Reinforcing saccadic amplitude variability in a visual search task

Human observers often adopt rigid scanning strategies in visual search tasks, even though this may lead to suboptimal performance. Here we ask whether specific levels of saccadic amplitude variability may be induced in a visual search task using reinforcement learning. We designed a new gaze-contingent visual foraging task in which finding a target among distractors was made contingent upon specific saccadic amplitudes. When saccades of rare amplitudes led to displaying the target, the $U$ values (measuring uncertainty) increased by 54.89% on average. They decreased by 41.21% when reinforcing frequent amplitudes. In a noncontingent control group no consistent change in variability occurred. A second experiment revealed that this learning transferred to conventional visual search trials. These results provide experimental support for the importance of reinforcement learning for saccadic amplitude variability in visual search.

Introduction

Humans use visual information to control a wide range of behaviors: to navigate in their environment, glean specific information, or find an object, for instance. Regularities often found in scan patterns (i.e., eye movement sequences) of observers viewing the same stimuli have been linked to low-level image properties like luminance, contrast, or color – the salient points being correlated with human fixations (Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002; Treisman & Gelade, 1980; Wolfe, 2007). However image features are typically kept constant across subjects in experimental situations and cannot account for the idiosyncratic nature of eye movement sequences. Indeed, stable interindividual differences have been observed in scan patterns of subjects required to repeat various visual tasks such as encoding and recognizing stimuli (Foulsham & Underwood, 2008; Noton & Stark, 1971), answering questions about photographs (Greene, Liu, & Wolfe, 2012), reading a text, or viewing simple patterns and complex natural scenes without specific instruction (Andrews & Coppola, 1999). Importantly, idiosyncrasies in eye movement sequences have also been reported in visual search tasks (e.g., Andrews & Coppola, 1999; Choi, Mosley, & Stark, 1995; Myers & Gray, 2010). High level factors such as task demand (Tatler, Wade, Kwan, Findlay, & Velichkovsky, 2010; Yarbus, 1967) or knowledge about the visual properties and the statistical regularities of the environment (Chen & Zelinsky, 2006; Jiang, Swallow, Rosenbaum, & Herzig, 2013; for reviews, see also Eckstein, 2011; Schütz, Braun, & Gegenfurtner, 2011; Tatler, Hayhoe, Land, & Ballard, 2011) are probably involved in these interindividual differences. In addition, it has been found that subjects maintain their specific scanning strategy across sessions even when it leads to suboptimal performance. This is evident in a series of studies from Boot and collaborators in which observers had to detect transient changes in dynamic displays (Boot, Becic, & Kramer, 2009; Boot, Kramer, Becic, Wiegmann, & Kubose, 2006). Detection accuracy was inversely correlated with the number of saccades. Despite these requirements, those subjects who frequently moved their eyes did not
spontaneously refrain from making saccades to increase their performance. Interestingly, Boot et al. (2006) and Boot et al. (2009) showed that only when given instructions and monetary reward were observers able to change their strategy, and consequently their detection rate.

Another example comes from a recent study from Morvan and Maloney (2012) who asked human observers to perform a visual search task with a display composed of three aligned tokens. Subjects had to emit one saccade in order to detect a change occurring in one of the side items. The separating distance or the size of these items were varied so that the optimal eye movement strategy abruptly changed, fixing the central item being optimal only for the smallest distances or largest token sizes. None of the subjects adapted their strategy to the stimulus configuration, and participants instead maintained their original search pattern (Morvan & Maloney, 2012).

However, there are some experimental evidences of modifications in visual search strategies as well. For instance a set of recent experiments probed the effects of reinforcement in visual search tasks and found that reward strongly affects eye movements’ patterns (Hickey, Chelazzi, & Theeuwes, 2010; Hickey & van Zoest, 2013; Hickey & van Zoest, 2012) even when the target is not visible (Chukoskie, Snider, Mozer, Krauzlis, & Sejnowski, 2013). That rigid strategies may be affected by reinforcement may be related to the exploration-exploitation tradeoff, a concept often used in the reinforcement learning approach of machine learning (Sutton & Barto, 1998). Exploration is thought to be needed because it allows exploiting new contingencies, a rationale that grounds empirical studies on behavioral variability (Page & Neuringer, 1985; for reviews, see Neuringer, 2002; 2009). There is in fact a long tradition of research probing the effects of reinforcing variable responses in a variety of behavior and species such as porpoises (Pryor, Haag, & O’Reilly, 1969), pigeons (Blough, 1966) or rats (Grunow & Neuringer, 2002). The main outcome of this research is that variability may be controlled by reinforcement and influences adaptation to environmental contingences (Neuringer, 2002, 2009). It has been shown that variability in saccades may also be affected by reinforcement (Madelain, Champrenaut, & Chauvin, 2007; Paeye & Madelain, 2011) but it remains unclear whether these effects would hold in visual search tasks.

Previous studies investigated the influence of reinforcement learning on target selection (for reviews, see Glimcher, 2003; Trommershäuser, Glimcher, & Gegenfurtner, 2009). In most learning experiments, saccades to specific locations are explicitly reinforced with extraneous consequences—food in monkeys or money in humans. Typically, monkeys are required to make a saccade to one of several targets, each one being associated with specific magnitudes (Platt & Glimcher, 1999) or probabilities (Sugrue, Corrado, & Newsome, 2004) of alimentary reward. A linear relationship, termed the matching law (Herrnstein, 1961), is observed such that the relative frequency of saccades to each target is close to the relative amount of reward obtained when choosing this target. Studies in humans demonstrated that monetary gains associated with particular targets influence the orientation of saccades made towards (Chen, Mihalas, Niebur, & Stuphorn, 2013; Hickey & van Zoest, 2012; Liston & Stone, 2008) and within these targets (Schütz, Trommershäuser, & Gegenfurtner, 2012; Stritzke, Trommershäuser, & Gegenfurtner, 2009), including in visual search tasks (e.g., Navalpakkam, Koch, Rangel, & Perona, 2010; see Anderson, 2013, for a review).

Some studies showed that visual consequences as well can act as a reinforcing consequence of target selection. For example, when human participants are required to choose between the presentations of simple or complex visual patterns by pressing either of two keys, the response rate on the key associated with the complex stimuli is higher than on the key associated with the simple stimuli (Berlyne, 1972), indicating that complex stimuli have a higher reinforcing value than simple ones. Recently, Collins (2012) found that humans choose more often to make a saccade towards a target that remained visible after the eye movement than towards a target systematically extinguished during the saccade.

Within the visual search framework, a statistical model proposed by Najemnik and Geisler (2005, 2008) highlights the importance of saccadic visual consequences on gaze allocation. Their model minimizes the number of saccades needed to locate the target and plans each saccade so as to maximize the information collected across successive eye movements. This model accounts for human performance, and suggests that visual information uptake could be viewed as a reinforcing consequence controlling successive eye-landing positions—even though this assumption was not directly tested by the authors (but see Eckstein, Schoonveld, & Zhang, 2010, for preliminary results).

Here, we tested whether a visual consequence can be used to reinforce specific levels of saccadic amplitude variability in a visual search task. We developed a new gaze-contingent paradigm in which finding a target among distractors depended on the relative frequency of the current saccade. Depending on the experimental condition, the target was either displayed if movement amplitudes varied across saccades or if saccadic amplitudes were more similar to each other. For an experimental group the target was visible only if the current saccadic amplitude was rare or frequent compared to previous ones, whereas for a control group the target presentation was independent of
saccadic amplitude (Experiment 1). We reasoned that if the contingency between eye movement consequences and saccadic variability may control variability, changes in amplitude variability would occur in the experimental group and not in the control group. The goal of the second experiment was to study the transfer of this reinforcement learning to conventional visual search trials (i.e., in which the target was present in the display independently of the observer’s saccades).

Methods

General methods

To test whether changes in saccadic amplitude variability may be driven by reinforcement, we had participants complete three successive experimental conditions—a baseline, a rare-contingency, and then a frequent-contingency condition—differing in the reinforcement criteria in force.

Regardless of the actual condition, subjects were required to find a target among distractors in items-configurations (illustrated in Figure 1A) consisting in 24 black circles with a diameter of 0.4° drawn on a uniform gray background (luminance 12.5 cd/m²). The distractors contained a dash tilted 45° to the left and the target a dash oriented 45° to the right. The 24 items were located on a 10 × 10 grid covering 23.18° × 23.18°. Each item was separated from neighboring ones by at least 2.32° and thus could not be identified in peripheral vision. The items-configurations were constructed and selected as follows: First, we generated hundreds of possible configurations of 24 items randomly located on the grid. For each of these preliminary configurations, we computed the frequency distribution of the distances separating each item with the others—there were 41 possible inter-item distances ranging from 2.32° to 22.95°. For our experiments we selected some of these configurations with the most uniform distributions of inter-item distances, so that the frequency of inter-item distances did not exceed 0.05. In the two experimental conditions of Experiment 1 and in Experiment 2, we used a unique configuration which was rotated and/or inverted (Figure 1B).

During the baseline condition, subjects performed conventional visual search trials in which a target item was displayed among 23 distractors from the beginning of the trial and independently of any eye movement.

The two subsequent experimental conditions consisted of several 10-trial sessions in which reinforcement contingencies aimed at increasing (rare-contingency condition) and then decreasing (frequent-contingency condition) the variability of saccadic amplitudes. As depicted in Figure 1A (upper left), at the beginning of each trial, 24 distractors were visible in the items-configuration, and the target item was not shown. After a saccade meeting the reinforcement criteria, the target was displayed at the item location closest to the eye landing position (Figure 1A, lower right). The time needed to display a target after saccades meeting reinforcement criteria was 25 ms on average. A target presentation was followed by a tone (450 Hz, 90 ms) signaling the end of the current trial. During each saccade all items were replaced by gray squares of identical size to equalize the visual transient due to stimulus substitution across all items (see Figure 1A, middle panel, and Figure 1C).

In the rare-contingency condition, increases in saccadic amplitude variability were favored by reinforcing the least frequent amplitudes—an operant reinforcement procedure similar to the one we used to increase saccadic endpoint variability (Paeye & Madelain, 2011; see also Denney & Neuringer, 1998). Saccades whose amplitude relative frequencies were below a specific threshold triggered the display of the target at the item location closest to saccade endpoint, whereas saccades of more frequent amplitudes did not. Conversely, in the frequent-contingency condition, saccades were reinforced by displaying the target only if the current amplitude relative frequency was higher than the frequency threshold. In this condition we expected a decrease in the amplitude variability.

To compute the amplitude relative frequencies we used the distribution of the 41 inter-item distances in the items-configurations. Distance frequencies were first computed using 0.39° (10 pixels) classes; then empty classes were concatenated with the immediately preceding ones, so as to obtain 32 unequal bins of similar inter-item distances count. Relative frequencies of saccadic amplitudes were computed over a temporal moving window encompassing the previous 100 trials. For the first 100 saccades of each session, we used the last 100 trials of the immediately preceding session—in the initial rare-contingency session, we used identical relative frequencies for all bins. Figure 2A illustrates the relative frequencies and reinforcement threshold for one saccade during the rare-contingency condition. In this example, a saccade of 4° or 16° would have been reinforced, but a saccade of 6° or 17° would have not.

A trial was successfully completed as soon as 10 out of the last 50 saccades had been reinforced. For each successfully completed trial, subjects gained 0.10 Euro. The total sum earned since the beginning of the session was displayed on the screen for 0.8 s. Subjects received the total amount of money at the end of the experiment. Three sessions per day were typically recorded, separated by 10-min breaks during which subjects were free to move. On average, 12 days
Figure 1. Stimuli used in Experiments 1 and 2. (A) Schematic illustration of a rare-contingency trial. Up left: as long as saccadic amplitude did not reach the reinforcement criteria (numbered dashed lines materialize such a sequence), only distractors were visible in the items-configuration. Two of these distractors are enlarged. Middle: during each saccade (here, the sixth one) gray squares masked each item. Bottom: once the amplitude reached the reinforcement criteria the target was displayed at the item location closest to saccade endpoint. (B) Items-configurations used in the experimental conditions, rotated by 90° (rows) and inverted (columns). (C) Time course of the first three saccades of the sequence illustrated in panel A. Gray shaded areas symbolize the presence of the mask.
elapsed between the beginning and the end of the experiment.

**Experiment 1. Is contingency necessary to control variability? Yoked-control**

To probe whether contingency between eye movements and saccadic amplitude variability is necessary to control this property, we had two groups of subjects perform this experiment. In the experimental group the target presentation depended on amplitude variability whereas in the control group the target appearance was independent of saccadic amplitude. For the sake of concision, the terms “rare-contingency” and “frequent-contingency” conditions will be used to refer to the first and second conditions respectively in the control group as well.

**Baseline**

Ten trials of conventional visual search were achieved using each of the 10 items-configurations specific to this condition.

**Rare-contingency condition: Reinforcement of high amplitude variability**

The number of 10-trial sessions (ranging from 24 to 28) depended on the time needed to reach a specific performance level defined as a $U$ value (a measure of variability detailed below, Equation 1) greater than 0.80 over three consecutive sessions. Table 1 presents the number of sessions performed by each subject as

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Baseline</th>
<th>Rare-contingency condition</th>
<th>Frequent-contingency condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean sac/trial</td>
<td>$N$ sac</td>
<td>$N$ sessions</td>
</tr>
<tr>
<td>Experiment 1 Experimental</td>
<td>S1</td>
<td>12.70</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>6.90</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>14.60</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>14.00</td>
<td>140</td>
</tr>
<tr>
<td>Experiment 1 Control</td>
<td>S5</td>
<td>13.40</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>S6</td>
<td>17.70</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>S7</td>
<td>16.80</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>S8</td>
<td>8.10</td>
<td>81</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>S9</td>
<td>12.25</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>S10</td>
<td>18.25</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>S11</td>
<td>14.13</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>S12</td>
<td>11.88</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>S13</td>
<td>17.50</td>
<td>140</td>
</tr>
</tbody>
</table>

Table 1. Number of sessions in each experimental condition (“$N$ sessions”), mean number of saccades per trial (“Mean sac/trial”), and total number of saccades per condition (“$N$ sac”) following the offline data analyses, and frequency threshold used in the last session (“Last threshold”) for each subject of the two experiments.
well as the mean number of saccades per trial and the total number of recorded saccades per condition. The frequency threshold used to deliver the reinforcer was set prior to each session according to individual performance. For the very first session, we used an initial threshold of 0.05. For the subsequent sessions it was progressively adjusted in order to adapt to changes in amplitude variability (see left part of Figure 2B and Table 1).

**Frequent-contingency condition: Reinforcement of low amplitude variability**

Subjects achieved about 25 sessions (except S3 who needed only 9, Table 1) depending on the number of trials required to reach a $U$ Value (Equation 1) below 0.50 over three consecutive sessions. The frequency threshold was initially set at 0.10 and progressively raised depending on subjects’ performance (right part of Figure 2B and Table 1).

In addition, the probability of a target presentation following a saccade amplitude meeting the reinforcement criteria (that is, a saccade amplitude with a relative frequency above the frequency threshold) was also manipulated in order to keep the reinforcement rate approximately constant between the two experimental conditions. Without this manipulation the reinforcement rates might have been different in the frequent-contingency condition and in the rare-contingency condition which could affect behavioral variability independently of the operant contingencies in force, an effect that has been reported in other preparations (Grunow & Neuringer, 2002; Lee, Sturmey, & Fields, 2007; Neuringer, 2002). Furthermore, to prevent participants from moving their eyes back and forth between the same two items, the target presentation could not occur if the participant returned to the item previously fixated.

**Yoked-control group**

Each subject of the yoked-control group was paired with one subject of the experimental group. The goal of this procedure was to probe the effects of the actual reinforcement contingency by ensuring that each subject in the yoked-control group received the exact same number of reinforcers, and followed the same number of saccades, as their paired experimental subject. For this group, target presentations (and monetary gain) were independent of the actual subject’s performance and instead matched the ones obtained by one of the experimental participants. For example, if the subject S1’s 12th and 20th recorded saccades had been reinforced, the target appeared after the 12th and 20th saccades detected in the corresponding paired subject (S5) regardless of her saccadic amplitudes. The items-configurations and the number of sessions in each condition also matched the situations experienced by the paired experimental subject (Table 1).

**Experiment 2. Do the variability levels induced by reinforcement learning transfer to regular visual search? Probe trials**

The transfer of learning to conventional visual search task was tested by introducing a few changes in the procedure. Only these modifications are listed below; otherwise the procedure was as in Experiment 1. No control group was used in this experiment.

**Baseline**

Eight trials of conventional visual search were achieved using the rotated or inverted items-configuration also used in the subsequent reinforcement conditions (Figure 1B).

**Rare-contingency condition: Reinforcement of high amplitude variability**

Subjects first experienced a rare-contingency condition identical to the one in Experiment 1. At the end of this condition, two probe trials per session were randomly interleaved with the 10 experimental trials. These probe trials were identical to the baseline trials, i.e., the target was displayed at one of the 24 possible item locations independently of the observer’s eye movements. Trials ended when the target was found.

Four to six sessions mixed experimental and probe trials (depending on the time needed to acquire at least 100 saccades in these probe trials). Overall, subjects performed 14–48 sessions in this condition (see Table 1 for details on individual numbers of sessions, mean numbers of saccades per trial, total numbers of saccades per condition, and frequency thresholds).

**Frequent-contingency condition: Reinforcement of low amplitude variability**

Subjects first experienced a frequent-contingency condition identical to the one in Experiment 1. Once the $U$ value was below 0.5 for three consecutive sessions, probe trials were interleaved with experimental ones, as in the rare-contingency condition. Three to seven sessions mixed experimental and probe trials (again depending on the time needed to acquire at least 100
saccades in these probe trials). Overall, 8 to 51 sessions were performed depending on the subject (Table 1).

**Subjects**

Thirteen subjects (18–28 years of age, eight in Experiment 1 and five in Experiment 2) participated in these experiments. They were naive as to the purpose of the study, had no previous experience in oculomotor experiments, and had normal or corrected-to-normal vision. They received 50 Euros for participating, plus an additional sum depending on their performance. The typical pay per subject was 88.50 Euros on average. Participants were instructed “to find as many targets as they could”; nothing was explained regarding the variability of their eye movements and its relation to monetary gains. All experimental procedures were reviewed and approved by the Institutional Review Board, and all subjects gave informed written consent.

**Apparatus**

Stimuli were generated using the Psychophysics Toolbox extensions for Matlab (Brainard, 1997; Pelli, 1997) and displayed on a video monitor (Iiyama HM204DT, 100 Hz) at a viewing distance of 60 cm. To minimize measurement errors, the subject’s head movements were restrained using a chin and forehead rest, so that the eyes in primary gaze position were directed towards the center of the screen. Eye movements were measured continuously with an infrared video-based eye tracking system (Eyelink, SR Research Ltd., sampled at 1000 Hz) and data were transferred, stored, and analyzed via programs written in Matlab running on a Windows XP computer.

Before each experimental session we calibrated the eye tracker by having the subject fixate a set of nine locations distributed on the screen. After each trial, the subject had to look at the center of the screen for a one-point calibration check until he or she pressed a key. A subject had to look at the center of the screen for a one-second period and move the eyes randomly at least once. If the subject did not fixate the center for more than one second, the trial was repeated. After each session, the subjects were asked to rate their comfort and their satisfaction with the experiment.

**Acquisition and data analysis**

Eye movements were recorded and measured throughout each trial. We used the Eyelink online saccade detector to identify saccades onset and offset, using 30°/s velocity and 8000°/s² acceleration thresholds. For offline analyses, saccades with duration longer than 100 ms were discarded. Intra-item saccades (i.e., if the item closest to the current eye landing position was the same as the previous one) and saccades landing further than 2° away from any item location were excluded (respectively 3.75% and 3.02% of the saccades on average).

We computed U values using the following formula:

\[
U = \frac{-\sum p_n \log_2(p_n)}{\log_b(N)},
\]

where \( p \) represents the relative frequency of a bin \( n \) and \( N \) the number of bins. Here we used the 32 inter-item distance bins that determined the reinforcement criteria (cf. Figure 2A). \( U \) value reflects the likelihood of a saccade amplitude falling in each bin. If the 32 bins contain an equal number of saccades, then \( U \) equals 1, indicating the maximum entropy. Conversely if all saccadic amplitudes fall within one single bin, \( U \) equals 0.

We assessed the evolution of amplitude variability by computing \( U \) values for each saccade over a temporal moving window encompassing the previous 800 saccades for each subject. To compare probe and experimental trials (i.e., the trials in which the target appearance was contingent on variability), we computed the \( U \) values over the saccades emitted during the probe trials and over the same number of saccades from the last experimental trials. We then computed percent changes in \( U \) value between the baseline and rare-contingency conditions as follow:

\[
\text{Change}_1 = \frac{U_r - U_b}{U_b} \times 100,
\]

as well as between the rare- and frequent-contingency conditions:

\[
\text{Change}_2 = \frac{U_f - U_r}{U_r} \times 100,
\]

and between the baseline and frequent-contingency conditions:

\[
\text{Change}_3 = \frac{U_f - U_b}{U_b} \times 100,
\]

where the subscripts \( b, r \) and \( f \) correspond to the
baseline, the rare-, and the frequent-contingency conditions, respectively.

We also measured saccadic peak velocities and intersaccade times.

Measures from the different conditions were compared within each subject using bootstrap methods (Efron, Jolivet, & Hordan, 1995), via either 95% percentile confidence intervals or a priori pair-wise comparisons (two-tailed tests).

**Results**

In both experiments, manipulating the reinforcement contingencies induced changes in the level of saccadic amplitude variability. On the contrary, in the control group in which displaying the target was independent of saccadic amplitude variability, no consistent change in variability was observed. In this visual search task, seeing the target acted as a reinforcer.

**Experiment 1. Effects of contingency on amplitude variability: Yoked-control**

Figure 3A illustrates the evolutions of $U$ values over saccades for two representative subjects (S1 and S2) from the experimental group and their paired control subjects (S5 and S6). For S1 and S2, the changes had idiosyncratic time courses, but both increased in the rare-contingency condition and decreased in the
frequent-contingency condition. By comparison, changes in variability were less pronounced in the control subjects, inconsistent across subjects, and occurred in opposite directions. $U$ value was significantly reduced during the rare-contingency condition in control subject S6 and remained low throughout the experiment, whereas no consistent trend was observed for control subject S5.

Figure 3B plots the frequency distributions of the last 800 saccadic amplitudes of S1’s experimental conditions. The abscissa corresponds to the 32 classes of inter-item distances used to determine the reinforcement criteria (see Figure 2A). There was a striking difference between the two distributions, the one for the frequent-contingency condition being much more populated for small amplitudes. From 0.57 in the baseline, the $U$ value increased to 0.91 at the end of the rare-contingency condition (indicating that the distribution was nearly uniform) and returned to 0.53 at the end of the frequent-contingency condition (which corresponds to percent changes of $+59.22\%$ and $-42.10\%$, respectively).

For the paired control subject (S5, Figure 3C), the distributions in the rare- and frequent-contingency conditions were more similar and the changes in variability much smaller. From 0.57 in the baseline, $U$ value decreased to 0.55 and then increased to 0.69 in the first and second experimental conditions (with respective percent changes of $-2.85\%$ and $+23.93\%$).

Bootstrap percentile confidence intervals (Figure 4A) indicate that for all experimental subjects the $U$ value at the end of the rare-contingency condition was significantly higher than during baseline with the percent change averaging $+55.48\%$ ($SD = 10.83$). Then $U$ value decreased by $-40.76\%$ on average ($SD = 2.03$) by the end of the frequent-contingency condition to reach a value significantly lower than during baseline (average percent change between baseline and frequent-contingency condition $= -8.05\%$, $SD = 3.96$). The greater absolute percent change in variability in the rare than in the frequent-contingency condition may be partly explained by our procedure: Because the target was not presented if participants returned to the previous location, subjects could not simply alternate saccades between two items, a situation which would lead to the smallest possible $U$ value.

In the control group the changes in $U$ value at the end of the experimental conditions, respectively $-10.22\%$ ($SD = 30.43$) and $+5.77\%$ ($SD = 12.28$), were lower and not consistent across all subjects. At the end of the first experimental condition, $U$ value significantly increased for only one out of the four control subjects (S7) and no significant decrease in variability occurred at the end of the second condition.

**Experiment 2. Variability levels induced by reinforcement learning transferred to conventional visual search: Probe trials**

The main purpose of this experiment was to test whether levels of amplitude variability induced by reinforcement contingencies would transfer to probe trials in which the target was displayed from the
beginning of the trial independently of saccades. As shown on the right side of Figure 4B, the amplitude variability in the probe trials at the end of the rare-contingency condition was significantly higher than during baseline (by 38.76% on average, SD = 17.23) in all subjects. Then it decreased by the end of the frequent-contingency condition (by −29.02%, SD = 14.79) to approximately reach the baseline level—except for subject S11, whose U value decreased below its baseline level.

Importantly, the significant changes in the probe trials according to each reinforcement condition indicate that the modifications induced by reinforcement learning transferred to a regular visual search task. Comparing the probe and experimental trials (Figure 5A) revealed that transfer was however not perfect as the percent changes in amplitude variability of the conventional search trials during the rare- and frequent-contingency conditions were 11.47 and 14.47 points lower than those of the experimental trials (average differences computed over the same number of trials).

This experiment also confirmed the findings from Experiment 1 (Figure 4B). In the experimental trials U value increased by 54.30% on average (SD = 7.23) between baseline and the end of the rare-contingency condition, and decreased by −41.66% (SD = 16.62) between the rare- and frequent-contingency conditions.

It should be noted that the inter-subject differences in the U values from the frequent-contingency condition were somehow greater than in Experiment 1. Indeed, between baseline and the frequent-contingency condition, U value decreased for two subjects (S11 and S13), increased for S10, and returned to baseline level for the two other subjects (S9 and S12)—corresponding to an average percent change of −10.21% (SD = 24.86).

We tested whether these inter-subject differences could be explained by disparities in initial levels of saccadic amplitude variability. Over the two experiments the increases in variability were not correlated with baseline U Values (Figure 4C, $r = 0.080$, $p = 0.795$). Moreover, the similarity between the baseline U values of the two experiments (two-sided Wilcoxon rank-sum tests, $p = 0.286$) indicates that the observed changes in saccadic amplitude variability may not be attributed to the fact that we used different items-configurations in the baseline and experimental conditions (as was the case in Experiment 1).

**Discussion**

The present study is the first to specifically test the influence of learned contingencies on saccadic amplitude variability in a visual search task. Amplitude variability increased or decreased depending on the reinforcement criteria used to trigger the target display in the experimental group. This was not the case in the yoked-control subjects for whom displaying the target was independent of amplitude variability. This result indicates that the contingency between saccadic eye movements and their amplitude variability is necessary to control this property. Moreover, our results show that visual consequences of saccades can reinforce various levels of amplitude variability.

**Transfer of learned saccadic amplitude variability**

The effects on amplitude variability during probe trials interleaved with reinforcement trials reveal that
reinforcement learning transfers well to conventional visual search trials in which the target was displayed from the beginning of the trial. This result is important because our gaze-contingent visual foraging task itself may have been responsible for the observed changes in saccadic amplitude variability.

We are therefore confident that a true operant conditioning occurred during our procedure. However, there were differences observed between the U values in the probe and experimental trials in our second experiment that could be explained by a specific voluntary strategy induced by our procedure. Indeed when the probe trials were interleaved with experimental ones, subjects could be able to discriminate between the two trial types after the first target detection: A probe trial ended immediately after target detection, whereas an experimental trial lasted until 10 out of 50 saccades had been reinforced with a target presentation. Therefore, in this latter case the first target presentation could signal that either the rare-contingency or the frequent-contingency condition was effective until the trial ends and subjects may have adapted their search strategy accordingly.

To test this assumption (see Figure 5B), we considered three types of saccades from the last sessions of the experimental conditions: saccades acquired during the probe trials, the same number of saccades generated before the first target detection in the experimental trials (“first saccades”), and the same number of saccades made after this first target detection (“last saccades”). We reasoned that if the first target detection triggered a specific strategy, the U values of saccades made before the first target detection and during the probe trials would not differ from each other whereas both would be different from the U value of saccades made after the first target detection.

Figure 5B presents the corresponding a priori pairwise comparisons. For each subject the bars plot in the probe and experimental trials. Horizontal brackets indicate significant comparisons (two-tailed bootstrap tests, Dunn-Sidak correction, p < 0.05). In a case of a strategic change in saccade variability, there should be a significant difference between the first and last saccades of the experimental trials (rightmost small brackets) and between the probe trials and the last saccades of the experimental trials (large brackets) but not between the probe trials and the first saccades of the experimental trials (leftmost small brackets). These differences were observed in only one case out of ten: the rare-contingency condition of participant S10. The lack of consistency in the other data argues against the effects of a specific change in strategy to account for the differences between the probe and experimental trials.

**Effects of saccades’ reinforcement**

The control of target selection by alimentary reward in monkeys (Platt & Glimcher, 1999; Sugrue et al., 2004) and gain of points or money in humans (Chen et al., 2013; Navalpakkam et al., 2010; Schütz et al., 2012) is one of the arguments for regarding saccadic eye movements as operant behaviors, i.e., behaviors selected, modified, and maintained by their own consequences (Skinner, 1953, 1981).

At the neuronal level, in monkeys, saccade task paradigms in which alimentary rewards were associated with specific target directions revealed neural activity changes correlated with reward in brain structures such as the basal ganglia (for a review, see Hikosaka, Nakamura, & Nakahara, 2006), and more precisely in the caudate nucleus (Lauwereyns, Watanabe, Coe, & Hikosaka, 2002; Watanabe, Lauwereyns, & Hikosaka, 2003) and the substantia nigra pars reticulata (Sato & Hikosaka, 2002), as well as in the superior colliculus (Ikeda & Hikosaka, 2003). Moreover, dopaminergic (Bayer & Glimcher, 2005), cholinergic, and serotoninergic (see Okada, Nakamura, & Kobayashi, 2011, for a review) brainstem neurons have been found to exhibit changes in their activity depending on reward delivery after saccades. Similar neuronal reward modulations have been observed in operant learning of other behaviors (Montague, Hyman, & Cohen, 2004; Schultz, 2006, 2010).

As mentioned previously, visual consequences have been found to influence choice of manual (Berlyne, 1972) as well as saccadic responses (Collins, 2012; Schroeder & Holland, 1969). The present findings provide further evidence by demonstrating that changes in saccadic amplitude during a visual search task can be induced by visual reinforcers. They extend the results of Boot and colleagues (2006, 2009), showing that instruction and monetary reward could be used to modify visual search strategies.

Independently of saccadic amplitude variability, our reinforcement procedure could have led participants to favor specific items or areas of the displays. Indeed, they might have associated the probability of seeing the target and specific spatial locations (although the target presentation did not depend on spatial criteria). To examine this possibility, we computed the proportion of saccades following a target presentation that landed in the same quadrant as the target. We reasoned that if target presentation had an influence on saccades directed towards a specific location, this proportion should be above chance level, i.e., 25% for each of the four quadrants.

Figure 6A shows the results of this computation for each participant. The 95% confidence intervals indicate that during baseline all experimental participants but one (S1) allocated their gaze independently of the
previous target location. During the last session of the rare-contingency condition, two of the nine participants from the experimental groups looked preferentially at the quadrant where the previous target was presented. This proportion increased to eight participants out of nine at the end of the frequent-contingency condition. On average, the proportion of saccades landing in the same quadrant as the previous target, from 0.30 during baseline and 0.26 at the end of the rare-contingency condition, reached 0.55 during the last session of the experiment. In the control group the bias favoring the previously reinforced spatial location was absent in the baseline (with a proportion of 0.26) but present in both the control-rare (0.48) and control-frequent (0.52) situations.

We cannot rule out the possibility that our procedure reinforced saccades made in a specific area of the display independently of their variability. However, the proportion of small (less than 5°) saccades increased as well from 30% in the rare-contingency condition to 69% in the frequent-contingency condition in the experimental group whereas it was about 80% in both conditions for the control group. Sequences of short saccades (more likely to fall within the same quadrant) might be the best way to quickly lead to a target presentation when frequent amplitudes are reinforced. 

Figure 6. Other effects of the reinforcement of amplitude variability procedure on saccadic properties. (A) Proportion of saccades following a target presentation that landed in the same quadrant as this target, computed for each participant over the baseline condition and the last session of both experimental conditions. Error bars: 95% confidence intervals. Horizontal gray line: chance level. (B) Percent change in $U$ value as a function of percent change in inter-saccade time, for both experimental groups (blue and green symbols) and the control group of Experiment 1 (orange and yellow symbols). (C) Individual example of the relation between saccadic amplitude and peak velocity for each experimental condition (subject S10) fitted via the Equation 4 (Lebedev et al., 1996). (D) Saccade kinematics fitted a parameters at the end of each experimental condition as a function of the a parameters during baseline (symbols as in panel A).
It is therefore difficult to distinguish between effects on saccadic-landing positions and repetition of saccades of similar small amplitudes.

Moreover, several studies indicate that operant conditioning involving nonvisual consequences can cause changes in other saccadic properties as well (see Madelain, Paeye, & Darcheville, 2011, for a review). In humans, auditory stimuli have been used to reinforce various distributions of saccadic gain (Madelain, Paeye, & Wallman, 2011; Paeye & Madelain, 2011)—i.e., the ratio of saccadic amplitude to target displacement—as well as saccadic latency (Madelain et al., 2007).

Some findings suggest that visual consequences as well can influence saccadic gain, latency, and peak velocity. First, in one experiment of Madelain et al. (2011, experiment 3), saccadic gain closer and closer to an arbitrary value was reinforced by displaying the target at the eye-landing position immediately after saccades met the amplitude criteria. This reinforcement procedure induced changes in median gain similar to those obtained with an auditory consequence.

Second, it is known that saccadic kinematics may also be altered by visual consequences. In an experiment carried out by Montagnini and Chelazzi (2005), subjects had to discriminate letters displayed for one monitor refresh period beginning at the median saccadic landing time. Consequently, subjects had to reduce their saccadic latency in order to discriminate the target. The latency shortening and the increase in peak velocity observed under these conditions suggest that the ability to see the target acted as a reinforcer.

In addition it has been found that the kinematics of saccades may be altered by reinforcement even if the reinforcer is contingent on saccade direction only. In monkeys saccadic peak velocities were higher, trajectories straighter, and latencies shorter for saccades in a rewarded compared to a nonrewarded direction (Ikeda & Hikosaka, 2003; Lauwereyns, Watanabe, Coe, & Hikosaka, 2002; Nakamura & Hikosaka, 2006; Sohn & Lee, 2006; Takikawa, Kawagoe, Itoh, Nakahara, & Hikosaka, 2002; Watanabe et al., 2003). These results are further established in humans by a recent study of Collins (2012) showing that saccades made towards targets that remained visible had shorter latencies and were harder to inhibit than saccades made towards transient targets. Xu-Wilson, Zee, and Shadmehr (2009) also found that saccades had higher peak velocities and shorter durations when they were made in anticipation of a visual stimulus with a high value (human faces) than a stimulus with a low value (random pixels). Reppert and collaborators (Reppert, Choi, Haith, & Shadmehr 2012) obtained similar results and observed a progressive decrease in peak velocity following the repeated presentation of each target. This modification was not related to changes in median amplitude and was interpreted as an effect of target devaluation over trials (see also Shadmehr, Orban de Xivry, Xu-Wilson, & Shih, 2010).

Because our reinforcement procedure as well could have influenced saccades kinematics, we analyzed the inter-saccade times and peak velocities of saccades in each experimental condition. Figure 6B plots the changes in inter-saccade time against the changes in U value between baseline and the end of each reinforcement condition. In most subjects we observed longer inter-saccade times at the end of the rare- and frequent-contingency conditions than during baseline (on average, this duration increased by 35.47% and 64.90% respectively), as illustrated by the data points plotting mostly on the right of the vertical gray line.

The changes in saccadic variability could have been related to the changes in inter-saccade time. Interestingly, a positive correlation between performance and saccadic latencies has been reported before (Schütz et al., 2012, Stritzke et al., 2009) when researchers asked subjects to make saccades within a color-coded patch containing two overlapping regions of different contrasts, one associated with monetary reward and the other with losses. They interpreted the increase in saccadic latency as the time necessary for the visual system to integrate reward-related information (see also Chen et al., 2013, for a discussion). However, in our present study the only significant—and negative—correlation between changes in inter-saccade time and U value was observed in the yoked-control group (orange and yellow symbols in Figure 6B, \( r = -0.903, p = 0.002 \)). In the experimental groups, this correlation did not reach significance in the rare-contingency condition (blue symbols in Figure 6B, \( r = 0.168, p = 0.665 \)) or in the frequent-contingency condition, even when discarding the most extreme value (green symbols, \( r = -0.379, p = 0.354 \)).

To test whether our current reinforcement procedure also altered saccade kinematics, we compared baseline saccades to saccades at the end of each experimental condition. Figure 6C plots an example of these comparisons for one subject (S10); the relation between saccadic amplitude and peak velocity was fitted using the equation proposed by Lebedev, Van Gelder, and Tsui (1996):

\[
a = \frac{V}{\sqrt{A}},
\]

where \( V \) and \( A \) represent saccadic peak velocity and amplitude, respectively.

Compared to baseline, an increase in saccadic peak velocity was observed in most subjects. This increase was reflected by data points situated above the equality line in Figure 6D which plots the \( a \) parameters for each experimental condition against the baseline \( a \) parameters. On average \( a \) parameters, from 135 during
baseline, increased to 157 and 165 at the end of the rare- and frequent-contingency conditions, respectively (two-sided Wilcoxon signed-rank tests, \( p = 0.033 \) and \( p = 0.001 \)). The difference between the two experimental conditions was not significant (two-sided Wilcoxon signed-rank test, \( p = 0.331 \)).

This increase in peak velocity argues against a fatigue hypothesis that could explain the increase in inter-saccade time. In addition, faster saccades might point to an arousal effect due to reinforcement (Killeen, Hanson, & Osborne, 1978; Schultz, 2010). This arousal effect may also account for changes in saccade kinematics previously attributed to reward value (Shadmehr et al., 2010; Reppert, Choi, Haith, & Shadmehr, 2012; Xu-Wilson et al., 2009). However, our experiments were not designed to disentangle between these two possibilities, and further studies are necessary to determine the respective contribution of value and arousal to changes in peak velocity.

### Operant control of saccadic variability

Our findings confirm and extend previous results from behavioral studies showing that reinforcement contingencies can control the variability of response sequences (for reviews, see Lee et al., 2007; Neuringer, 2002, 2009). Indeed, various levels of variability in sequences of key pecks and levers presses have been reinforced in animals (e.g., Machado, 1989, 1992; Page & Neuringer, 1985) as well as in manual response sequences in humans (e.g., Hopkinson & Neuringer, 2003; Lee, McComas, & Jawor, 2002; Neuringer, 1986; Neuringer & Voss, 1993; Souza, Pontes, & Abreu-Rodrigues, 2012).

The operant control of variability has also been demonstrated in a large variety of properties of other behaviors such as inter-response time of key pecks in pigeons (Blough, 1966), forms of human constructions (Goetz & Baer, 1973), or spoken words (Lee et al., 2002). It is worth noting that variability in oculomotor responses is often viewed as the outcome of either a dedicated mechanisms of gratuitous randomization (e.g., Carpenter, 1981) or noisy accumulation of information (e.g., Ratcliff, 2001) operating at a low level. Although our current paradigm is quite different from the ones used in conventional saccade studies and could be regarded as involving higher cognitive process, our results are in agreement with previous studies on reflexive saccades showing that the spread of saccadic latency (Madelain et al., 2007) and saccadic gain (Paeye & Madelain, 2011) distributions can be modified independently of their median by manipulating saccadic consequences.

According to the operant behavior theory, variability is necessary for adaptation (Donahoe, Burgos, & Palmer, 1993; Skinner, 1981). This idea led to the concept of exploration, as opposed to exploitation, in formal reinforcement learning models (Sutton & Barto, 1998) and some experimental data are consistent with this hypothesis. For instance, animals could emit difficult-to-learn target sequences only after reinforcement contingent on variability whereas this was not the case when variability had not been reinforced previously (Grunow & Neuringer, 2002; Neuringer, 1993; Neuringer, Deiss, & Olson, 2000). Recently it has been shown that higher levels of motor variability predict faster movement adaptation in arm trajectories (Wu, Miyamoto, Castro, Ölvézcky, & Smith, 2014). Interestingly, Boot et al. (2009) found a significant correlation between the variability in saccade rate observed across various visual search tasks and overall accuracy, suggesting that those subjects exhibiting more variability in their eye movement sequences were able to adjust their search patterns—to some extent—to the task requirements.

Our results, showing that variability of saccadic amplitude may be increased by reinforcement learning, seem to diverge from those obtained by Myers and Gray (2010) who had participants searching several times in a row for a target among distractors in similar displays. They reported an increase in scan pattern similarity attributed to learning due to repeated exposure to similar stimulus arrangements. In this situation, the reinforcement contingencies would enforce the systematic relation between an eye movement sequence and the target detection at the end of every trial. This systematic reinforcement might lead to repeating the same saccade sequences, and it is well established that continuous reinforcement per se (i.e., when every response is followed by a reinforcing consequence) increases response stereotypy (Antonitis, 1951; Boulanger, Ingebos, Lahak, Machado, & Richelle, 1987; Lee et al., 2007). In these conditions, varying may be costly as it increases the risk of missing a reinforcer (Gharib, Gade, & Roberts, 2004). Our experiments provide the first evidence that the decrease in variability over repeated visual search is not systematic, but may depend on constraints set by learned contingencies.

### Conclusion

Extending previous findings suggesting that saccades are highly sensitive to rewards, we show that saccadic amplitude variability is strongly affected by operant conditioning in a visual search task and that
this learning transfers to conventional search trials. We used a new gaze-contingent visual foraging task allowing the reinforcement of specific eye movement characteristics while subjects search for a target among distractors. We suggest that saccadic variability in visual search is actively regulated, a process that sheds a new light on the importance of reinforcement contingencies for active vision and provides potential means for improving training and rehabilitation.

**Keywords:** visual search, saccades, reinforcement learning, variability

## Acknowledgments

This research was funded in part by a scholarship from the Ministry of Research (CP), Agence Nationale pour la Recherche Grants ANR-JC09 494068 (CP & LM) and ANR-13-APPR-0008 (LM), and a Fulbright Fellowship (LM).

Commercial relationships: none.

Corresponding author: Laurent Madelain.

Email: laurent.madelain@univ-lille3.fr.

Address: Université Charles-de-Gaulle Lille III Domaine universitaire du Pont de Bois, Villeneuve d’Ascq Cedex France.

## References


Herrnstein, R. J. (1961). Relative and absolute strength
Madelain, L., Champrenaut, L., & Chauvin, A. (2007). Control of sensorimotor variability by conse-


