Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model

R. Calen Walshe (r.c.walshe@sms.ed.ac.uk)
Antje Nuthmann (antje.nuthmann@ed.ac.uk)
University of Edinburgh, Psychology Department,
7 George Square, Edinburgh EH8 9JZ

Abstract

Several computational models explaining fixation durations in scene viewing (Nuthmann, Smith, Engbert, & Henderson, 2010) and in reading (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998) assume that saccade programming is completed in two stages: an initial, labile stage that is subject to cancellation and an subsequent, non-labile stage in which the program can no longer be cancelled. This distinction is motivated by findings from double-step experiments that used much simpler situations than scene viewing or reading. Here, we adopt a classic double-step paradigm to a scene-viewing context. In a Static condition targets are presented to the left or right of a central fixation cross along a horizontal axis while in a Scene condition targets are presented in a gaze contingent manner along a trajectory defined by the location of recent fixations. We found evidence in support of the claims that saccade cancellation occurs within a naturalistic scene-viewing context and that saccade cancellation can account for increases in observed fixation duration distributions. The duration of the non-labile stage was estimated to be longer in the Scene condition compared to the Static condition.

Keywords: Double-step; Scene viewing; Saccade programming; Mixed-effects modelling

Introduction

There is a long history of utilizing the double-step paradigm to explore the lower level details of the programming and execution of eye-movements (Westheimer, 1954). Classic variations of the double-step paradigm involve presenting participants with two targets along a horizontal axis with a varying inter-stimulus interval separating the two targets. For instance, in one classic study of saccade programming that utilized double-step stimuli, Becker and Jürgens (1979) had a condition in which a first target was presented at 15° to the left or right of fixation with a second target presented at 30° in the same direction at delays of 50, 100, 150 and 200 ms. The participants task was to fixate the target as quickly as possible, which meant that in order to fixate the more distal target, a saccade program initiated to the first target was put in competition with a program to the second target.

By studying the characteristics of the response pattern, the paradigm affords numerous avenues to investigate the processes underlying the programming of saccades. One method of formalizing double-step data has been to produce what is called an amplitude transition function (ATF) (Becker & Jürgens, 1979). The ATF provides a measure of the saccade amplitude resulting from the stimuli as a function of the delay (D) which measures the time elapsed between the onset of the second target step and the first measured response saccade. Therefore, in this analysis only those trials in which both targets appeared prior to the first response saccade are analysed.

D can therefore be thought of as the amount of time available to the saccadic system to reprogram an eye-movement to the second target. Frequently replicated results demonstrate that when reprocessing time in low (short values of D) saccades are typically directed towards the first target step, and when reprocessing time is high (high values of D) then saccades compensate for the updated target position and move to the second target step (Ludwig, Mildinhall, & Gilchrist, 2007).

From inspection of the ATF it is also apparent that there is a point at which the appearance of the second target step can no longer have an influence on saccade programming. This region of the distribution represents a "point of no return" in the saccade processing to the first target and as such the program to the first target is executed despite the availability of countermanding information from the second target. The point in processing at which a saccade program can no longer be modified by a second target is also referred to as saccadic dead time (SDT) and has been estimated at approximately 80 ms prior to the execution of a saccade (Ludwig et al., 2007).

The double-step paradigm has been a fruitful one in elucidating the basic properties of the occulomotor system. The principles derived from such investigations have formed the basis of several models of eye-movement control in a variety of fields. These investigations have proven particularly useful in models that attempt to explain the mechanisms that control how long aspects of the visual environment are fixated. For instance, Nuthmann et al. (2010)'s CRISP model which explains fixation durations in scene viewing, utilized a two stage saccade programming mechanism. In the first labile stage of programming a saccade could be cancelled and reprogrammed, while a program that had moved into the non-labile stage could no longer be cancelled. In the CRISP model architecture saccade cancellation acts as a causal mechanism that accounts for systematic delay in fixation durations. The theoretical dichotomy between a labile and nonlabile stage of programming was first introduced in Reichle et al. (1998)'s E-Z Reader model of eye-movement control in reading. While these models borrow the distinction from classic double-step results, it has never been formally tested within the domains to which the models apply. In the current study, a classic approach to studying double-step stimuli is adapted to a naturalistic scene viewing context.

In summary, the scene-based double step experiment has several concrete aims. Firstly, the assumption that is inherent in several influential models both in scene viewing and in reading is that delays in the latency of fixations can be

partially accounted for by the time required to cancel and reinitiate a saccade program to a novel stimulus. These models often cite basic research into saccadic programming although little work has been done to verify these findings for the more naturalistic case. Therefore, our study investigates saccade programming within scene viewing by adapting a classic double-step paradigm to the scene viewing context. Secondly, by including a classic version of the double-step paradigm as an experimental condition we are able to directly compare performance across tasks.

Method

Participants were all University of Edinburgh undergraduate students that were paid £7 in compensation for their time. Each participant was presented with 100 trials in each condition (Static vs Scene). In the scene condition 100 unique colour photographs were presented at a resolution of 800x600. Stimuli were presented on a 21-inch CRT monitor and participants were seated at a distance of 67cm from the monitor. Eye-movements were recorded with an SR Research Eyelink 1000 desktop system operating at 1000Hz. Out of the 16 subjects tested, 4 were rejected for poor data quality. Of the remaining 12 participants the mean age was 23 and 10 participants were female and 2 were male.

Experiment Overview

Double-step experiments typically involve having a participant fixate to a location while a stimulus is displayed at a distal location. At varying delays, this target is then shifted to a new location. The participant is instructed to make a fixation to the final location of this double-stepped stimuli. At short delays, the participant is frequently able to interrupt whatever processing may have been made to the first target, and instead program a saccade to the second target location. Important aspects of the saccade motor system can be derived by looking at the time course of the response. In order to investigate double-step performance in a more naturalistic environment we adapted a single experimental condition from a classic double-step experiment (Becker & Jürgens, 1979) to a context in which participants received the double-step stimulus while they were actively viewing a natural scene. Furthermore, a replication of the Static double-step condition was included as a baseline measure.

Static Condition

Participants fixated a cross located in the centre of the screen. The first target step was presented after a variable delay of between 2000-3000 ms. The first target step was presented to either the left, or to the right of the fixation cross. The target step delay (TSD), the duration elapsed between presentation of the first and second targets, was either 50, 100, 150, or 200 ms. The presentation side and delay durations were counterbalanced. Furthermore, in order to ensure that participants did not simply postpone their responses and wait until the presentation of the second target step had appeared before making a response, 20% of trials were single step trials

in which only a single target was presented to the first target location. These single steps also provide a baseline for saccade response parameters that can be compared to those on the double-step trials. The first target was always presented at exactly $7°$ on the horizontal axis and the second target was always presented at $14°$ along the same axis.
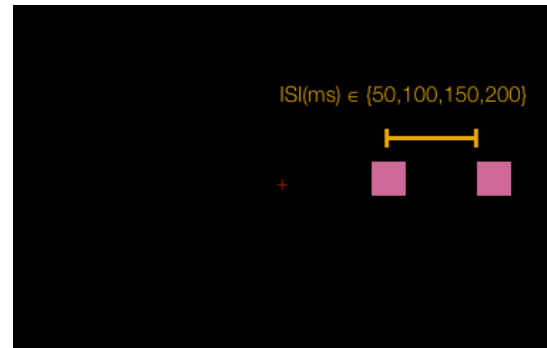
Figure 1. Target steps in the Static condition move in the horizontal plane either to the left or right of fixation cross. The first target step moves to an eccentricity of $7°$ and the second target step moves in the same direction to an eccentricity of $14°$. The second step follows the first at a delay of either 50, 100, 150 or 200 ms. In no-step trials, the first target step to $7°$ is not followed by a movement to the second target step at $14°$.

The instructions to the participant were that they were to "chase the pink box" with their eyes. Their task was to fixate the box as quickly as possible.

After 1000 ms the trial was terminated and a new trial was initiated once the fixation detection procedure had assured that the participants' gaze was directed towards the central fixation cross.

Scene Condition

The instructions to the participant in the scene condition were that they would be required to memorise a scene for a later recall test. However, this recall test was never applied. Furthermore, participants were instructed that they would see pink boxes appear while memorising the image, and that when they see these pink boxes that they should "chase the pink box" with their eyes. The memorisation task was included in order to observe performance in a more naturalistic and cognitively demanding context. All temporal characteristics of the double-stepped stimuli were identical to those of the Static condition. In the scene condition 20% of trials only had a single step at an eccentricity of $7°$ in order to avoid participants making anticipatory saccades to the second target location at $14°$. The first target step was presented once 15 saccades had been made and the scene had been viewed for at least 4 seconds. The first target step was always presented while a fixation was in progress, and this fixation could be at any possible location on the screen. A further difference from the Static double-step condition is that steps were not simply placed on the central horizontal plane as was done in the Static condition. In order to maximize the similarity be-

tween the Static and Scene conditions, while also adapting the study to a naturalistic context, the double-step manipulation trajectory was determined by the line intersecting the current fixation and the last recorded fixation. The first target was presented at 7° along this line in the same direction as the eye-movement plotted from the nth and n-1th saccade. The second target was presented at 14° on the same line. As was done in the Static condition the second target was presented in the same direction as the first (See Figure 2 for details). In circumstances such that projecting the targets along the line of presentation would result in a target being presented off the dimensions of the screen, the presentation procedure was delayed until a fixation occurred such that the presentation of the targets would not occur off screen.

The decision to place the targets along any trajectory intersected by the most recent two fixations was done for two reasons. Firstly, we wanted to control for the effect that angular changes of successive saccades has on resulting fixation durations (Tatler & Vincent, 2008). Furthermore, it is known that saccades in scene viewing are primarily executed along a horizontal axis (Nuthmann & Henderson, 2010). Due to such a bias it was expected that manipulations would primarily be placed along the horizontal axis and this was confirmed with a post-hoc analysis.

Figure 2. In the scene condition targets are presented at 7° and 14°. Unlike the Static condition targets can be placed on any axis within the image. The angle at which the boxes are presented is determined by the location of the current and previous fixations and was presented in the direction of the eye-movement. The delay between target presentations is either 50, 100, 150 or 200 ms. As in the Static condition 20% of trials consisted of only a single step to 7°.

Gaze contingent fixation detection

In order to present targets to participants within the scene condition it was necessary to accurately detect the presence of a fixation with as much temporal precision as possible. The native Eyelink gaze contingent algorithms were used in order to detect the onset of fixation. Once the conditions for presenting the first target had been met, and the Eyelink detected a fixation, the first target was presented to the participant. Delays in the online detection of fixations resulted in the targets being presented after the onset of fixation at a delay (ms) of

$\mu = 45.2; \sigma = 19.0$.

Results

Amplitude transition function in the Scene condition

The aim of the first analysis is to provide evidence in support of the hypothesis that saccade cancellation does indeed occur within a more natural scene viewing context than is typically studied with double-step stimuli. In order to construct an ATF, only trials in which both the first and second target steps occurred during a single critical fixation were analysed. Trials were also rejected when the response saccade was not made in a direction consistent with the target steps. Therefore, in the Scene condition 33 trials were removed due to movement of the eyes prior to presentation of the second target, and 6 were removed due to detection of a misdirected saccade. In the Static condition 22 and 2 trials were removed, respectively.

The amplitude transition function for the scene condition was constructed by fitting a four parameter logistic function with a form:

$$y = a + \frac{b-a}{1 + e^{c(d-x)}} \qquad (1)$$

where a represents a lower bound for the sigmoid, b represents an upper bound, c scales the response to x (Delay) about the midpoint and d is the inflection point of the sigmoid.

Model fitting   The data were fit with a nonlinear mixed effects model (NLME) (Pinheiro & Bates, 2000). Firstly, we added fixed effects which allow us to directly estimate the effect of experimental condition (Static vs Scene) on the parameters of the nonlinear response function described in (1). Secondly, random effects were included in the model in order to reduce the effect of unreliable differences between participants due to unbalanced observations and individual variability in task performance.

In the analysis of the Scene condition we fit a model which included the effect of only a single condition on the parameters a, b, c and d. Random effects of participant on the parameters a, b, c and d were also included in the model. For the comparison between the Static vs Scene conditions the model was extended to include a fixed effect of condition on the four model parameters. The R statistical programming language (R Development Core Team, 2008) and the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2013) were used to conduct the analysis.

Effect of Scene on model parameters   The parameters of the best fitting model are reported in Table 1. From the scatter the typical ATF evoked by double-step stimuli is observed with a characteristic sigmoidal shape. The horizontal dotted lines indicate the location of target 1 at 7° and target 2 at 14°. Furthermore, the scatter confirms that within the Scene condition lower values of D are typically associated with saccades directed towards the location of the first target step whereas at larger values of D reprogramming occurs and saccades are directed towards the second target location.

Table 1. Mixed effects model parameters

Effect	*M*	*SE*	*t*	*p*
a	6.47	.160	40.48	(< 0.01)
b	12.35	.351	35.23	(< 0.01)
c	0.12	.027	4.33	(< 0.01)
d	113.18	5.46	20.72	(< 0.01)

Summary of the estimated values of the fixed-effects parameters along with their means (M), standard errors (SE), *t* and *p* values, units of the parameters are reported in milliseconds. The parameters *a* and *b* are respectively the lower and upper asymptote of the sigmoid while *c* scales the response about the midpoint and *d* is the inflection point of the sigmoid.
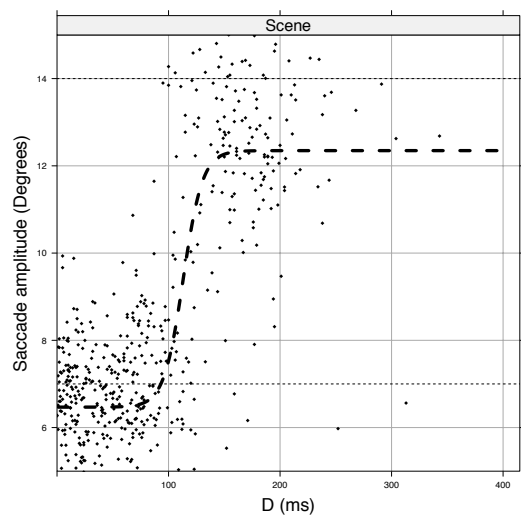
Figure 3. Amplitude transition function constructed from responses in the Scene double-step condition. *D* represents the amount of time elapsed between the onset of the second target stimulus and detection of the first response saccade. Horizontal lines represent the locations of target steps 1 and 2.

The results of the model fits estimated that the lower and upper bounds of the saccadic endpoints were $6.47°$ (*SE* = .160, *t* = 40.48) and $12.35°$ (*SE*= 0.35, *t* = 35.23) respectively. While there was undershoot for saccades targeting both the initial and final target steps, the undershoot to the final step was larger ($0.53°$ vs $1.65°$).
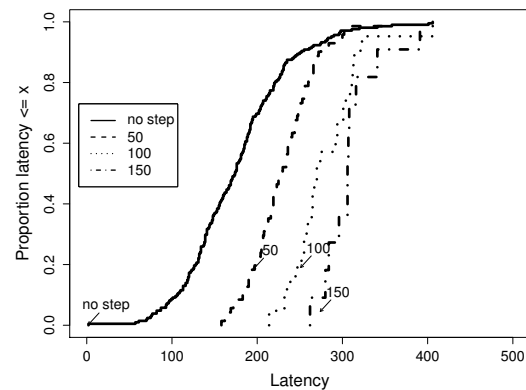
Cumulative distribution function of saccade latencies

Saccade programming latencies were compared on trials in which there was no target step (no-step) with trials in which there was a target step (50, 100, 150, 200 ms). This analysis aims to investigate whether trials in which a saccade was reprogrammed from the first target to the second target require longer latencies when compared to no-step trials in which no such reprogramming occurs.

In no-step trials saccade latency was calculated as the

elapsed time (ms) from the appearance of the $7°$ target and the first observed response saccade. For trials of all other delays, latencies were analysed for saccades in which the first response saccade occurred after the second target step had appeared and in addition that the first response saccade compensated for the second target step. Compensation was defined such that the saccadic endpoint was within a distance $2°$ of visual angle from the second target location. Latency was calculated as the elapsed duration between the onset of the first target step and the onset of first response saccade.

A cumulative distribution function (CDF) was fit to compare latencies between delays of different lengths. Latencies from 200 ms TSD trials were excluded as too few compensated saccades were observed. Latencies for compensated saccades are clearly longer than those of saccades elicited by no-step stimuli indicating that in order to incorporate the second target step into the response, increased latencies are required. Furthermore, we observe that as the TSD increases a corresponding increase in latency is also observed.

Figure 4. Cumulative distribution function of latencies at target step delays of 50, 100, 150 ms compared to the no-step latencies. The no-step latencies are constructed from latencies on trials in which there was only a step to the first target response.

Comparison of Static vs Scene conditions

While the primary aim was to provide evidence that results from static double-step conditions generalize to a more dynamic scene based context, our dataset also offers an opportunity to directly compare performance differences between the Static and Scene conditions. A description of the model used to fit the Static vs Scene data can be found in the *Model fitting* section of the Scene only analysis.

Comparing the scatter in the Static vs Scene condition (Fig. 5) it is apparent that there is considerably more variability in the data that comprise the ATF in the Scene as compared to Static condition. Due to the more dynamic nature of the Scene task this is to be expected. For instance, in the Static condition participants stay fixated on a central cross while they wait for the target stimuli to appear. It is therefore likely that any anticipatory processes preparing future eye-movements are suppressed. In contrast, during the Scene con-

Table 2. Mixed effects model parameters

Effect	M	SE	t	p
a (Intercept)	6.63	0.14	48.62	(< 0.01)
a (Scene)	-0.17	0.16	-1.1	(= 0.27)
b (Intercept)	13.0	0.11	119.9	(< 0.01)
b (Scene)	-0.74	0.20	-3.82	(< 0.01)
c (Intercept)	0.14	0.02	7.57	(< 0.01)
c (Scene)	-0.05	0.02	-2.18	(= 0.03)
d (Intercept)	76.14	2.30	32.91	(< 0.01)
d (Scene)	33.86	2.51	13.44	(< 0.01)

Summary of the estimated values of the fixed-effects parameters along with their means (M), standard errors (SE) and t and p values, units of the parameters are reported in milliseconds. The parameters a and b are respectively the lower and upper asymptote of the sigmoid while c scales the response about the midpoint and d is the inflection point. The intercept indicates the estimated parameter in the Static condition, while (Scene) indicates the influence of condition Scene.

dition participants are actively engaged in search, the display and measurement of their double-step response is likely to incorporate processes involved in preparing an eye-movement prior to the presentation of the double-step stimuli.
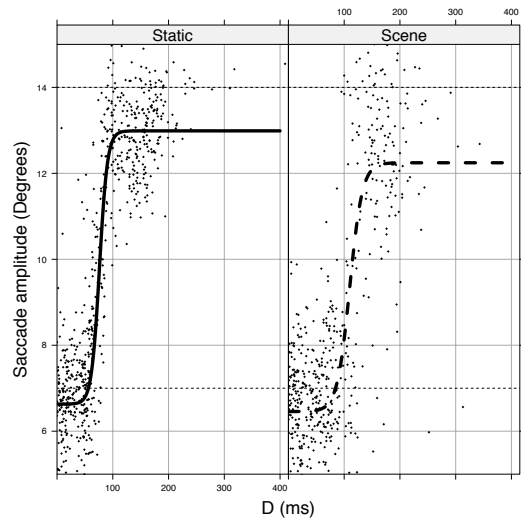
Figure 5. Comparison of the amplitude transition functions in Static vs Scene conditions. D represents the amount of time elapsed between the onset of the second target stimulus and detection of the first response saccade. Horizontal lines represent the locations of target steps 1 and 2.

Effect of experimental condition (Scene vs Static) on model parameters Model parameters are summarized in Table 2. The lower and upper bounds of the fitted functions measure the saccadic endpoints of responses to either the first (lower bound) or second (upper bound) targets. No effect of condition was observed on the lower bound ($t = -1.1$, $p = 0.27$) however, there was an effect of condition on the

amplitude of responses targeting the final location ($t = -3.82$, $p < 0.01$), indicating greater undershoot in the Scene condition.

A final observation can be made regarding the markedly slower compensation response in the Scene condition. The ATF in the Scene condition appears to be shifted to the right, and this reflects that increasing values of D are required to make a response of corresponding amplitude to that of the Static condition. Furthermore, we observe a significant effect of condition on d (inflection point) ($t = 13.44$, $p < 0.01$) supporting the observation that responses of comparable amplitude require longer values of D in the Scene condition as compared to the Static condition.

Ludwig et al. (2007) have referred to saccadic dead time (SDT) as the last moment at which a new stimulus can modify a saccade program currently under preparation. They describe that the SDT may be extracted from the ATF by estimating the point at which the compensation function begins to incorporate the location of the second target step. The SDT was extracted from our ATF by deriving the point on the curve which represents a cumulative increase of 5% from the lower asymptote. We estimated this point in the Static condition as 55 ms and in the Scene condition as 77 ms.

Discussion

The research question that this paper addresses is whether established results utilizing double-step stimuli to explore saccade programming can be extended to scene viewing. It has been argued that the ATF constructed from double-step responses provides evidence for a distinction between a labile and nonlabile stage of saccadic programming. A target stimulus is only able to modify the current goals of a saccade while it is in the labile stage of programming and can no longer have an influence once the program becomes nonlabile. The sigmoidal shape of the ATF (Fig. 3) reveals that when the target stimulus is presented shortly before the saccade (low values of D) that processing had reached the nonlabile threshold and therefore had no influence on the resulting saccade. When the second target is presented in earlier stages of saccade preparation (higher values of D) we see saccades that compensate for the second target location due to programming still being within the labile stage. These results have been previously established in double-step studies utilizing static conditions (Becker & Jürgens, 1979; Ludwig et al., 2007), and our study provides evidence for an analogous process occurring within a condition more akin to naturalistic scene viewing.

Cancellation has also been suggested as a causal mechanism for the systematic increase in observed fixation durations. For instance, in the CRISP model, saccade cancellation accounts for the increase in fixation durations that is observed directly following a delay of stimulus onset (see Figure 7, Nuthmann et al., 2010 for details). The CDF (Fig. 4) illustrates that latencies are increased in trials in which a reprogramming of a saccade is likely to have occurred. However, caution must be taken when assigning a causal interpre-

tation to the role of saccade cancellation in observed latency increases. The TSD trials analysed are specifically those for which a saccade was not executed prior to the appearance of the second saccade target. Therefore, we expect to see a complementary increase in latency alongside increases in TSD. One possibility is that increased latencies are observed specifically because compensated saccades are those in which the programming to the first target progressed slowly enough to wait out and incorporate the appearance of the second target. While this analysis does provide confirmatory evidence that saccade cancellation is consistent with increased fixation durations, it does not necessarily shed light on the causal connection between cancellation and increased latencies.

The comparison between the Static and Scene condition also indicate the presence of several notable differences. Ludwig et al. (2007) analyse a concept termed saccadic dead time (SDT) which corresponds closely to the concept of a nonlabile stage of programming. The SDT corresponds to the last point in time at which a saccadic eye-movement may be modified. We estimated SDT as 55 ms in the Static condition and 77 ms in the Scene condition. Differences in the SDT across experimental conditions have been observed in prior work (Ludwig et al., 2007). An important implication of this result with regard to models of gaze control in naturalistic scene viewing is that it provides an empirical bound for the duration of the nonlabile stage. In CRISP for instance, a mean duration of 40 ms was assigned to the duration of the nonlabile stage. This value was determined from classic double-step results conducted under static conditions. This value is also roughly consistent with the duration of the nonlabile stage estimated in our own Static condition (55 ms) but represents an underestimate when compared to the nonlabile duration in the Scene condition. As CRISP is a model of fixation durations in scene viewing it is likely that the estimated mean duration of the nonlabile stage in our Scene condition represents an improvement over the corresponding Static estimate.

It should be noted that the comparison reported here may still reflect important differences not solely attributable to the influence of scene processing. For instance, the Scene but not Static condition double-step targets were presented on any axis. Future work may consider including a task in which performance in our Scene condition is compared directly to a similar task but one in which the scene is replaced by a noise filtered image.

Further comment is warranted on the applicability of the data reported here to models such as CRISP that claim a causal interpretation for saccade cancellation in observed fixation duration delays. In CRISP, when a saccade program is within the labile stage of programming a cancellation signal may interrupt the current program. The time required to reinitiate a new saccade program results in a delay to the current fixation duration. In an alternative formulation aimed at explaining saccade latencies within a Static double-step context, Camalier et al. (2007) suggest that cancellation occurs

due to a race between a process initiated to execute a saccade (GO) and process initiated to cancel that saccade (STOP). While there is some similarity in the assumptions between these two models the race model does not insist on a non-labile/labile dichotomy. Rather, the race model accounts for double-step performance with reference to the timing of the GO and STOP accumulation processes. In order to further explore the role of saccade cancellation it may be of interest to directly compare the predictions of the saccade programming mechanisms in the CRISP model with those of the race model described in Camalier et al. (2007).

References

Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, *19*(9), 967-983.

Camalier, C., Gotler, A., Murthy, A., Thompson, K., Logan, G., Palmeri, T., et al. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, *47*(16), 2187-2211.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777-813.

Ludwig, C. J. H., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology*, *97*(1), 795-805.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 1-19.

Nuthmann, A., Smith, T., Engbert, R., & Henderson, J. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382-405.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2013). nlme: Linear and nonlinear mixed effects models [Computer software manual]. (R package version 3.1-108)

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org (ISBN 3-900051-07-0)

Reichle, E., Pollatsek, A., Fisher, D., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125-157.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18.

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology*, *52*(6), 932-941.