the models by calculating the Euclidean distance between the experimental data and the simulated results. A perfect match gives a value of 0, while the furthest possible separation gives a value of 2. The performance predicted by the MAC model (distance = 0.15) more closely reflects the behaviour recorded during the experiments than the classic Q-learner (distance = 0.78). Inclusion of a motivation variable reproduced the lower d' observed during the Initial segment of the session, and gradually changed towards a consistent discrimination level in the Middle and Final segments of the session. The model was also able to capture the change in task participation whilst maintaining discrimination sensitivity for the Middle and Final segments, as seen in the experimental data.

We devised a ROC space classification scheme to assign motivational state (Fig. 3b), guided by the ROC literature²¹. Accordingly, when we examined the Hit/FA pairs within ROC space from the experimental test session (Fig. 3a: Far Right, bottom panel), we found that behaviour early in the session mostly fell within the section reflecting “high or over-motivation” (red shading) while the Hit/FA pairs from the later stages of the session fell within the purple and green shading regions (“unbiased” and “low motivation”, respectively).

An alternative explanation might posit that mice spend the Initial segment of the session re-learning the task contingencies, until they are able to perform again at a high standard. We compared the typical motivational variation shown by mice during the learning stage of the behavioural training to get a better understanding of the interaction of these two processes. Figure 3c shows 6 consecutive training sessions over which the d' improved to fulfil the criterion level of performance (First session $d' = 0.25$, Final session $d' = 1.62$). Unlike the learnt stage of behaviour where performance followed a typical trajectory, the performance during the learning phase showed substantial variation in which portion of motivation ROC space the Hit/FA pairs occupied over the sessions. Hit/FA rate values during learning were rarely found in the extreme upper right of the ROC space (red shading). Thus the deprivation regime did not result in overmotivated performance during the learning phase of behavioural training.

Our modelling results thus strongly support a causal role for motivational state in determining the performance trajectory observed on a GO/NOGO task within individual behaviour sessions. In the next section we seek to define these changes in a more quantitative way by considering the *activating* aspects of motivation.

Quantitative measures of response vigour reliably demonstrate motivational changes within individual sessions.

Response vigour, or the rate at which instrumental actions are undertaken, has recently been used as a readout of motivational state^{3,4}. We quantified response vigour using three complementary measures: lick frequency, lick ‘efficiency’ (or relative licking) and lick latency (Table 1) to characterise changes to motivational state. Licks occurring at specific, well-defined points within a trial can be indicative of different motivational factors and/or changes to behavioural policy when examined in relation to overall lick rates.

We measured the licking frequency ($LFreq$) of the mice during the criterion sessions by counting the absolute number of licks that each mouse performed per trail normalised to the trial length. We then calculated the mean $LFreq$ for the ‘Initial’, ‘Middle’ and ‘Final’ segments for both individual mice and the group (Fig. 4a). The variability

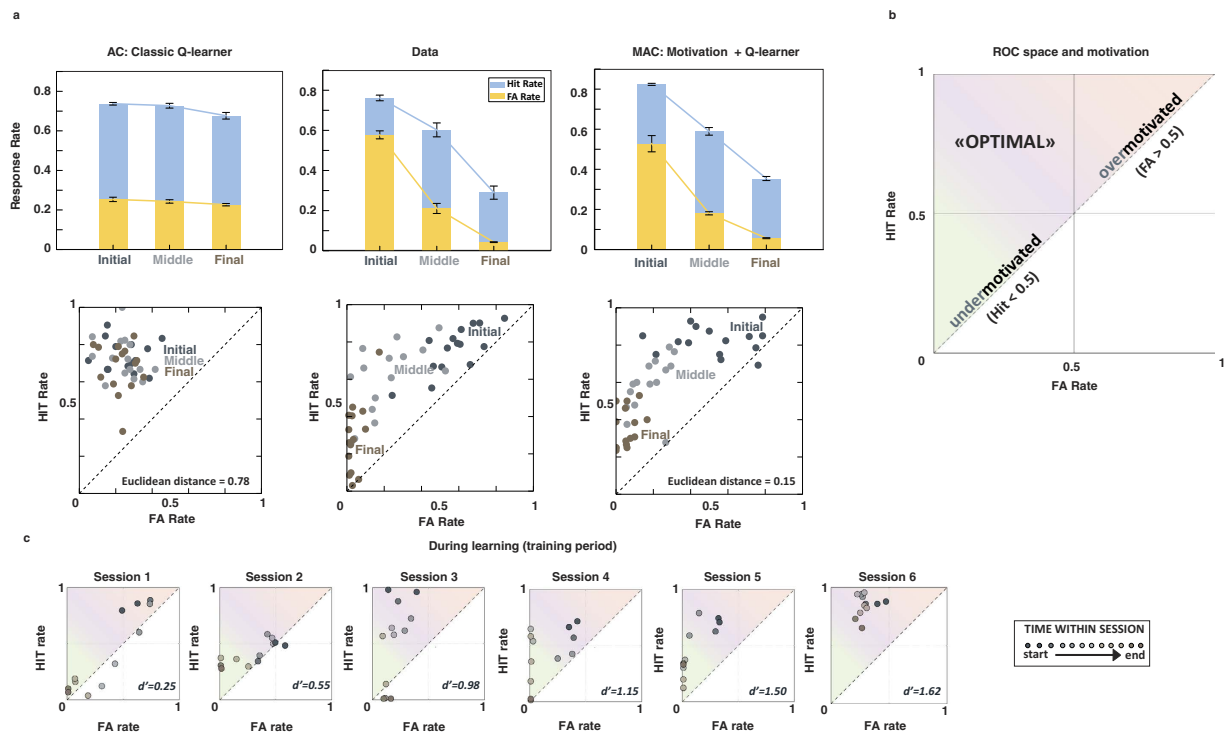


Figure 3. Introducing a motivation variable into the Actor-Critic (AC) model can account for our observations. (a) Introducing a motivational factor into the classic Q-learner framework results in performance predictions similar to the changes we observed in the experimental results (Far right panel). Top: while the AC model showed relatively stable sensitivity throughout the entire session, the MAC model recreated the characteristic differences between HIT and FA rates. Bottom: HIT/FA pairs projected into ROC space demonstrate tight clustering for the classic model but spread similar to the experimental results for the motivation model. Modeled data, $n = 15$ iterations, initialized to Q-values of 1 under the assumption that task contingencies had been learnt (see Methods for details). MAC performed better than AC as measured by the Euclidean distance between the values predicted by the model and the experimental results (AC = 0.78; MAC = 0.15). (b) Illustrative ROC space: red region shows overmotivated performance, green undermotivated, and purple freedom from extremes of motivational bias. (c) The performance of a typical mouse (experimental data) over 6 successive sessions during a training period with full-field stimulation. Each datapoint corresponds to the Hit/FA pair from a given time-point within the session, projected into ROC space. The points move away from the $d' = 0$ diagonal with learning. The session trajectory varies considerably between sessions, but Hit/FA pairs rarely fall in the overmotivated (red) region of ROC space.

Lick Frequency (<i>LFreq</i>)	Number of licks per trial, normalised to the total length of the trial.
Relative Lick Rate (<i>LRel</i>)	Number of licks within defined sections of a trial (eg. Buffer, ITI, referred to as time-periods hereafter) normalised to the lick frequency (<i>LFreq</i>).
Lick Latency (<i>LLat</i>)	Exact timing of first lick within a trial relative to trial onset.

Table 1. Quantitative measures of response vigour used to measure changes in motivation.

of absolute *LFreq* values within the group was large for the ‘Initial’ segment. We found no systematic trend between the size of these values and deprivation parameters (results not shown) suggesting that this variation was more indicative of an intrinsic baseline of individual animals. Despite this variability all mice followed a group trend, with *LFreq* decreasing over the course of the session ($LFreq_{Initial} = 4.50 \pm 0.46$, $LFreq_{Middle} = 3.22 \pm 0.42$, $LFreq_{Final} = 1.01 \pm 0.10$). This is consistent with the motivational state making a transition between over and undermotivation, where under high motivation animals invigorate an instrumental action (activating effect) that leads to a desirable outcome (directional effect)⁴.

We also calculated the relative lick rate (*LRel*) during different time-periods, i.e. the number of licks occurring during a particular time-period divided by the total number of licks in the trial. By this definition, a highly efficient animal would concentrate licking activity into time-periods when the action could lead to triggering or collecting the water and not engage in adjunctive licking in other time-periods. Highly motivated but inefficient animals, on the other hand, may show a more general increase in instrumental response²² by increasing licking indiscriminately in all time-periods.

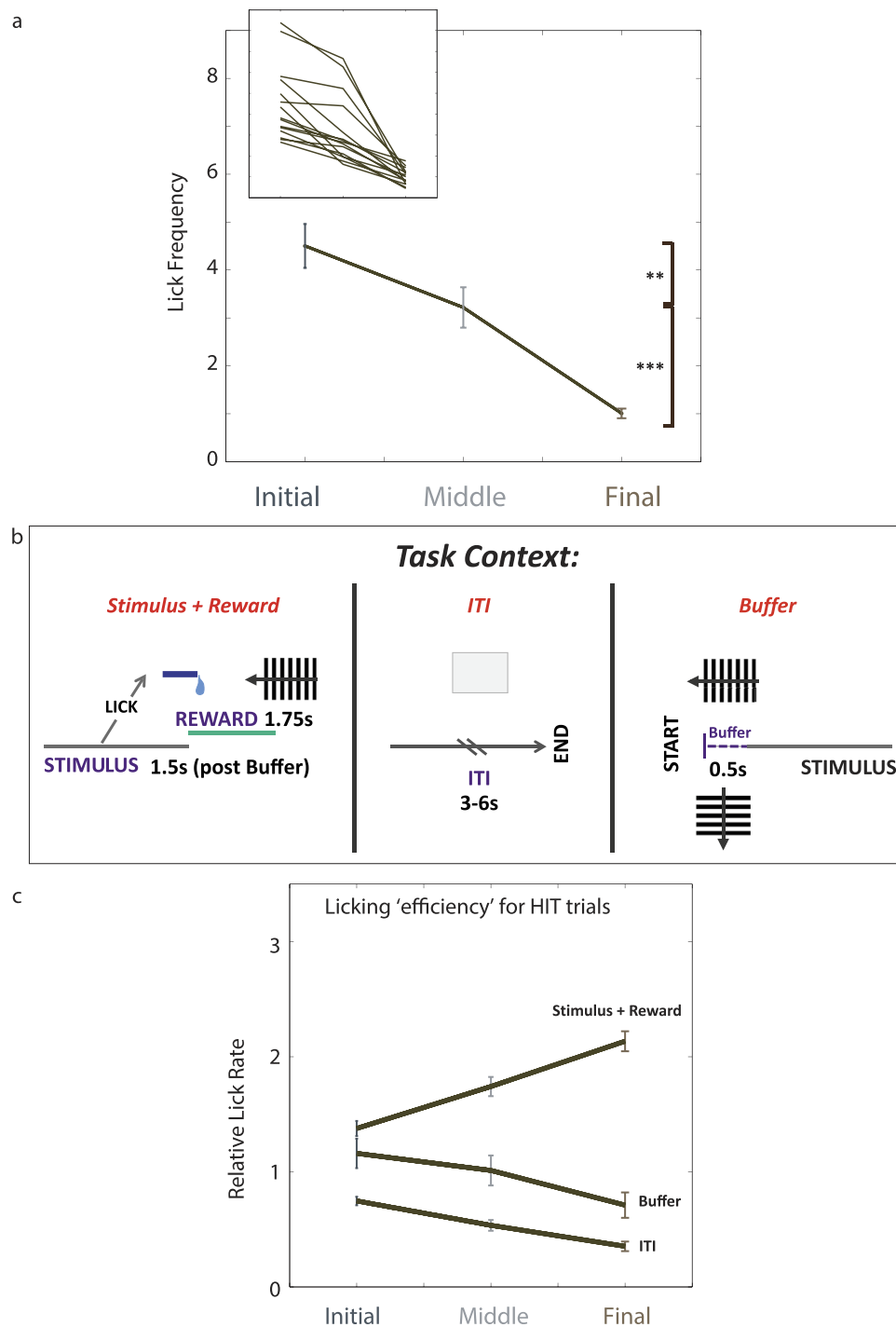


Figure 4. Variation of the lick frequency depends on session progression and task context. (a) The rate of the licking action ($Lfreq$) per trial decreases over the session. ($n = 15$, data represent mean \pm s.e.m. $**p < 0.0056$, $***p < 1.05 \times 10^{-9}$, repeated measure ANOVA with post-hoc comparisons). Inset: $Lfreq$ for individual mice. Background shading on the main panel corresponds to the 3 different segments of the session (dark grey = Initial, light grey = Middle, brown = Final). (b) The lick rate for different periods of a trial normalised to the overall $Lfreq$ (ie., $LRel$ = Relative lick rate). Upper panel shows a schematic of the periods within the trial over which the $LRel$ values were calculated: Hits, ITI, Buffer. All $LRel$ values were calculated using the trials with a “Hit” outcome. (c) $LRel$ for Hit, Buffer and ITI time periods over the course of a session. $LRel$ for the Buffer and Stimulus + Reward periods are not significantly different during the Initial segment ($p = 0.75$, repeated measure ANOVA with post-hoc comparisons).

For each stage, we compared how $LRel$ changed over the session for rewarding time-periods versus those that carried no instrumental gain (Stimulus + Reward vs Buffer or inter-trial interval (ITI) for “Hit” trials, Fig. 4b).

Figure 4c shows that at the start of the session, the Buffer period (Initial: 1.16 ± 0.13) had similar $LRel$ values to the Stimulus + Reward (Initial: 1.38 ± 0.07) period, which were not significantly different ($p = 0.75$). Over the course of the session, these two values separated and the animals licked much more efficiently. This is shown in the significantly higher $LRel$ values for the Stimulus + Reward condition than the time-periods where licking held no behavioural advantage ($LRel_{Final}$), indicating that animals limited their licking to the instrumentally rewarding periods of the trial.

Comparisons with the time-periods where licking did not lead to a water reinforcement highlighted an interesting difference between the two conditions. We noticed that while $LRel$ values during the “hits” time-periods gradually increased, $LRel$ in the Buffer period decreased over the course of the session (Initial: 1.16 ± 0.13 , Middle: 1.01 ± 0.13 , Final: 0.62 ± 0.11). In contrast, the $LRel$ values for the ITI period showed more stability over the session and less variability between mice (Initial: $0.75(\pm 0.04)$, Middle: $0.53(\pm 0.05)$, Final: $0.35(\pm 0.04)$), suggesting they represented a baseline licking tendency while Buffer licking was more prone to motivational effects.

The response vigour measures of $LFreq$ and $LRel$ confirmed that motivation changed in a quantifiable, consistent pattern between all animals to provide a readout of motivational level during a session. Next, we sought to determine whether an analysis of lick timing might allow us to gain a more nuanced insight into the behavioural strategy governing the “activating” licking rates.

First-lick latencies reveal the balance of Impulsive and Instrumental influences on behaviour.

While the licking frequency and the distribution of lick rates within specific trial sections reflect the “activation” aspects of motivation, the latency of an instrumental response can probe more deeply into the decision making process. Latencies have been used as indicators of behavioural impulsivity^{23–25} and the ability to inhibit inappropriate responses in decision-making tasks²⁴.

We calculated first lick peri-stimulus-histograms (PSTHs) pooling latencies from all trial outcomes to identify the general trends of lick timing for a behaviour session. Figure 5a shows these group latencies for the monocular test session. We decomposed the response into two episodes: a primary lick response, defined by a sharp early peak immediately after stimulus onset, and a broad, multi-peaked secondary response which occurred at the end of the Buffer period, clustered around the 0.5 s time-bin. The sharpness of the primary peak indicates a more stereotyped behaviour that is inflexible from trial to trial, typically seen in Pavlovian and over-trained habitual instrumental responses due to the availability of precise temporal information. To avoid habitual responses, animals were taken off daily training regimens immediately after attaining criterion and were kept on a reduced “reminder” regimen until testing. Comparisons with previous cohorts who formed habits after training for longer periods on easier stimuli confirmed the success of this strategy (data not shown). In the case of rodent lick-based tasks, licks can represent both an active indication of choice or the manifestation of an acquired ‘adjunctive’ consummatory behaviour, or Pavlovian response^{26–29}. Our use of the term Impulsive (which we could also refer to as Pavlovian within the scope of anticipatory and consummatory behaviours that are directionally aligned towards an appetitive goal), is intended to capture the broader relevance of these results to other experimental paradigms where choice and consumption responses are differentiated.

The spread in lick latency values seen during the secondary response on the other hand reflects more flexible decision-making. In this case, an animal’s GO/NOGO choice is based on the prior knowledge of the task contingencies in view of the trial-specific stimulus (S+ or S−) that is shown. We therefore describe the primary response as an “Impulsive” contribution to the behaviour and the secondary response as the “Instrumentally Driven” component. As the main features of the PSTHs were observed within the 0–2 sec time range (ie, during the Stimulus state), we focussed on comparing the latency profiles for the two outcomes where the animals made a ‘Go decision’ (Hits for S+ trials and FA for S− trials) and therefore necessarily made their first lick before 2 sec.

In order to relate the differences to our other performance indicators, we split the sessions into 3 equal segments as with the previous analyses (Fig. 5). We extracted all Hit and FA trial latencies from within each segment and used a time range of between 0 and 2 seconds to construct PSTHs for the lick latencies. The lick latency PSTHs for both the Hit and FA trials changed considerably over the course of a session (Fig. 5b). The difference between Hit and FA licks (Fig. 5c) reveal that responses during these two trial types increasingly diverge in the secondary response (post buffer period), towards the end of a session. There is little difference between the primary response for both trial types, which might be expected given that licks during the buffer period had no consequence. This implies that the lick responses of the mice during the earliest part of the session show little difference irrespective of the stimulus presented (dark grey line in Fig. 5c) and that discriminative licking patterns emerge later in the session (lighter grey lines, Fig. 5c).

Peaks in the PSTH indicate the time at which it was most common for mice to register their initial lick, relative to the maximum response. The latency of the Impulsive response remained stable at 0.04 sec throughout the session. For the Instrumental component, latencies changed over the course of the session from 0.52 sec to 0.56/0.84 sec for Hit trials during the Initial, Middle and Final stages respectively. The differences in the timing of the two response components reinforces the suggestion that they are supported by different underlying processes, with the primary component showing more stereotyped, automated timing, while the secondary component remains prone to variation, suggesting a flexible decision making process. Interestingly, the latency of the peaks for the secondary component of the FA trials showed two widely separated modal peaks, at 0.48 sec and 1.44 sec. These groupings suggest two underlying error types, ‘Failure to Stop’, and ‘Failure to Wait’, which were previously reported by Mayse *et al.*²⁴. Such nuanced errors are indicative of a cost/benefit judgement of performing an action. The emergence of these lick latencies towards the end of the task reinforces the idea that mice start to rely on a behavioural strategy to solve the task in a goal directed manner.

We calculated the ratio between the peaks of the secondary and primary responses to quantify which grouping (Impulsive or Instrumental) contributed more to first-lick responses. To estimate the ratio we first quantified the

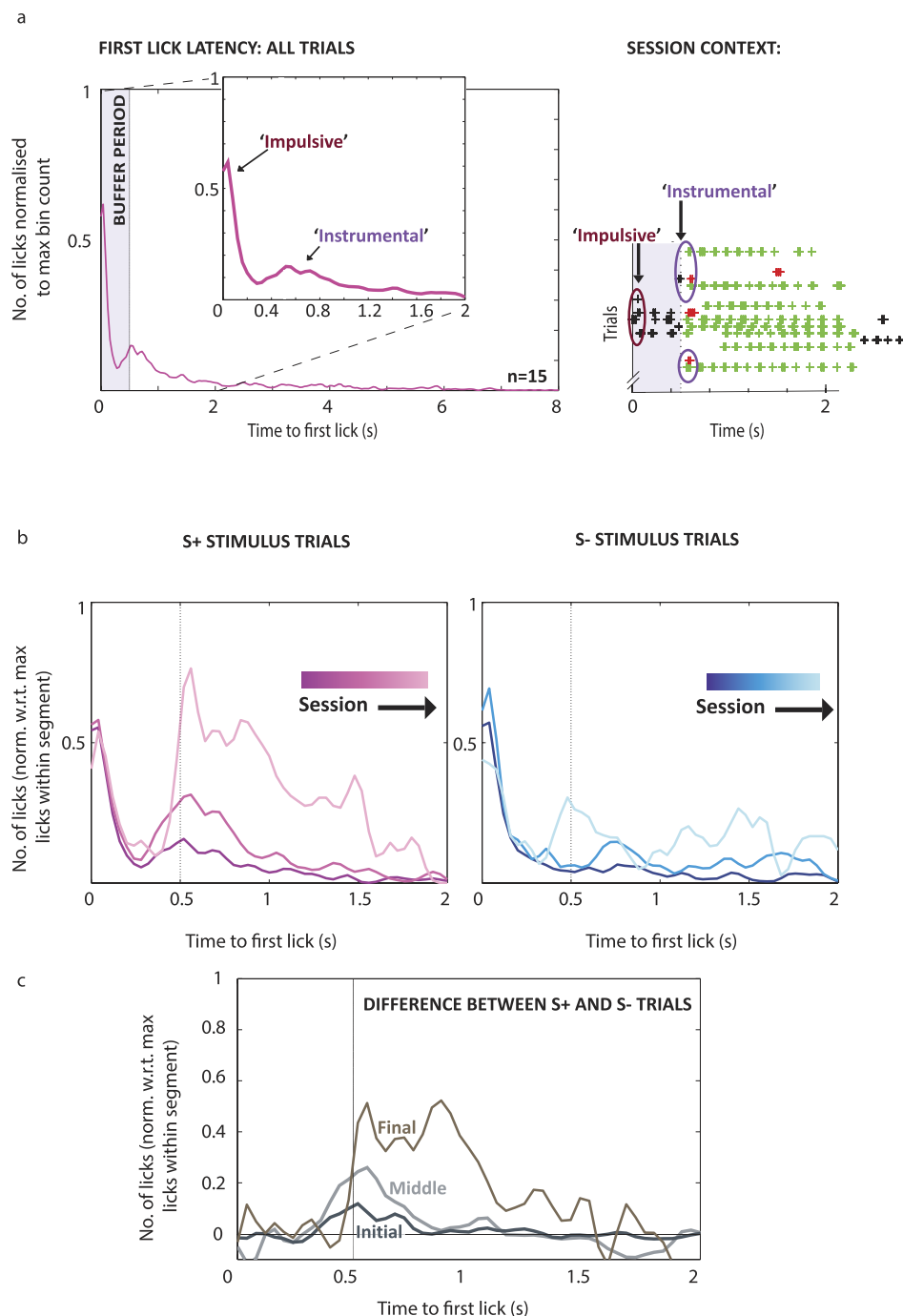


Figure 5. First-lick latencies reveal Instrumental and Pavlovian components of behaviour. (a) Left, peri-stimulus time histogram (PSTH) of first licks collapsed over all outcomes for the whole session for all mice ($n = 15$). The distribution is bimodal with an immediate sharp peak (the Pavlovian component) following the trial onset, followed by a more dispersed secondary peak (the Instrumental component) at the end of the Buffer period (0.5 s after trial onset). Inset: Enlarged view of the first-lick responses over the first 2 s of the trials, providing the lick profile for just FA and Hit trials. Right, Part of a session report highlights some of the licks that contribute to the two prominent peaks seen in the lick profile. (b) The changes to first-lick profiles over the course of a session, calculated for the Initial, Middle and Final segments for S+ and S- trials, left and right respectively. (Initial = dark line, Final = pale line) Dashed arrows highlight the emergent changes to the post-Buffer lick profile towards the end of the session. The proportion of first licks occurring as part of an Instrumental response increases over the course of the session. (c) The differences between the lick profiles for S+ and S- trials for the Initial (dark grey), Middle (light grey) and Final (brown) segments obtained by subtracting the S- lick profiles from the S+ (raw values were used). PSTH values were normalised to the time-bin with the highest number of first licks over the whole session (a) or for that particular session segment (b,c). All PSTHs were convolved with a Gaussian filter for smoothing, hence capping of highest peaks below 1 (see Methods for detail).

licks within the Impulsive and Early Instrumental components by calculating the area under the curve for the lick PSTH between the times 0–0.2 sec and 0.4–0.6 sec respectively. The Impulsive component of the behaviour dominated the Initial segment for all mice, and this was reflected in the ratio of 2.88 for the peaks of the two response types (~3 Impulsive first licks for every 1 Instrumental first lick). By the Middle segment, the ratio shifted to 1.39 reflecting the shift to goal-directed behaviour. Nearer to the end of the session, Instrumental licking behaviour began to dominate (ratio = 0.76). Over the course of the session, the Impulsive component gradually became less prominent, and the most frequently occurring first lick values belonged to the second, Instrumental grouping. Although this result might imply that the behavioural strategy is only optimised in the Final segment, an examination of the participation levels of the animals suggests otherwise. Namely, while the separation between Hit and FA values remains large and gives an appropriately high sensitivity value in the Final segment, the high proportion of Misses (71% (± 0.05)) demonstrates that the mice were no longer well engaged in the task.

Overall, our analysis of the lick latency distribution revealed that as motivation changes, the balance of the behavioural processes contributing to decision-making shifts between impulsive and Instrumentally determined actions. Thus a highly motivated state is dominated by an Impulsive component, which obscures the optimal performance that Instrumental actions would otherwise result in. As a final test that high levels of motivation are in fact detrimental to performance on a discrimination task, we compared the performance of a mouse on consecutive days under satiation and deprivation.

Deprivation may facilitate learning but prevents optimal performance once a task is learnt.

Both the simulated results of the Actor-Critic “learning” (Fig. 3b) and the instability of the ROC trajectories during learning (Fig. 3c) suggested it was unlikely that the stereotypical poor performance of the mice during the Initial phase could be explained by learning. Furthermore, the Actor-Critic model predicted that performance should be stable over the course of a behaviour session if motivational state does not play a determining role (clustering of Hit/FA pairs in top panel of Fig. 3b). To further test our hypothesis of the behaviour being motivation dependent, we compared the performance on consecutive days under a state of relative satiation on day 1 and normal deprivation on day 2. Figure 6 gives a typical example of these conditions for a single mouse to demonstrate the persistent vulnerability of Instrumental behaviour to “masking” by motivation irrespective of prior performance.

Once criterion was reached, we established satiation by pre-feeding. Pre-feeding has previously been described as an effective short-term manipulation of motivation for reward incentive and responding^{30,31}. During the first session from the ‘Satiated’ state (Fig. 6a, left) we found that the Instrumental component of the behaviour remained stable, with high d' values throughout ($d' = 1.90$ – 2.01). In contrast, when we returned to normal procedures for the session on the following day (Fig. 6a, right), we confirmed that over-motivation affected even the best performing animals during the ‘Initial’ segment of the session. There was a negligible difference between the Hit and FA rates at the start of the session (Fig. 6a, right: Initial), indicating that the animal was responding without any differentiation between the two stimuli ($d'_{\text{Initial}} = 0.04$, $\text{Hit}_{\text{Initial}} = 0.78$, $\text{FA}_{\text{Initial}} = 0.77$). Once the influence of motivation normalised, the animal returned to levels of response and performance indicators of a level that had been previously demonstrated ($d'_{\text{Middle}} = 1.54$ $d'_{\text{Final}} = 2.56$). We also assessed the lick-based and response bias measures for the two sessions. As expected, the ‘Deprived’ session (Fig. 6c) followed the same trajectories as described in the group results (Fig. 4a). The session under relative satiation, in contrast, showed remarkable stability over the lick-based indicators ($L\text{Freq}$ and $L\text{Rel}$) indicating that the motivational influences described for the cohort were much reduced or completely absent. Specifically, the slope of the decrease followed by the $L\text{Freq}$ over the 3 segments was minimal (Fig. 6c), there was no crossover for $L\text{Rel}_{\text{Hit}}$ and $L\text{Rel}_{\text{BUFF}}$ (Fig. 6b).

We also explored the relative contributions of the Impulsive and Instrumental components to the first-lick latencies for ‘Satiated’ and ‘Deprived’ sessions. The profiles of the PSTHs (for the whole session) were noticeably different (Student’s t-test: $p < 0.01$) despite the relatively low number of trials contributing to the statistics. An examination of the general trends revealed that for ‘Satiated’ states, the majority of licks occurred in time bins after 0.5 sec and were likely to be goal-directed (main panel and inset, Fig. 6d). This second response was multi-peaked with the majority of the maxima occurring between 0.7–1 sec. The ‘Deprived’ session, showed a considerably different pattern on the other hand (grey shading in Fig. 6d highlights this difference between the two sessions). As with the ‘Satiated’ session, there was some evidence of a secondary response grouping close to 0.5 sec. These latencies correspond to the responses of the mouse that occurred in the Final segments towards the end of the session (data not shown). The primary response directly after trial onset, however, holds proportionally more of the latency values from the whole session. This grouping captures the Impulsive licking that the animals engage in when overmotivated. The inset of Fig. 6d shows the absolute heights of the two response groupings. The ‘Satiated’ session had a dominant secondary response and a reduced Impulsive component. During ‘Deprived’ conditions, on the other hand, the contribution of Impulsive licking was proportionally much greater than the Instrumental.

Discussion/Conclusion

We have shown that overmotivation can impair performance on a learnt visual discrimination task and, if performance is assessed by an average across a session, the true sensitivity of mice can be masked. The consistency of the changes to performance across subjects demonstrates that motivation affects behaviour in a stereotypical and predictable manner.

Our demonstration of the performance of a trained animal over consecutive days with ‘Satiated’ and ‘Deprived’ starting states respectively, shows that degradation in discrimination performance at the start of sessions is caused by overmotivation. Our results suggest that while deprivation may be necessary to drive initial learning during the training phase of a behavioural task, it can be detrimental to performance once a task is learnt. A detailed

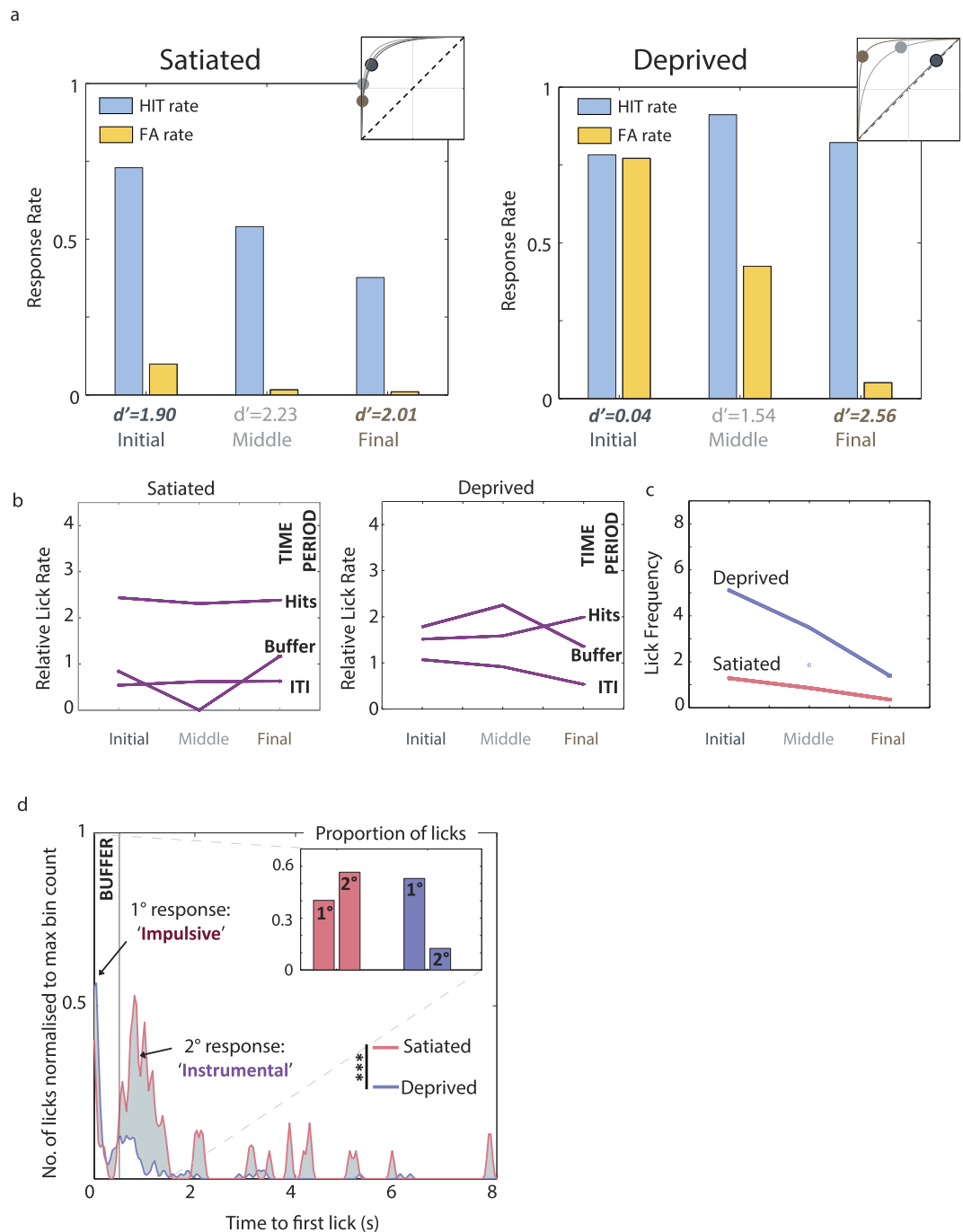


Figure 6. Satiated states result in better performance once a task has been learnt. (a) ROC plots for consecutive sessions by one mouse in the absence of (Satiated, 114 trials) and vulnerable to changes in motivation (Deprived, 168 trials). A brief free-water period prior to the Satiated session resulted in a stable instrumental performance throughout. Returning to the usual water restriction regimen for the Deprived test session on the following day reinstated the vulnerability of instrumental responding to motivational influence. (b) Relative lick rate (*LRel*) for 3 different time periods within trials are shown for this mouse: Hits, Buffer, ITI (see Fig. 4b). Left, during the Satiated session, lick efficiency for Hits remained high throughout the session. Right, Buffer *LRel* remained during the Deprived state until the Final segment of the session. (c) The frequency of licks (*LFreq*) for the Deprived state followed the typical trajectory shown by the group, gradually decreasing over the course of the session while the Satiated *LFreq* showed little difference between the 3 session segments. (d) First-lick latencies (*LLat*) over all outcomes for the whole session. The majority of the first licks for the Satiated session occurred in the post-Buffer time period, suggesting the dominance of goal-directed responding. In contrast, the latency profile of the Deprived session showed a typical sharp peak immediately after trial onset and a much smaller secondary peak centered on 0.5 s. (Satiated vs Deprived *LLat* profile, *** $p < 0.0001$, Student's *t*-test). 1°: primary, 2°: secondary peak. Inset: the difference between the Pavlovian and Instrumental components for both sessions.

Parameter	Value
P	0.5
N	150
R	15
T	1
Q_{mit}	0
α	0.1
M_0	1
α_m	0.06

Table 2. Model Parameters.

characterisation of changes to the distribution of first-lick latencies gave further insight into the cause of the variation to performance. The bimodal distribution of first-licks suggested that this variation was due to differences in the relative contribution of Pavlovian and Instrumental factors in states of high versus low motivation. At the start of a session, when animals are overmotivated, their actions are almost entirely driven by Pavlovian mechanisms, with only a small Instrumental component. In contrast, at the end of the session, the balance shifts, with most licks falling into the goal-directed grouping, but engagement with the task often at a relatively low level. Optimal performance is realised during the middle part of a behaviour session, where both Pavlovian and Instrumental responses are engaged to a high level, resulting in high participation and good discrimination. We conclude that the best performance on discrimination tasks requires a combination of these components.

By comparing the average d' for the session with a segmented (within-session) analysis (Fig. 2b) we were able to show that the former, which is the typical performance indicator used in behavioural neuroscience^{8,9}, misrepresents the true abilities of mice on a GO/NOGO task. Averaging the d' over the session underestimates the mouse's discrimination capability, due primarily to reduced sensitivity in the initial part of a behaviour session. Instead, we recommend fitting d' to data collected across the session, excluding the early (over-motivated) trials, by tracking the animal's progress across the ROC throughout the session in a sliding or shifting window. This builds on previous suggestions of rank-biserial correlation between trial number and responses³² and modelling by Wichmann and colleagues (Wichmann and Hill 2001; Fründ *et al.*) that suggested statistical methods to overcome non-stationarity.

A comparison of 2 different modifications of the popular Actor-Critic model^{2,3,20} demonstrated that the performance curves observed in our behavioural data were most closely replicated by introducing a motivation variable into the framework. This further supports the idea that the “masking” of high discrimination ability is caused by short-term motivational influences, rather than the persistence of a dynamic learning phase throughout the testing period. Guided by the previous work of^{3,4}, we quantified the activational aspect to motivation with the licking frequency, showing a linear decrease over the course of the session. We also showed that licking became differentially more focused towards the rewarding times within the trial (LRe^I_{Hits} vs LRe^I_{Buff} comparison, as has previously been suggested with a different behavioural paradigm with rodents¹⁴. We consider it likely that a Pavlovian response is enhanced by the display of the visual stimulus at the start of trials, as it has previously been suggested that Pavlovian responses can undergo a generalisation effect that dissociates them from the specific details of a stimulus²⁹. Therefore, licking under Pavlovian influence should be disproportionately increased during the Buffer for all stimuli in the early stages of a session. Towards the end of a session, actions (licks) are deployed in a more strategic manner, supporting our claim that the balance of behavioural strategy shifts towards goal-directed as the trial progresses.

As such tasks become more widespread, the behavioural analyses described here may be a useful resource for neurophysiologists investigating the neurobiological basis of behaviour. Our results may help with task design and appropriate analysis constraints. A growing number of groups are using GO/NOGO paradigms with SDT performance measures. Such studies use neurometric readout and psychophysical performance to interpret perceptual thresholds and sensory coding^{6,32–34}. Our behavioural results show that the ongoing balance of Pavlovian and Instrumental factors is likely to shape the extent to which an animal relies on sensory information in order to guide its behaviour¹⁷. Licking behaviour is but one example of a commonly employed instrumental action that suffers from the difficulty of separation between Pavlovian and instrumental components, showing the wider relevance of our results. Approach behaviour towards appetitive reinforcers can interfere with the interpretation of paradigms that use spatially separated levers and/or maze-choice arms in operant chambers for freely moving animals³⁵ (Goltstein *et al.*). A recent study using human subjects has also described similar Pavlovian to instrumental imbalance in decision-making (Chumbley *et al.*). We hope that these results can guide others towards the accurate classification of behavioural components in order to be confident in their interpretations of neurometric readout. By using quantifiable measures of the “activating” aspects of motivation, researchers can set thresholds and decide on reproducible and consistent criteria, that best suit the question they are investigating to systematically compare behaviour between groups or within individual animals.

Future work might expand on the measures presented by our group to consider alternative readouts of motivational state. This could include further analysis of session data, for instance by the use of breakpoint analysis. From the perspective of behaviour as a whole-body process, this might also involve monitoring physiological variables such as pupil diameter, muscle tension and heart rate, in order to optimise the behavioural regime for each subject and session.

Methods

Subjects. All experiments used female C57BL/6 (Harlan or Charles River Laboratories) mice that were 4–6 weeks of age at the time of headplate surgery ($n = 15$). Animals were group housed on a reversed 12 hour light-dark cycle and experiments were carried out in the dark phase. Mice were weighed and handled daily throughout water restriction. Food was available *ad libitum* and animals received daily water supplementation either during behavioural training or in their home cage on rest days to maintain weight above 85% of the starting value. All procedures were carried out in accordance with UK Home Office Animal Care Guidelines, and approved by the Imperial College Animal Welfare and Ethical Review Body under Project License 70/7355.

Headplate implantation. Prior to water restriction and behavioural training all animals underwent a headplate implantation surgery. All surgeries were carried out using aseptic technique under isoflurane anaesthesia. Throughout surgery, body temperature was monitored with a rectal probe. Immediately after induction of anaesthesia, 0.1 ml of carprofen (1 mg/ml, Rimadyl) was injected subcutaneously. After exposing the skull, a drop of 5% hydrogen peroxide was used to clean the skull surface. Custom designed metal headplates were attached using Histoacryl (TissueSeal LLC, TS1050071FP), leaving access to a 5 mm diameter site centered on monocular V1 (2.5–3.2 mm lateral to lambda)^{8,9}. The craniotomy site was filled with silicon elastomer (Kwiksil, World Precision Instruments, Inc) and covered with a layer of coloured nail varnish to discourage the group-housed animals from interfering with the site prior to electrophysiology experiments. A small surgical screw (Precision Technology Supplies, M1.2 × 2.4) was implanted contralaterally to act as a ground. The headplate attachment was reinforced with dental acrylic and the cut edges of the skin were re-attached using Histoacryl. Throughout the surgery, the eyes were covered with Vaseline or eye ointment to prevent drying.

Following the procedure, animals were transferred to a heated (~35 °C) incubation chamber, where they were allowed to recover to normal activity with *ad libitum* water and food access. The water restriction protocol was started 1–2 days following surgery.

Behavioural apparatus. We used an 800 mm (i.d.) polystyrene toroid (Ecclestone & Hart Ltd) to create an immersive visual environment based on the designs of previous virtual-reality environments^{36,37}. All programming of training and behavioural paradigms used LabVIEW System Design Suite coupled to a National Instruments Data Acquisition Board (NI-DAQ). Each sub-division of the task (task state) was matched to a visual background projected onto the interior of the dome. Visual backgrounds, including the equiluminant GO and NOGO stimuli, were generated using an OpenGL framework with C++ syntax. The resulting texture was mapped to a polygon on the computer screen using specified coordinates of a grid-like mesh to the dome interior (Tomazzo Muzzu, in preparation). A digital projector (ViewSonic PJD 6533w, 120 Hz refresh rate, output brightness 3500 Lumens) relayed the dome-adjusted image onto a rectangular mirror mounted at 35°, which reflected onto a 180° spherical mirror (Viso MS180 Indoor Dome Mirror) to ensure the image was projected onto the appropriate coordinates of the dome interior (Fig. 1). Water rewards were delivered through a blunted 19 G syringe needle attached to 0.5 mm bore silicone tubing (RS Components, Stock No. 667–8438). Water delivery was controlled by a TTL pulse to a single channel peristaltic pump (Campden Instruments, Product No. 80204-0.5). Licks were detected with a photomicrosensor (Omron Electronics, EE-SX4070) sampled at 50 Hz. All interruptions to the beam emitted by the sensor were counted as licks. Airpuff delivery was controlled by a 2-2 way solenoid valve (Shako Company, PU220AR-01, 1/8 inch) connected to an air pressure regulator, allowing manual adjustment of air pressure (10–15 psi). Airpuffs were delivered via a 1 mm copper piping angled towards the mouse from the side opposite to the monocular projection.

Training paradigm. Time spent in each phase of training was allocated on an individual basis governed by experimenter experience and pre-determined criteria. Throughout training and testing, the number of trials per session typically ranged between 100–300 trials. Length of session was determined by mouse participation levels; sessions were terminated after ~10–20 non-participatory trials. These non-participatory periods were not included in the analysis.

Pre-training phase: Mice were gradually habituated to head fixation over several days, starting with 5 minute sessions. **Autoshaping:** The S+ stimulus (reward-coupled) was a leftward drifting squarewave vertical grating (Temporal Frequency (TF): ~3.3 Hz). Only S+ was shown at this stage, alternated with blank (black) screen every 15 trials. For the first session, water reward was automatically delivered following the presentation of S+, before transitioning to the instrumental contingency where reward state was lick-triggered by the animal. Licking during the blank screen presentation was punished with an airpuff (0.3–0.5 sec) to the upper body to prevent compulsive over-licking. Mice progressed to the next stage when they displayed higher lick efficiency by targeting their licks to water rewarded time periods of the trial. **Full task (Full-field):** S+ and S– were alternated in a pseudorandom pattern determined by a random number generator in LabVIEW with a rule preventing the presentation of the same stimulus more than three times in a row. The S– stimulus (airpuff-coupled) was a downward drifting horizontal grating (TF: ~0.5 Hz). Progress during full-task training was evaluated using Signal Detection Theory (SDT) measures. When high discrimination sensitivity was observed in three of five consecutive sessions, mice entered the final phase of training where a monocular projection region was used (outer ~60–135° of the right visual field). Despite similarity of full-field and monocular stimuli (same SF and TF), performance level of the animals typically dropped at this stage of training. Animals took much longer to reach criterion performance levels, in some cases requiring adjustment of stimuli parameters (TF) to make them more easily discernable. Criterion performance was evaluated using a combination of mean/optimal d' values and observing the profile of Correct Rejection rates over session. When high sensitivity (as indicated by d' values >1.0) was observed in three of four sessions, the mice were considered well-trained. After reaching criterion animals were taken out of the daily training regime to prevent overtraining to habitual responding. Thereafter, animals were kept on a regular

“reminder” regime limiting sessions to one every 2/3 days, which enabled us to ensure a high level of performance on the task. These criteria were established in accordance with current behavioural literature^{14,18,35,38,39}.

“Satiated” state. Satiation was established by pre-feeding, which consisted of free water access for 1–2 minutes before the behaviour session.

Timings and State Flow of Task. Each trial started with a Stimulus State (2 sec total), which consisted of an initial 0.5 sec “Buffer period” and a 1.5 sec “Response Window”. During the Buffer, licks made by the animal did not count towards a decision/state change. The Buffer period was used within the trial structure to allow mice enough time to attend to the stimulus in view of the rapid succession of trials, and for mice to have time to bring licking under control, as guided by previous behavioural task literature^{9–11}. For every single mouse, FA trials had more licks in the Buffer period than Hit trials (see Supplementary Figure 1). S+ trials qualified as Hits when the animal licked within the Response Window of the stimulus presentation, and Misses otherwise. During the Reward State (1.75 s), S+ continued to be projected onto the screen while a 1–4 µl water reward was delivered to the mouse. S– trials were designated as Correct Rejections (CR) if no licks occurred during the Response Window and False Alarms (FA) otherwise. Punishment for FAs consisted of a mild airpuff (0.5 sec) to the upper body and a 5 sec timeout period while a black screen was projected onto the dome. Any licks during the timeout period triggered additional brief (10 ms) airpuffs. Misses and CRs were not negatively or positively reinforced. A variable length (3–6 sec) Inter-trial Interval (ITI) State immediately followed the Reward, Punishment or Miss/CR States.

Data analysis. All analyses were carried out in MATLAB using scripts developed in house and built-in functions. Over a given session, a sliding window over 30 trials was used to calculate Hit and FA rates (step size: 1 trial). Hit rate was calculated as (Hits/ total S+ trials) and FA rate (FA/total S– trials). We clipped values of Hit and FA rates that were >0.99 or <0.01 ⁴⁰. Performance was evaluated using the SDT sensitivity measure (d') from the inverse normal cumulative distribution of the Hit and FA rates: $d' = z(\text{Hit}) - z(\text{FA})$. We split sessions into 3 segments (Initial, Middle, Final) and projected Hit/FA pairs into Receiver Operator Characteristic (ROC) space. Statistical tests were Student’s t-test for parametric pairwise and unpaired comparisons and repeated measure ANOVA with post-hoc comparisons; error bars show s.e.m.

Lick latency peri-stimulus-time-histogram (PSTHs). We calculated histograms (time bins = 0.04 sec) of first licks per trial for the group. All PSTHs were convolved with a Gaussian filter for smoothing and normalized to the maximum bin count for the session (whole session PSTH) or for the segment (trial outcome PSTHs) to determine the relative likelihood of licks occurring within particular time bins. We classified the bimodal distribution of the lick latencies, into primary (1°) “Pavlovian” (0–0.2 sec) and secondary (2°) “Instrumentally Driven” responses (0.4 sec⁺).

Reinforcement Learning Modelling. To model behavioural responses, we used the classic Q learning model^{41,42}. In this model, the agent takes probabilistic actions in certain states, given a set of real values called Q-values. Our version of the model uses two actions (L, NL) and two states (GO, NOGO). The states correspond to two different stimuli, S+ and S–, for Go and NoGo respectively, and the two actions are L (Lick) and NL (No Lick). In total there are two Q-values in each time step, each corresponding to a given state ($Q_{1,t}$:S+ at time t and $Q_{2,t}$:S– at time t). The higher the Q-value the more likely the agent will take the lick action. Conversely the lower the Q-value the less likely the agent will take the no-lick action. To model mouse decision making, we used the non-deterministic Softmax decision rule,

$$P(A_{i,t} = 1) = \frac{e^{Q_{i,t}}}{e^{Q_{1,t}} + e^{Q_{2,t}}}, \quad (1)$$

where $p(A_{i,t} = 1)$ is the probability of taking the lick action given that the mouse is in the i -th state at time t , $Q_{i,t}$ is the value of licking in the i -th state in the t -th time bin, and T is a “temperature” parameter that was fixed to 1 in all our simulations. A session corresponds to $N = 150$ trials where the first action is randomly assigned to 0 or 1 with a probability of $p = 0.5$. We used $R = 15$ repetitions of this protocol and computed the hit and false alarm rate each time. This choice of parameters reflected the experimental protocol.

We compared two different models, which we have denoted AC and MAC. The AC model is the classic actor-critic Q-learner, which learns at the single session level, with the Q-values being modified as

$$Q_{i,t+1} = Q_{i,t} - \alpha(R_t - Q_{i,t}) \quad (2)$$

where Q_i is the value at a given time t , α is the learning rate, and R_t the reward provided to the agent at time t . This updated rule guarantees that the Q values converge towards the mean reward when the agent takes the action in this state. For AC model, the reward is 1 when the mouse licks during a GO stimulus or when it refrains from licking during a NOGO stimulus, otherwise the reward is -1 . Figure 3a shows results from the AC model where Q-values are initialized to 1 under the assumption that the mice have learnt the task contingencies during discrimination training.

The MAC (Motivated Actor-Critic) model is similar to AC in all respects, except that we have introduced a real valued variable M which accounts for the motivation or thirstiness of the agent. M diminishes by a fixed step m each time the agent received a reward. The Q-values we used are now biased by a motivation variable M ,

$$Q_{i,t} = Q_{i,t} + M_t. \quad (3)$$

M is modified if the mouse receives a liquid reward during a GO stimulus:

$$M_{t+1} = M_t - \frac{R_t M_0}{N_{\text{trials}}}. \quad (4)$$

The code used to generate Fig. 3 is available as a module (motivBCS2015) that can be downloaded from our website (link to be provided). All the parameters used in our model are presented in Table 2.

References

- Guitart-Masip, M., Talmi, D. & Dolan, R. Conditioned associations and economic decision biases. *NeuroImage* **53**(1), 206–214 (2010).
- Dayan, P. & Niv, Y. Reinforcement learning: the good, the bad and the ugly. *Curr. Opin. Neurobiol.* **18**(2), 185–96 (2008).
- Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191**(3), 507–20 (2007).
- Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**(3), 470–85 (2012).
- Huberman, A. D. & Niell, C. M. What can mice tell us about how vision works? *Trends. Neurosci.* **34**(9), 464–73 (2011).
- Carandini, M. & Churchland, A. K. Probing perceptual decisions in rodents. *Nat. Neurosci.* **16**(7), 824–31 (2013).
- Fu, Y. *et al.* A cortical circuit for gain control by behavioral state. *Cell* **156**(6), 1139–52 (2014).
- Glickfeld, Histed, M. H. & Maunsell, J.H.R. Mouse Primary Visual Cortex Is Used to Detect Both Orientation and Contrast Changes. *J. Neurosci.* **33**(50), 19416–19422 (2013).
- Lee, S.-H. *et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**(7411), 379–83 (2012).
- Pinto, L. *et al.* Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nat. Neurosci.* **16**(12), 1857–63 (2013).
- Bracey, E. F., Pichler, B., Schaefer, A. T., Wallace, D. J. & Margrie, T. W. Perceptual judgements and chronic imaging of altered odour maps indicate comprehensive stimulus template matching in olfaction. *Nat. Comm.* **4**, 2100 (2013).
- Zhang, S. *et al.* Long-range and local circuits for top-down modulation of visual cortex processing. *Science* **345**(6197), 660–665 (2014).
- Busse, L. *et al.* The detection of visual contrast in the behaving mouse. *J. Neurosci.* **31**(31), 11351–61 (2011).
- Komiyama, T. *et al.* Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464**(7292), 1182–6 (2010).
- Mayack, C. & Naug, D. Starving honeybees lose self-control. *Biol. Lett. -UK* **11**(1), 20140820 (2015).
- Andermann, M. L., Kerlin, A. M. & Reid, R. C. Chronic Cellular Imaging of Mouse Visual Cortex During Operant Behavior and Passive Viewing. *Front. Cell. Neurosci.* **4**, 16 (2010).
- Stüttgen, M. C., Schwarz, C. & Jäkel, F. Mapping spikes to sensations. *Front. Neurosci.* **5**, 125 (2011).
- Rivalan, M., Valton, V., Serié, P., Marchand, A. R. & Dellu-Hagedorn, F. Elucidating poor decision-making in a rat gambling task. Ravel, N. ed. *PLoS One* **8**(12), e82052 (2013).
- Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proc. Natl. Acad. Sci. USA* **101**, 13124–31 (2004).
- Dayan, P. & Balleine, B. W. Reward, Motivation, and Reinforcement Learning. *Neuron* **36**(2), 285–298 (2002).
- Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**(8), 861–874 (2006).
- Dickinson, A. & Balleine, B. Motivational control of goal-directed action. *Anim. Learn. Behav.* **22**(1), 1–18 (1994).
- Mayrhofer, J. M. *et al.* Novel two-alternative forced choice paradigm for bilateral vibrotactile whisker frequency discrimination in head-fixed mice and rats. *J. Neurophysiology* **109**(1), 273–84 (2013).
- Mayse, J. D., Nelson, G. M., Park, P., Gallagher, M. & Lin, S.-C. Proactive and reactive inhibitory control in rats. *Front. Neurosci.* **8**, 104, doi: 10.3389/fnins.2014.00104 (2014).
- Bari, A. & Robbins, T. W. Inhibition and impulsivity: behavioral and neural basis of response control. *Prog. Neurobiol.* **108**, 44–79 (2013).
- Falk, J. L. The nature and determinants of adjunctive behavior. *Physiol. Behav.* **6**, 577–588 (1971).
- Rescorla, R. A. & Solomon, R. L. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychol. Rev.* **74**(3), 151–82 (1967).
- Schultz, W. Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* **57**, 87–115 (2006).
- Dayan, P. & Berridge, K. C. Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cogn. Affect. Behav. Neurosci.* **14**, 473–92 (2014).
- Aberman, J. E. & Salamone, J. D. Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement. *Neuroscience* **92**(2), 545–52 (1999).
- Salamone, J. D. *et al.* Haloperidol and nucleus accumbens dopamine depletion suppress lever pressing for food but increase free food consumption in a novel food choice procedure. *Psychopharmacology* **104**(4), 515–521 (1991).
- Stüttgen, M. C. & Schwarz, C. Psychophysical and neurometric detection performance under stimulus uncertainty. *Nat. Neurosci.* **11**(9), 1091–9 (2008).
- Nienborg, H., Cohen, M. R. & Cumming, B. G. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu. Rev. Neurosci.* **35**, 463–83 (2012).
- Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* **6**(5), 363–75 (2005).
- Smith, K. S., Virkud, A., Deisseroth, K. & Graybiel, A. M. Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **109**(46), 18932–7 (2012).
- Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**(7266), 941–6 (2009).
- Hölscher, C., Schnee, A., Dahmen, H., Setia, L. & Mallot, H. A. Rats are able to navigate in virtual environments. *J. Exp. Biol.* **208**(Pt 3), 561–9 (2005).
- Horner, A. E. *et al.* The touchscreen operant platform for testing learning and memory in rats and mice. *Nat. Protoc.* **8**, 1961–84 (2013).
- Izquierdo, A. *et al.* Basolateral amygdala lesions facilitate reward choices after negative feedback in rats. *J. Neurosci.* **33**, 4105–9 (2013).
- Miller, J. The sampling distribution of d' . *Percept. Psychophys.* **58**(1), 65–72 (1996).
- Watkins, C. J. C. H. Learning from Delayed Rewards. PhD thesis, Cambridge University, Cambridge, UK (1989).
- Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. No.1. Cambridge, MIT press Vol. 1. (1998).

Author Contributions

A.B. performed the experimental work described, R.D.C. performed the modeling work described, A.B. prepared the figures, and A.B., R.D.C. and S.R.S. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Berdichevskaia, A. *et al.* Performance in a GO/NOGO perceptual task reflects a balance between impulsive and Instrumental components of behaviour. *Sci. Rep.* **6**, 27389; doi: 10.1038/srep27389 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>