
Article:
Thirkettle, M., Walton, T., Shah, A. et al. (3 more authors) (2013) The path to learning: Action acquisition is impaired when visual reinforcement signals must first access cortex. Behavioural Brain Research, 243. 267 - 272. ISSN 0166-4328

https://doi.org/10.1016/j.bbr.2013.01.023

The path to learning: Action acquisition is impaired when visual reinforcement signals must first access cortex

Martin Thirkettle[1*], Thomas Walton[1], Ashvin Shah[1], Kevin Gurney[1], Peter Redgrave[1], Tom Stafford[1]

[1]Department of Psychology, University of Sheffield Western Bank Sheffield S10 2TN United Kingdom

[*]Corresponding Author, Email: m.thirkettle@sheffield.ac.uk, Tel: +44 (0) 1142 226565, Fax: +44 (0) 114 276 6515

Published as

Thirkettle, M., Walton, T., Shah, A., Gurney, K., Redgrave, P., & Stafford, T. (2013). The path to learning: Action acquisition is impaired when visual reinforcement signals must first access cortex. ***Behavioural Brain Research, 243***, 267–272. doi:10.1016/j.bbr.2013.01.023

Abstract

Animals, interacting with the environment, learn and exploit the consequences of their movements. Fundamental to this is the pairing of salient sensory input with recent motor output to form an action-outcome pair linking a performed movement with its outcome. Short-latency dopamine (DA) signalling in the basal ganglia has been proposed to support this crucial task. For visual stimuli, this DA signalling is triggered at short latency by input from the superior colliculus (SC). While some aspects of the visual signal (e.g. luminance), are relayed directly to the SC via the retinotectal projection, other information unavailable to this subcortical pathway must take a more circuitous route to the SC, first submitting to early visual processing in cortex. By comparing action-outcome pairing when the visual stimulus denoting success was immediately available to the SC, via the retinotectal pathway, against that when cortical processing of the signal was required, the impact this additional sensory processing has on action-outcome learning can be established. We found that action acquisition was significantly impaired when the action was reinforced by a stimulus ineligible for the retinotectal pathway. Furthermore, we found that when the stimulus was eligible for the retinotectal pathway but evoked an increased latency, action acquisition was not impaired. These results suggest that the afferent sensory pathway via the SC is certainly primary and possibly instrumental to the DA neurons' role in the discovery of novel actions and that the differences found are not due to simple sensory latency.

Keywords

Action outcome learning; Visual Pathways; Superior Colliculus; S-cone stimuli

Introduction

From discovering and refining the ‘knack’ of opening a sticky lock to mastering a new piece of technology, the successful identification of which recent motor behaviours caused a sensory outcome and the later exploitation of these action-outcome pairings is an essential part of interacting with our environment. Computationally, this skill demands the identification of the component of behaviour responsible for producing the incoming sensory information from all the aspects of motor output recently performed. This is a fundamental problem faced by all animals, and one which the basal ganglia (BG) have been suggested to underpin [1].

The BG are a group of sub-cortical nuclei that have afferent connections from and efferent connections to both sensory and motor areas. They are remarkably well conserved evolutionarily, and therefore the behavioural competencies in which they are thought to play a critical role must be of crucial and enduring importance [2]. Converging evidence suggests these include action selection [3] and reinforcement learning [4, 5]. We have proposed that short latency phasic activity of dopaminergic (DA) neurons within the BG acts to reinforce the coincidence of surprising sensory events and recent motor output [6]. Such encoding would allow the identification of potential action-outcome associations, and subsequent trial and error repetition would identify the components of behavioural output responsible for causing the initially unpredicted outcomes. Once action-outcome associations have been acquired they could then be refined and stored for later deployment.

Any system is constrained by its inputs, and this is particularly relevant in the case of a system seeking to associate motor output with the input from a novel sensory consequence. The superior colliculus (SC) is a sub-cortical sensorimotor structure strongly associated with visual orienting and low-level sensory processing [7-9]. While the SC is a multisensory area, it also functions

as an important source of short-latency visual input to all basal ganglia input nuclei, including DA neurones in the ventral midbrain [10, 11]. Visually responsive neurones in the intermediate and deep layers of the SC have direct projections to ventral midbrain DA neurones [6] and therefore are thought to play a critical role in the visual reinforcement of action-outcome learning [10].

While the output of all three retinal photoreceptor cone classes in response to visual stimuli are relayed to the visual cortex via the retino-geniculo-cortical pathway, a subset of photoreceptor output is also directly passed to the SC via the retinotectal pathway. Specifically, the visual signal carried by the retinotectal pathway is almost entirely monochromatic and luminance based [12], containing little or no input from the short-wave-sensitive cones [13] and also lacking colour opponency information. The direct tectonigral projection from the SC to the substantia nigra within the BG [14] means that this sensory signal is directly passed to the dopamine neurons in the BG, raising the implication that the function of these neurons may be constrained by the sensory priorities of the SC. While it would be tempting to make the further prediction that visual information not passed to the SC via the retinotectal pathway would remain unavailable to the DA neurons and the action-outcome learning circuits of the BG, recent work has shown that colour-based signals from the cerebral cortex do eventually reach the SC, arriving approximately 30ms after simple luminance based information [15, 16]. Therefore, these signals may also become available as BG input [10], possibly using the same tectonigral pathway. Indeed, cortically based visual processing has been shown to activate DA reinforcement mechanisms [17], but by slower, less direct route(s). Nevertheless, the division in visual signal processing provides the opportunity to investigate and quantify the relative efficacy of the two streams of visual sensory information.

The fact that the direct retinotectal pathway only processes a subset of the sensory information that is processed by the cortical pathway affords the use of psychophysical methods to investigate their relative efficacy in a

behavioural task focussed on action-outcome learning and action acquisition [18]. Our strategy was to compare action acquisition using calibrated, colour-based, visual stimuli (available only to the indirect cortical pathway) as the reinforcing stimulus with action acquisition using luminance-based visual stimuli (available to the direct retinotectal pathway). This approach allows us to directly capture the impact any difference in efficacy between the two pathways at a behavioural level.

Visual stimuli calibrated to isolate the response of the short-wave-sensitive cones (S-cone stimuli), and so avoid the retinotectal pathway, have been used to investigate phenomena of low-level vision and the workings of the oculomotor system [19-22]. However, here we use the lack of input from short-wave-sensitive cones to the SC in order to test whether visual reinforcement of action acquisition is supplied exclusively by the direct retinotectal projection to the SC [10]. Effective reinforcement based on isoluminant changes in stimulus chromaticity, isolating the response of the short wave-sensitive cones, would preclude this pathway and necessitate early cortical processing. Therefore colour based reinforcement should be less efficient than luminance-based reinforcement available to the direct tectonigral route [23]. In order to determine whether any impairment in performance should be attributed to the different sensory pathways taken by the reinforcing signals or simply due to the delayed arrival at the SC of the non-retinotectal signal, two experiments were conducted: The first contrasted action outcome learning performance using S-cone stimuli as a the reinforcing signal against that found with a luminance based reinforcing signal, the second manipulated the latency of sensory signal at the SC along the same pathway by comparing performance with luminance based reinforcing signals of differing luminance intensity.

Changes in luminance have previously been demonstrated to alter response latencies in the SC itself [24], and these in turn have been shown to be due to longer latencies in retinal processing for reduced luminance resulting in the delayed arrival of the sensory signal at the SC [25]. The effect of this on

behaviour has been captured by Piéron's law, which relates reaction time to stimulus intensity [26-28]. In experiment 2, by measuring each participant's psychometric and chronometric functions for luminance stimuli we were able to create a pair of stimuli, one corresponding to the luminance stimulus in the previous experiment, and one calculated using Piéron's law to produce a 25ms faster reaction time. Using these two stimuli of different luminance intensities, which invoke different processing delays at the level of response latency [29] but share a common sensory pathway, we can determine whether the effect on action learning performance found in experiment 1 is due to signal latency or to the different sensory pathways taken by the reinforcing stimuli.

If our findings showed that a less intense luminance stimulus produced the same degraded learning as the similarly delayed S-cone stimulus (when compared to a baseline luminance signal) the difference in performance could be attributed to the time of arrival at the SC of the sensory information. However, we find that the different luminance stimuli were equally effective in reinforcing the acquisition of novel actions, despite their different arrival times at the SC, and so we must conclude that the sensory processing of the retinotectal pathway produces a superior learning signal compared to that available to the S-cone stimuli.

Methods

Participants

Forty undergraduate psychology students (University of Sheffield, UK) served as subjects for course credit, 20 in experiment 1 (11 female, mean age 19.1), and 20 in experiment 2 (18 female, mean age 19.3yrs). All reported normal or corrected to normal vision without any colour vision impairment and, successfully completed the calibration phase of the experiment (which, in the case of experiment 1, would have detected any colour vision impairment had one been present). All participants were naive to the purposes of the experiment and the experiments were approved by University of Sheffield

Research Ethics Board and therefore conducted in accordance with the 1964 Declaration of Helsinki.

Materials

All stimuli were created using Matlab to control a Cambridge Research Systems Visage graphics board which was in turn driving a calibrated Mitsubishi Diamond Pro 2070sb 22” monitor screen at 160Hz. This apparatus was used for stimulus calibration and presentation in the experimental tasks. A chin rest ensured the participants remained seated 57cm from the screen throughout. All experiments were conducted in a darkened room with no light source other than the screen. All experimental control and data analysis programs were also conducted using Matlab.

Stimulus Display

Figure 2: Examples of reinforcing stimuli showing the spatial arrangement of the display and signals. The circles making up the central ring of the annulus were flashed in a change of either chromatic value (left) or luminance (right) to denote movement of the joystick into the hidden reinforcement area.

The visual display and reinforcement signal consisted of an annulus centred on the centre of the screen to prevent any spatial information from being conveyed (figure 2), and the small circles making up the annulus meant the luminance and chromaticity of portions of the display could be controlled independently. Stimuli were presented against a grey background (CIE 1976 coordinates 0.197 u', 0.442 v', 25cd/m^2). Each of the small circles making up the annulus had the same chromaticity as the background (when not presenting the S-cone signal) and changed luminance value each 6.5ms (1 frame) to a new, randomly selected value in the range 24-26cd/m^2. A white cross was placed at the centre of the screen for the participant to fixate upon throughout the experiment. The interior edge of the annulus was presented at 5˚ eccentricity from this cross and the external radius was at 9.5˚ eccentricity. The reinforcing signals comprised a change of either the mean luminance (which is available to the retinotectal pathway) or the chromaticity (which is not available to the retinotectal pathway) of the central 2˚ of the annulus for 12.5ms (2 frames). While there is some research suggesting that the inability of the retinotectal pathway to process the S-cone signal is not total, the lack of colour opponency in the retinotectal pathway means any S-cone signal can be effectively masked using luminance noise [30]. By presenting the signal for two frames, while luminance intensity fluctuated each frame, the change in chromaticity was indeed masked in luminance noise. The luminance signal was presented by increasing the luminance of the constituent circles within the signal region by a value set to be above that of the rest of the annulus by the participant's calibration. This was done concurrently with the frame-by-frame fluctuation of the luminance of all the circles of the annulus, so the circles making up the luminance signal still fluctuated to the same extent as the rest of the annulus, but around a raised luminance value.

Procedure

Experiment 1 – S-cone Vs Luminance

Calibration procedure

Participants first completed calibration and stimuli validation tasks before proceeding to the action-outcome learning task in a subsequent session. The calibration of chromaticity and luminance changes were calculated individually for each participant using techniques established previously [31]. As far as possible, the calibration tasks all used variations on the annulus display, and in each case the displays were arranged so that the measurements were taken using the same visual eccentricity as the presentation of the subsequent reinforcing signals. The calibration experiments used transient tritanopia to measure the participant's tritan line [31]. The minimum motion technique [32] was used to measure the point of equiluminance between the different chromatic values. Responses to the various calibration tasks were recorded with a dedicated response box (Cedrus RB-530). These procedures were then followed by a simple experiment to equate detectability between the two classes of stimuli. The annulus was split into four and the participant had to identify which quadrant of the annulus flashed during a presentation window of 1.8sec. Flashed stimuli lasted 12.5ms (2 frames) and the luminance of the circles making up the annulus jittered in each frame. An adaptive method to calculate psychometric performance was employed [33] where 80% thresholds were taken from the resulting functions for the S-cone and luminance stimuli using the 'psignifit toolbox' for Matlab [34].

Stimuli Validation - Reaction time experiment

After completing the calibration procedures, the bespoke S-cone and luminance stimuli, matched for detectability for each individual, were validated using a simple manual reaction time task. Previous research [19] reports reliably slower manual response times to S-cone stimuli than luminance stimuli. Reaction times were measured using the same annulus display with luminance and chromatic changes presented on either the left or right halves. Participants were required to respond using a button box (Cedrus RB-530) as to which half, left or right, of the annulus had changed. Changes lasted for 2 frames in

duration (12.5ms) as in the reinforcing stimulus and occurred randomly within a 2 second window. The side of the annulus which changed was randomised, and the two stimulus types were presented in separate blocks. The reaction time experiment was repeated twice, for a total of 160 responses to each stimulus. The order of the stimuli blocks was counterbalanced for each participant across the two sessions. Participants were instructed to respond quickly and accurately to the task and received feedback on their accuracy. Anticipatory responses given before stimulus presentation and slow responses greater than 1.8 seconds were discarded and the particular stimulus condition repeated. As in previous work [19] no further cleaning of the reaction time data was necessary as the median score proved robust to outliers. Participants demonstrated the characteristic slower response times to S-cone stimuli ($t(19) = 2.3$, $p < 0.05$). Median reaction times across participants were 292.3ms for Luminance stimuli and 312.6ms for S-cone stimuli, making a difference of 20.3ms.

Experiment 2 – Luminance Intensity

Calibration procedure & Learning task

Piéron's law relates stimulus intensity to response time [26], and was used to generate a luminance stimulus that would produce a reaction time difference comparable to that found between the luminance and S-cone signals of experiment 1. Here we combined a measurement of the participant's psychometric function for luminance stimuli detection with a measurement of their chronometric response to the same stimulus. Because two functions were being measured, the adaptive procedure used in the calibration procedure for experiment 1 was replaced with a more extensive method of constants approach. Using a 2AFC design, participants were required to respond as quickly and accurately as they could to a half annulus presented at one of five different levels of luminance. Participants had to press the button on the same side of the annulus which flashed. Psychometric and chronometric curves were then fitted to the data. Once these two functions had been fitted the 80%

detection threshold was used to form the control stimulus, and, using the chronometric function of the participant, the increased intensity necessary produce a 25ms decrease in reaction time was calculated. An increase in intensity was used because; while decreasing intensity would also produce a change in reaction time it would have been accompanied by a dramatic change in detectability. Increasing luminance intensity, and therefore detectability, beyond threshold toward asymptote should produce a much smaller, if any, change in task performance.

Experiments 1 & 2 - Action Outcome Learning Task

The effect of stimulus type on action acquisition was tested using the joystick paradigm [18]. In this task participants moved a joystick (Logitech Extreme 3D Pro joystick, P/N: 863225-1000) to discover the location of a randomly determined hidden target area (‘hotspot’), entry to which evoked a visual stimulus (figure 1). The search space was defined as a square of 1000 by 1000 units, and, after piloting, the hotspot was defined as a circle with an area 1.5% of the total search area (in this case a circle with a radius of 70 units). Participants started from the centre of the area, and searched for the hotspot by moving the joystick freely. The position of the joystick was monitored at 1000Hz. During their search, participants were instructed to maintain fixation on the centre of the screen. The location of the reinforced hotspot was randomly selected for each trial and was not allowed to overlap either the centre of the search area or any edge.

Neither the current location of the joystick nor the target locations were represented on screen during the trial, thereby precluding the use of direct visual feedback to identify the target location. The visual display was designed to reinforce movements that take the joystick into the target zone without providing any spatial cues as to the location of the target. However, the annulus of circles was constantly displayed, flickering in luminance at 160Hz as in the calibration tasks. Importantly, the participant received no feedback whatsoever until they encountered the target area at which point one of their

individually calibrated visual stimuli, either S-cone or luminance, was presented by changing either the chromaticity or mean luminance of the central band of the annulus of circles for a 12.5ms on-screen signal. Subsequent signals, if the participant remained within the target area, were separated by a 30ms refractory period, producing a maximum presentation frequency of 23.5Hz, and presentation of the visual signals ceased if the joystick was moved out of the target area. Discovery of the target location was determined by the participant maintaining the joystick within the target area consistently enough to receive 14 reinforcing signals within a 1s window. This criterion equated to 595ms of a 1 second window spent within the target area. The trial then ended and a new trial begun with a new, randomly located, target. Therefore, a particular advantage of this task is that it can accommodate repeated measures designs to investigate action acquisition. This task produces rich behavioural data in the form of the movement record of the joystick, allowing the discovery of a novel action to be recorded and studied (figure 1).

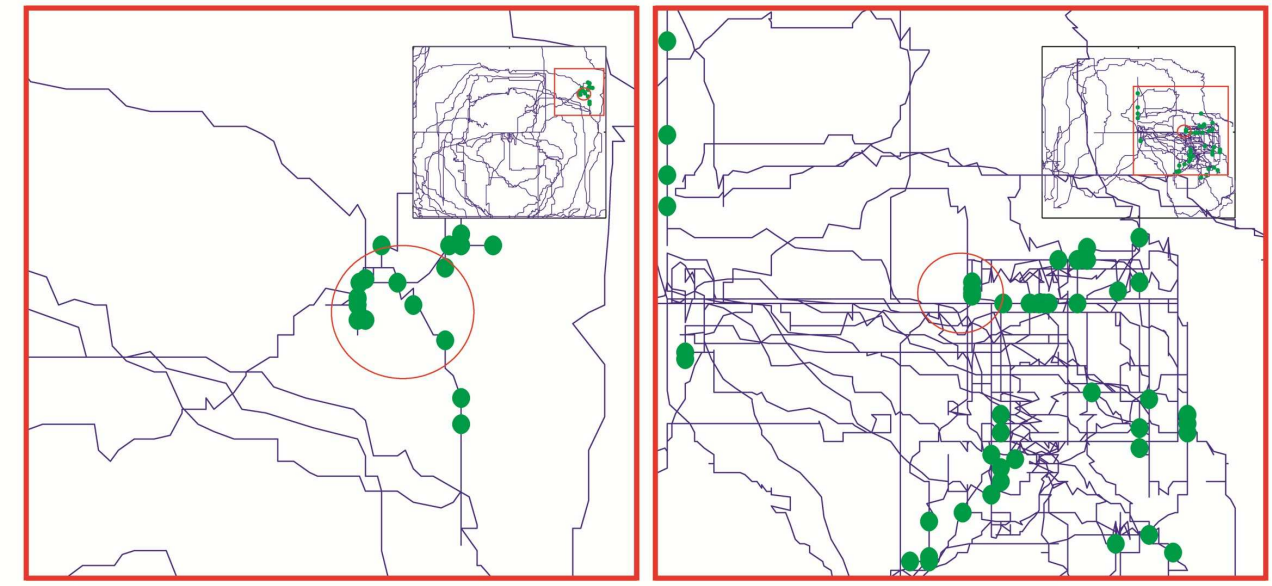

Figure 1: Example joystick movement traces from single trials from one participant at low, 75ms, delay (left) and high, 375ms, delay (right). The path of the joystick is shown in blue, the target area is circled in red and

the green dots represent the location of the joystick at the moment reinforcement for previously entering the hotspot was delivered. Inset plots show the entire search space, main plots show enlarged view of area bordered in red, centred on the target.

A fundamental challenge for action learning is correctly assigning responsibility for sensory events to components of your own behaviour – the 'credit assignment problem' [35]. Delays between action and outcome exacerbate this problem by allowing contamination of the record of potentially causal motor outputs with irrelevant action components, which, because of their temporal proximity to the sensory event, are also reinforced [36, 37]. In line with this, behavioural evidence from action-outcome learning paradigms has shown that human and non-human animals suffer a decline in reinforcement efficacy with increasing intervals between action and reinforcement signal [18, 38, 39]. We therefore expected that performance should be vulnerable to the insertion of an additional external delay between successful actions and reinforcer presentation. Participants therefore completed trials where the visual reinforcer was presented either immediately upon target encounter, or after a predetermined delay. Six delay conditions were used: 0ms, 75ms, 150ms, 225ms, 300ms and 375ms with participants completing each condition 3 times giving a total of 18 trials per stimulus type. The order of the delay condition trials were randomised for each stimulus and the different stimuli trials - luminance and S-cone, or high and low luminance – were presented in separate blocks. The order of testing these blocks was counterbalanced between participants and a self-paced break was provided for the participants between blocks. Learning performance was recorded as the time taken between the initial encounter with the target, and maintenance of the joystick in the target location sufficiently long to achieve the criterion rate of reinforcement (14 flashes in a second) - the 'homing period'. For each subject in experiment 1,

the entire experiment, including calibration and RT experiment, lasted about two hrs, while experiment 2 lasted around an hour in total.

Results

Experiment 1

Figure 3 shows that the participants were able to acquire novel actions when the reinforcing signal was unavailable to the retinotectal pathway, and that this learning was reliably impaired when compared against that based upon a luminance signal directly available to the SC. The time required by subjects to discover the target location was significantly longer when the task was reinforced by S-cone visual stimuli ($F(1,19) = 8.27$, $p < 0.01$). While we were not able to resolve this performance difference at each delay level, on average across delay levels it represented an approximately 2-3 second decrement in learning performance on the joystick task (Fig 3, inset). It is unlikely, therefore, that an additional 20-30ms delay in the arrival of colour signals to the SC [15] would be sufficient to account for the increases in homing period observed. Participants required significantly more S-cone reinforcements than luminance ($F(1,19) = 7.59$, $p < 0.05$) to learn target location, despite the two being equated for detectability. While this could imply that luminance, compared with colour-stimuli, are more effective at attracting attention, previous research has shown that this is not the case. Sumner and colleagues [22] reported that S-cone signals are equally effective in attracting attention. Therefore, we suggest that the S-cone signal provides a comparatively less effective input to the reinforcement learning mechanisms of the BG.

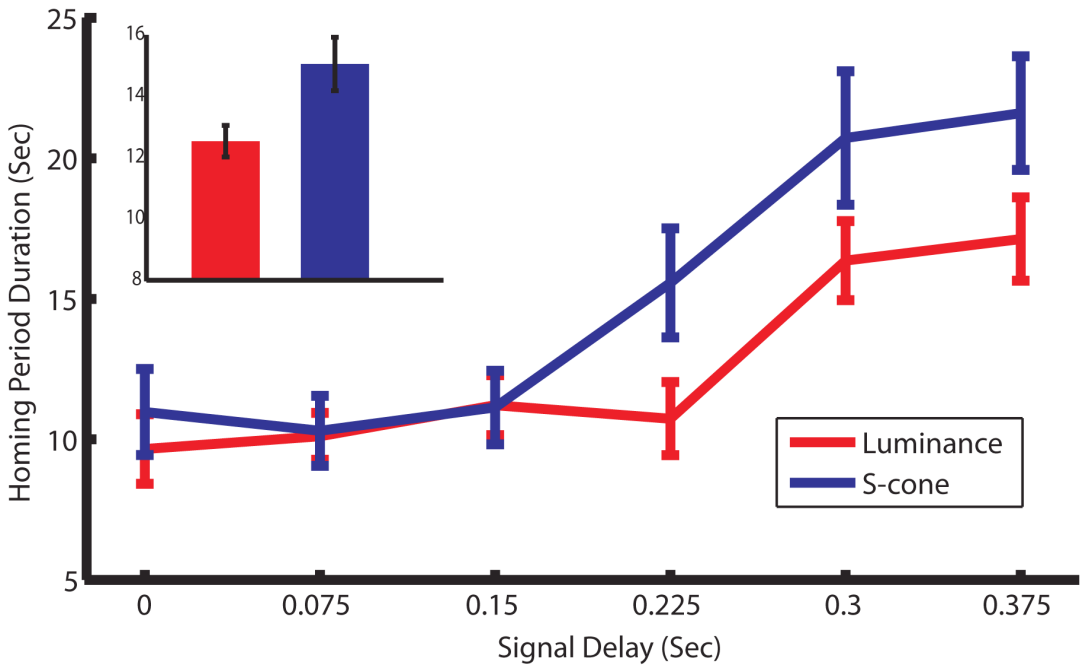

Figure 3: S-cone stimuli are less efficient than luminance stimuli at reinforcing action acquisition. Learning across stimulus conditions and signal delay conditions show decrements in performance with the S-cone stimuli. Inset shows average learning performance collapsed across signal delay.

An expected finding was that increasing the imposed delay between the movement evoking visual reinforcement and the reinforcing visual signals had a dramatic effect on learning. The longer the delay, the longer it took to acquire the action successfully ($F(5,95) = 14.9$, $p < 0.001$). This was also true of the distance participants moved the joystick while searching for the target area ($F(1,19) = 6.77$, $p < 0.05$). As previously discussed, this demonstrates the difficulty in assigning credit for a caused event when the motor output is temporally separated from the outcome. The effect of reinforcement delay on learning performance becomes striking for S-cone and luminance reinforcement signals at delay levels exceeding 150ms and 225ms respectively. Beyond this point the relationship between additional delay and

increases in homing period duration is remarkably constant across the two stimulus types suggesting a fixed cost of the use of reinforcement signals ineligible for the direct pathway to the SC. While at first glance the difference in the number of reinforcement signals between conditions could suggest that the performance impairment can be explained in terms of a cumulative effect of additional sensory latencies, the size of this difference in absolute terms is not comparable to the delay required to produce such a difference. Although it may be the case that the additional sensory delay of the S-cone signal is analogous to an imposed external delay, it does not appear that this is of sufficient extent to explain the differences we report here.

Experiment 2

We found no difference between learning performance, as measured by homing period duration, between the two luminance stimuli ($F(1,19) = 0.132$, $p > 0.05$). Neither did we replicate the significant difference in number of signals required for success when two luminance signals of differing intensity were used ($F(1,19) = 3.77$, $p > 0.05$). Thus we find no evidence to suggest that the impairment in learning performance found with S-cone stimuli in experiment 1 is due to the additional latency the non-retinotectal signal incurred in its progression to the SC, and so we suggest that the performance difference found in experiment 1 results from the S-cone stimuli's lack of direct access to the SC.

Discussion

Results from both studies are consistent with our underlying hypotheses that the BG is the neural locus of action acquisition [6] and the SC is the primary sensory input for this fundamental process [10]. Experiment 1 shows that for the task of action acquisition, sensory information available to the more direct pathway to the SC is a more effective reinforcing signal. Experiment 1 also replicates the well-established finding that temporal contiguity between action and outcome is important for reinforcement. When sensory latency within a

single pathway was manipulated in experiment 2, we find that it is the pathway travelled by the sensory signal, not the sensory latency that is responsible for the performance impairment found in experiment 1.

Until recently [17], the latency of sensory-evoked phasic DA responses (70-100ms following stimulus onset in primates) seemed to preclude a contribution from cortical sensory processing [6]. This view was supported by a range of anatomical and physiological data implicating the SC as a supplier of afferent visual information to DA neurons at sufficiently short latencies [23, 40]. The present study in humans was designed to test the relative efficacies of visual stimuli, processed cortically and sub-cortically, in reinforcing the acquisition of novel actions. Our results show that S-cone stimuli which can be processed only by cortical visual systems, and cannot access the direct, retinotectal, pathway to the SC, can still act as effective reinforcers. Thus, the previous suggestion that sub-cortical visual processing may be the exclusive source of short-latency visual input to DA action learning mechanisms [10, 23], now needs to be expanded to include early visual processing by cerebral cortex.

However, those stimuli requiring cortical processing were inferior for reinforcing action acquisition when compared to luminance-based stimuli that access the SC directly via the retinotectal pathway. Current, as yet unpublished data from our laboratory, suggest that activation of DA neurones by visual cortex does so by accessing the evolutionary prior tectonigral projection. The present results therefore show that cortically processed visual signals relayed to DA neurones, possibly via the SC, are utilised less efficiently to support action acquisition than luminance information that can be processed directly by the SC. Experiment 2 demonstrates that increased sensory latency is not sufficient to explain this performance decrement. It is therefore likely that the sensory pathway taken by afferent signals to the DA neurons is of importance to the validity of the reinforcing signal. One possibility is that the behavioural difference reflects an attenuated dopamine signal within the BG in response to cortical visual input. Future studies are planned to explore this suggestion.

Conclusions

Our data demonstrate the behavioural consequences of the afferent sensory connections of the BG on one of its most critical functions: encoding the juxtaposition of motor output with incoming sensory information to establish contingency. Previous work has established the BG's, evolutionarily sensory input from the SC is well conserved, and here we exploited the sensory processing characteristics of the SC in order to probe BG function. We show that sensory information not readily available to the preferred input of the BG is less effective for action-outcome learning, and demonstrate that this difference cannot be explained by simple sensory latency alone. We conclude that signals ineligible for the retinotectal pathway produce impaired learning, not because they are slow to reach the SC, but because they present a degraded signal to the action-outcome learning mechanisms of the BG.

Acknowledgements

The Authors wish to thank Dr Petroc Sumner for help and advice in the creation of the S-cone stimuli. Thanks to Peter Dayan, Jen Lewis, Nicolas Vautrelle, Nicol Harper, and Stuart Wilson for useful discussion of the results and manuscript. The research leading to the results presented here was supported by the European Community's Seventh Framework Programme FP7/2007-2013, "Challenge 2 - Cognitive Systems, Interaction, Robotics'', under grant agreement No FP7-ICT-IP-231722, project ''IM-CLeVeR - Intrinsically Motivated Cumulative Learning Versatile Robots''. TW was supported by an EPSRC DTC grant.

References

[1] Redgrave P, Gurney KN, Reynolds J. What is reinforced by phasic dopamine signals? Brain research reviews. 2008;58:322-39.
[2] Stephenson-Jones M, Samuelsson E, Ericsson J, Robertson B, Grillner S. Evolutionary conservation of the Basal Ganglia as a common vertebrate mechanism for action selection. Current Biology. 2011;21:1081-91.

[3] Redgrave P, Prescott T, Gurney K. The basal ganglia: a vertebrate solution to the selection problem? Neuroscience. 1999;89:1009-23.
[4] Schultz W. Predictive Reward Signal of Dopamine Neurons. Journal of Neurophysiology. 1998;80:1-27.
[5] Redgrave P, Vautrelle N, Reynolds JNJ. Functional properties of the basal ganglia's re-entrant loop architecture: selection and reinforcement. Neuroscience. 2011;198:138-51.
[6] Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? Nature Reviews Neuroscience. 2006;7:967-75.
[7] Perenin MT, Jeannerod M. Subcortical vision in man. Trends in Neurosciences. 1979;2:204-7.
[8] Isa T. Intrinsic processing in the mammalian superior colliculus. Current Opinion in Neurobiology. 2002;12:668-77.
[9] Sparks DL. Conceptual issues related to the role of the superior colliculus in the control of gaze. Current Opinion in Neurobiology. 1999;9:698-707.
[10] Dommett E, Coizet V, Blaha CD, Martindale J, Lefebvre V, Walton N, et al. How visual stimuli activate dopaminergic neurons at short latency. Science. 2005;307:1476-9.
[11] Coizet V, Schulz JM, Reynolds JNJ, Overton PG, Redgrave P. Sensory reinforced corticostriatal plasticity in the anaesthetised rat. Society for Neuroscience Abstract Viewer and Itinerary Planner. 2008;38.
[12] Boehnke SE, Munoz DP. On the importance of the transient visual response in the superior colliculus. Current Opinion in Neurobiology. 2008;18:544-51.
[13] Marrocco RT, Li RH. Monkey superior colliculus: properties of single cells and their afferent inputs. Journal of neurophysiology. 1977;40:844-60.
[14] May PJ, McHaffie JG, Stanford TR, Jiang H, Costello MG, Coizet V, et al. Tectonigral projections in the primate: a pathway for pre-attentive sensory input to midbrain dopaminergic neurons. The European journal of neuroscience. 2009;29:575-87.
[15] White B, Boehnke S, Marino R, Itti L, Munoz D. Color signals in the primate superior colliculus. Journal of Vision. 2010;9:396-.
[16] Vautrelle N, Leriche M, Dahan L, Redgrave P, Overton PG. Cortical control of dopaminergic neurons. Society for Neuroscience Abstract Viewer and Itinerary Planner. 2011;41.
[17] Nomoto K, Schultz W, Watanabe T, Sakagami M. Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. The Journal of Neuroscience. 2010;30:10692-702.
[18] Stafford T, Thirkettle M, Walton T, Vautrelle N, Hetherington L, Port M, et al. A Novel Task for the Investigation of Action Acquisition. PLoS ONE. 2012;7:e37749.
[19] Bompas A, Sumner P. Sensory sluggishness dissociates saccadic, manual, and perceptual responses: an S-cone study. Journal of Vision. 2008;8:1-13.
[20] Bompas A, Sumner P. Oculomotor distraction by signals invisible to the retinotectal and magnocellular pathways. Journal of Neurophysiology. 2009;102:2387-95.
[21] Smithson HE, Mollon JD. Is the S-opponent chromatic sub-system sluggish? Vision Research. 2004;44:2919-29.
[22] Sumner P, Adamjee T, Mollon JD. Signals invisible to the collicular and magnocellular pathways can capture visual attention. Current Biology. 2002;12:1312-6.
[23] Comoli E, Coizet Vr, Boyes J, Bolam JP, Canteras NS, Quirk RH, et al. A direct projection from superior colliculus to substantia nigra for detecting salient visual events. Nature Neuroscience. 2003;6:974-80.
[24] Bell AH, Meredith MA, Van Opstal AJ, Munoz DP. Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. Experimental Brain Research. 2006;174:53-9.
[25] Valjakka A. The reflection of retinal light response information onto the superior colliculus in the rat. Graefes Archive for Clinical and Experimental Ophthalmology. 2007;245:1199-210.

[26] Stafford T, Gurney KN. The role of response mechanisms in determining reaction time performance: Pieron's law revisited. Psychonomic Bulletin & Review. 2004;11:975-87.
[27] Palmer J, Huk AC, Shadlen MN. The effect of stimulus strength on the speed and accuracy of a perceptual decision. Journal of Vision. 2005;5:376-404.
[28] Piéron H. The sensations: Their functions, processes, and mechanisms. London: Müller; 1952.
[29] Jaśkowski P, Sobieralska K. Effect of stimulus intensity on manual and saccadic reaction time. Attention, Perception, & Psychophysics. 2004;66:535-44.
[30] Mollon JD. Color vision. Annual review of psychology. 1982;33:41-85.
[31] Smithson HE, Sumner P, Mollon JD. How to find a Tritan Line. Normal and defective colour vision, London: Oxford University Press; 2003. p. 279-87.
[32] Anstis SM, Cavanagh P. A minimum motion technique for judging equiluminance. Color vision: Physiology and psychophysics. 1983:156-66.
[33] Kontsevich L, Tyler C. Bayesian adaptive estimation of psychometric slope and threshold. Vision Research. 1999;39:2729-37.
[34] Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Perception & Psychophysics. 2001;63:1293-313.
[35] Minsky M. Steps toward artificial intelligence. Proceedings of the IRE. 1961;49:8-30.
[36] Izhikevich EM. Solving the Distal Reward Problem through Linkage of STDP and Dopamine Signaling. Cerebral cortex. 2007;17:2443-52.
[37] Sutton RS, Barto AG. Reinforcement Learning - an introduction. Cambridge, MA: MIT Press; 1998.
[38] Elsner B, Hommel B. Contiguity and contingency in action-effect learning. Psychological Research. 2004;68:138-54.
[39] Walton T, Thirkettle M, Redgrave P, Gurney K, Stafford T. Very short reinforcement delays impair action discovery in humans. 2012;Submitted.
[40] McHaffie JG, Jiang H, May PJ, Coizet V, Overton PG, Stein BE, et al. A direct projection from superior colliculus to substantia nigra pars compacta in the cat. Neuroscience. 2006;138:221-34.