Setting things straight: a comparison of measures of saccade trajectory deviation

Article

Accepted Version

Tudge, L., McSorley, E. ORCID: https://orcid.org/0000-0002-2054-879X, Brandt, S. A. and Schubert, T. (2017) Setting things straight: a comparison of measures of saccade trajectory deviation. Behavior Research Methods, 49 (6). pp. 2127-2145. ISSN 1554-351X doi: https://doi.org/10.3758/s13428-016-0846-6 Available at https://centaur.reading.ac.uk/68658/

To link to this article DOI: http://dx.doi.org/10.3758/s13428-016-0846-6

Publisher: Psychonomic Society

Setting things straight:
A comparison of measures of saccade trajectory deviation

Luke Tudge[1,2], Eugene McSorley[3], Stephan A. Brandt[4], & Torsten Schubert[1,5]

1. Institut für Psychologie, Humboldt-Universität zu Berlin, Germany
2. Berlin School of Mind and Brain, Germany
3. School of Psychology and Clinical Language Sciences, University of Reading, UK
4. Charité Universitätsmedizin Berlin, Germany
5. Martin Luther-Universität, Halle, Germany

Abstract

In eye movements, saccade trajectory deviation has often been used as a physiological operationalization of visual attention, distraction, or the visual system's prioritization of different sources of information. However, there are many different ways to measure saccade trajectories and to quantify their deviation. This may lead to non-comparable results and poses the problem of choosing a method that will maximize statistical power. Using data from existing studies and from our own experiments, we use Principal Components Analysis (PCA) to carry out a systematic quantification of the relationships among eight different measures of saccade trajectory deviation and their power to detect the effects of experimental manipulations, as measured by standardized effect size. We conclude: 1) that the *saccade deviation* measure is a good default measure of saccade trajectory deviation, as it is somewhat correlated with all other measures and shows relatively high effect sizes for two well-known experimental effects; 2) more generally, measures made relative to the position of the saccade target are more powerful; 3) measures of deviation based on the early part of the saccade are made more stable by an eyetracker with a high sampling rate. Our recommendations may be of use to future eye movement researchers seeking to optimize the design of their studies.

Introduction

When a new object appears in our field of view, we may make a quick eye movement (a saccade) to bring our gaze to that object. During these saccades, the path that our gaze follows across our field of view is rarely a straight line from our current point of regard to the location of the new object. Instead, saccades describe a curved path, and do not always land exactly on target (Erkelens & Sloot, 1995; Viviani, Berthoz, & Tracey, 1977). This deviation is systematically influenced by the presence of other objects that we have not chosen to look at, termed distractors (for reviews see Van der Stigchel, 2010; Walker & McSorley, 2008). This phenomenon may be termed saccade trajectory deviation.

A widely-accepted explanation of saccade trajectory deviation is that it occurs because the visual system prepares eye movements to both the target and the distractor, and the resulting eye movement is an average or combination of the two different planned movements at the moment when the saccade is initiated (McPeek & Keller, 2001; McPeek, Han, & Keller, 2003; Port & Wurtz, 2003; Tipper, Howard, & Paul, 2001; White, Theeuwes, & Munoz, 2012). To the extent that the planned eye movement to the distractor has not been fully suppressed by the time the saccade is executed, the trajectory of the saccade will deviate towards the distractor. Conversely, deviation away from the distractor may reflect an ‘overinhibition’ of the planned eye movement to the distractor (McSorley et al., 2006).

Saccade trajectory deviation provides a convenient quantification of the allocation of attention to the distractor. By varying the content of the distractor or of the target, and by varying the conditions under which participants view the two objects, we may learn what priorities and strategies the visual system employs. Saccade trajectory deviation has been widely used in this way as an operationalization of attention and cognitive control in investigations of diverse phenomena, such as phobias (McSorley & Morriss, 2015), the processing of word meaning (Weaver, Lauwereyns, & Theeuwes, 2011), emotion (McSorley & van Reekum, 2013), social behavior (Laidlaw, Badiudeen, Zhu, & Kingstone, 2015), cognitive decline in the elderly (Campbell, Al-Aidroos, Pratt, & Hasher, 2009), and participants’ preparedness for the task (Tudge & Schubert, 2016).

When studying saccade trajectory deviations, it is necessary to quantify the extent of a saccade’s deviation. There exists no single agreed-upon method for doing so. Rather, different studies have quantified deviation in different ways (for an overview, see Van der Stigchel, Meeter, & Theeuwes, 2006). If these different measures reflect slightly different aspects of saccade planning, or if some measures are better suited than others to detect the effects of experimental manipulations, then studies using different measures may not be easily comparable, or may in fact be drawing conclusions about different underlying phenomena. Our aim in the present study is to systematically compare different measures of saccade trajectory deviation, in order to find out which of them are likely to reflect the same underlying phenomenon, and which are most sensitive to certain experimental manipulations. We hope that this information will help future researchers in choosing an optimal measure for a planned study, and help to better compare the findings of studies that use different measures.

There are several different features of a saccade trajectory that might reflect its apparent deviation from a straight path. A widely-cited review of research with saccade trajectory deviations lists eight methods of measuring deviation (Van der Stigchel et al., 2006). In the present study, we compare these eight measures. It is therefore important to describe them briefly before continuing. The measures are also summarized in **Table 1**, and illustrated in **Figure 1**.

	Description	**Classification**	**Units**
overall direction	Angle between straight line from saccade start to saccade end, and straight line from saccade start to target position.	late, target-based	angular degrees
saccade deviation	Mean of all angles formed by lines drawn from saccade start to each sample point in the saccade compared to straight line from saccade start to target position.	full-sample, target-based	angular degrees
overall initial direction	As *saccade deviation*, but using only first sample point occurring 10 ms after saccade start.	early, target-based	angular degrees
maximum curvature	Maximum perpendicular distance of saccade from straight line from saccade start to saccade end, divided by saccade amplitude (i.e. length of saccade).	subsample, endpoint-based	dva
area curvature	Area between saccade trajectory and straight line from saccade start to saccade end, estimated using midpoint rectangles.	full-sample, endpoint-based	dva^2
initial direction	As *overall initial direction*, but with angle calculated relative to straight line from saccade start to saccade end instead of to target position.	early, endpoint-based	angular degrees
initial average curvature	As *maximum curvature*, but average of perpendicular distances to samples occurring within 10 ms of saccade start.	early, endpoint-based	dva
quadratic curvature	Saccade samples are standardized and rotated so that straight line from saccade start to saccade end is horizontal and runs from -1 to 1. Quadratic polynomial function is fitted to saccade trajectory. *Quadratic curvature* is the coefficient for quadratic term of fitted function.	full-sample, endpoint-based	dva

Table 1: Summary of saccade measures. **Classification** column categorizes measures according to the distinctions drawn in the main text of the article (target-based, endpoint-based, full-sample, subsample, early and late). **Units** column gives the units of measurement, where ‘angular degrees’ are degrees of rotation on the two-dimensional surface of the computer screen, and ‘dva’ are degrees of visual angle.

Overall direction (OD) is the angle between a straight line from saccade start to target position and a straight line from saccade start to saccade end. It measures the extent to which a saccade lands to one side of its target, and does not take into account any part of the saccade apart from its landing point.

Saccade deviation (SD) is the mean of all the angles formed between a straight line from saccade start to target position and straight lines from saccade start to each sample within the saccade. Like *overall direction*, it measures the extent to which the saccade deviates to one side of its target, but averaged over the entire trajectory.

Overall initial direction (OID) is the angle between a straight line from saccade start to target position and a straight line from saccade start to a point 10 ms after saccade start (i.e. early in the saccade). Again, it measures deviation relative to the target, but does so only for the earliest part of the saccade.

Maximum curvature (MC) is the maximum perpendicular distance of the saccade trajectory from a straight line from saccade start to saccade end. It measures the curved shape of the trajectory.

Some previous studies have standardized *maximum curvature* by dividing it by saccade amplitude (Doyle & Walker, 2001). This is intended to correct for the fact that longer saccades have more space within which to describe a larger curve. We also follow this standardization procedure in our analyses.

Area curvature (AC) is an estimate of the area between the trajectory of the saccade and a straight line from saccade start to saccade end. Different studies have estimated this area in slightly different ways. In all methods, rectangles drawn along the straight line from saccade start to saccade end and located between saccade samples are used to approximate the area of the curve. These rectangles may extend either to each sample (see e.g. Figure 1 in Ludwig & Gilchrist, 2002) or to a point half way between each sample and the previous sample (Walker et al., 2006). We have used the latter procedure in our analyses (see **Figure 1**, right panel). Like *maximum curvature*, this measure is often standardized to saccade amplitude (Walker et al., 2006), and we follow this standardization procedure in our analyses.

Initial direction (ID) is similar to *overall initial direction* in that it measures an angle to a saccade sample 10 ms into the saccade. The difference is that this angle is measured relative to a straight line from saccade start to saccade end and not to the target position.

Initial average curvature (IAC) is similar to *maximum curvature*. It measures the perpendicular distance of saccade samples from a straight line from saccade start to saccade end, but instead of the maximum such distance, it is the mean of distances to samples within the first 10 ms of the saccade. This measure is a variant of a measure that has been termed simply *initial average*. There has been some inconsistency of terms in the literature on saccade trajectory deviations regarding *initial average*. To our knowledge, the first occurrence of a measure with this name is in the work of Sheliga and colleagues (e.g. Sheliga, Riggio, Craighero, & Rizzolatti, 1995). The authors describe a measure that averages perpendicular distances from a straight line from saccade start in an absolute direction (up, down, left or right, depending on where the target is located). Later, Ludwig & Gilchrist (2002) describe a measure called *initial direction,* and reference the description by Sheliga et al. (1995), but in fact describe a very slightly different process of calculation, using perpendicular distances from a straight line from saccade start to saccade end. In the present study, we follow the method from Ludwig & Gilchrist (2002), but use the novel term *initial average curvature* to avoid confusion with the slightly different method described as *initial average* in Sheliga et al. (1995). To avoid further confusion, it is also important to note here that the term *initial average* also appears in Van der Stigchel et al. (2006), with another very slightly different method of calculation. The authors describe *initial average* as the average of angles between the saccade trajectory and a straight line from saccade start to saccade end. We have not used this method of calculation in the present study.

Quadratic curvature (QC) is calculated by fitting a quadratic polynomial to the saccade samples after normalizing the amplitude of the saccade onto a scale from -1 to 1. The quadratic coefficient of the fitted curve is the *quadratic curvature*, and measures the curved shape of the trajectory (Ludwig & Gilchrist, 2002).

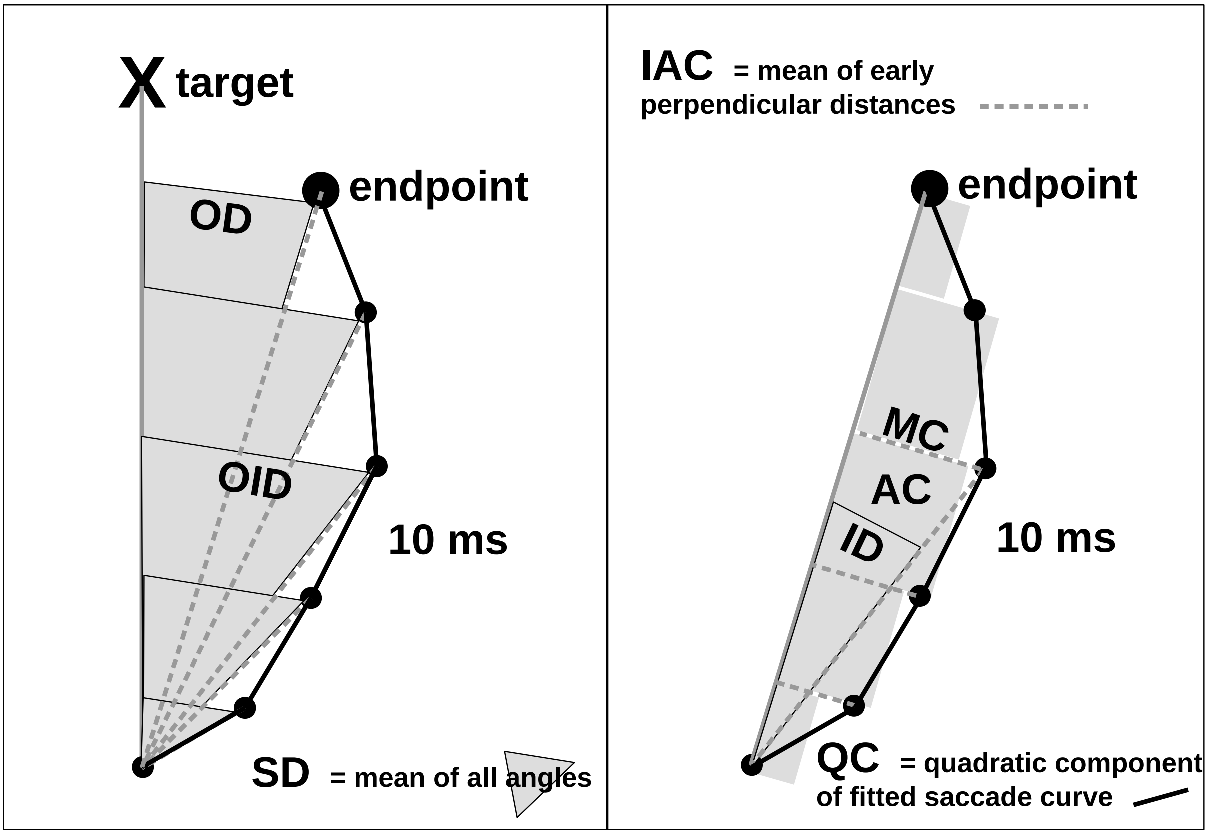

Figure 1: Measuring saccade trajectory deviation. **Left panel**: Target-based measures; all angles calculated relative to straight line to target. OD is angle for saccade endpoint. OID is angle for first sample after 10 ms. SD is mean of angles for all gaze samples. **Right panel**: Endpoint-based measures; all angles / perpendicular distances calculated relative to straight line to endpoint. ID is angle for first sample after 10 ms. MC is distance to furthest sample point. AC is estimated area under saccade trajectory. IAC is average of distance for sample points earlier than 10 ms. QC is quadratic coefficient of estimated normalized saccade trajectory.

In order to give some structure to this list of measures, we have classified them according to three features. The first is the choice of ideal straight line to which the saccade trajectory is compared. *Overall direction*, *saccade deviation*, and *overall initial direction* are calculated relative to a straight line from the start of the saccade to the correct target position. We term these ‘target-based’ measures. The other measures are calculated relative to a straight line from the start of the saccade to the end of the saccade. We term these ‘endpoint-based’ measures. These two categories have sometimes been termed ‘deviation’ and ‘curvature’, respectively. We have not followed this convention here, since the term ‘deviation’ is also commonly used to refer to the overall notion of distortions of saccade trajectory, both target-based and endpoint-based (e.g. in McSorley et al., 2006), and it is in this more general sense that we also intend the term ‘deviation’ in this article.

Target-based measures quantify the extent to which the saccade misses its target, whereas endpoint-based measures quantify the curved shape of the saccade trajectory, irrespective of whether it is on target or not. It is in principle possible that these two types of measure be independent of one another; a saccade may be on target but have reached the target via a very curved trajectory, or conversely a saccade may be a long way off target but have an entirely straight trajectory. However, there is some evidence to suggest that this independence is not realized in practice. McSorley, Haggard, and Walker (2004) found that *overall direction*, a target-based measure, is positively correlated with *area curvature*, an endpoint-based measure, though only for saccades that are directed upwards, not downwards (see Figure 6 in McSorley et al., 2004). Similarly, Van der Stigchel, Meeter, and Theeuwes (2007) found that *overall direction* and *initial direction* are strongly positively correlated.

The second feature concerns the amount of information that the measure makes use of. An eye tracking device samples gaze position at many different points along the trajectory of the saccade. *Saccade deviation*, *area curvature*, and *quadratic curvature* make use of all these samples, by averaging or integration. We term such measures ‘full-sample’ measures. The other measures make use of only one sample or a subset of samples that are deemed to be of particular importance, for example the first few samples after saccade start, the endpoint of the saccade, or the point at which deviation reaches a maximum. We term these ‘subsample’ measures.

It has been argued that full-sample measures are preferable, because combining multiple samples may help to average out measurement error in the eye tracking system (Ludwig & Gilchrist, 2002). Although plausible on theoretical grounds, to our knowledge this assertion has not been tested. If it is the case that different features of a saccade reflect different underlying phenomena, then it may nonetheless be preferable to focus only on a subset of samples, if these are the samples most likely to reflect the phenomenon of interest. In addition, it is not necessarily the case that measurement error is of the same magnitude throughout a saccade. For example, gaze might be measured more noisily while the eye is in motion than when it has stopped moving, which could make *overall direction* less noisy than full-sample measures despite being based on only one sample.

The third distinction is between ‘early’ and ‘late’ measures of saccade trajectory deviation. An early measure of deviation is a type of subsample measure that takes its subsample from the beginning of the saccade. These measures therefore reflect the state of the saccade shortly after initiation, before any corrective processes have brought the trajectory closer in line with the target (Van der Stigchel et al., 2006). *Overall initial direction*, *initial direction*, and *initial average curvature* are early measures, since they use only samples within the first 10 ms of the saccade. The use of 10 ms as a cutoff for the early part of a saccade is an arbitrary choice, and its appropriateness will depend on the expected duration of saccades in a given experiment. Some previous studies have used 8 ms (e.g. Ludwig & Gilchrist, 2002), 10 ms (e.g. Sheliga, Riggio, Craighero, & Rizzolatti, 1995), 12 ms (e.g. Van der Stigchel & Theeuwes, 2005) or 20 ms (e.g. Van der Stigchel & Theeuwes, 2006).

Conversely, late measures take their subsample from the end of the saccade. Only one measure, *overall direction*, is explicitly based on a subsample taken from the end of the saccade, and as such is the only strictly late measure. Many measures are neither early nor late, either because they are full-sample measures or because they are based on a subsample that may occur anywhere during the saccade, for example the *maximum curvature*.

The fact that so many different measures are in use to quantify saccade trajectory deviation raises two potential problems. The first is the issue of comparability. If different studies on similar topics make use of different dependent measures, it remains unclear to what extent their findings are comparable. Studies of saccade trajectory deviation may in fact be investigating different phenomena if they employ different methods of measurement. Saccade trajectory deviations may be the outcome of a process with several different components, such as selecting the target, inhibiting the distractor, deciding when to execute the saccade, and correcting the saccade trajectory ‘online’, i.e. while underway (Quaia, Lefèvre, & Optican, 1999). Different features of a saccade trajectory may be measuring some of these components but not others. For example, early measures are made before much online correction has taken place, and may therefore reflect more closely the initial amount of attention allocated to the distractor, whereas late measures may additionally reflect the success or failure of online correction.

If the different measures are strongly correlated with one another, then we may be more confident that they all reflect broadly the same phenomenon. One previous study reported the correlations of some measures, and found these to be generally high (between .70 and .98; Ludwig & Gilchrist, 2002). However, this study only investigated endpoint-based measures, and correlation does not of itself guarantee that the measures will respond identically to experimental manipulations.

In order to more systematically address the problem of comparability, we employ Principal Components Analysis (PCA) with all eight measures. PCA reduces a set of correlated variables to a smaller number of underlying components that describe most of the variance in the data (Hotelling, 1933). If it can be established that particular subsets of measures are likely to reflect the same underlying phenomenon, then we may be more confident in comparing the results of studies using different measures from within one subset. Conversely, where discrepant findings arise, we may be able to explain these as a consequence of having employed two different measures of deviation that may reflect different underlying phenomena.

The second problem is the issue of selecting a measure that maximizes statistical power. All else being equal, we wish to use a measure that gives us the best chance of detecting the effects of our experimental manipulation. The power of a particular measure to detect a particular effect depends on the magnitude of the effect on that measure, relative to the measure's variance. To quantify the power of each measure, we use the standardized effect size generalized eta-squared (η^2_G), as a metric that is comparable across different study designs (Olejnik & Algina, 2003). If it can be established that a certain measure reflects more clearly the effects of experimental manipulations, then that should be the preferred measure for future studies.

Saccade trajectory deviations have been used as the dependent measure for a wide variety of experimental manipulations. Since it is not feasible to investigate effect sizes for all of these manipulations, we instead restrict the investigation to two well-established experimental paradigms. The first is arguably the simplest target-distractor paradigm possible, one in which a target and a distractor are presented simultaneously. The participant's task is to make a saccade to the target as quickly as possible. The target and the distractor are distinguishable only by virtue of their shapes (e.g. one is a cross and the other a circle, as in McSorley et al., 2006). In this paradigm, the effect of interest is the negative relationship of saccade trajectory deviation to saccade latency. Saccades that occur very soon after the stimuli appear tend to deviate more towards the distractor, whereas saccades that occur later show less deviation towards the distractor, and may even deviate away from it (McSorley et al., 2006).

The negative relationship between deviation and latency is typically explained as the result of competition between target and distractor, as described above. When target and distractor appear, the oculomotor system generates planned eye movements to both of them. If a saccade is initiated while both of these eye movement plans are still active, the resulting eye movement trajectory will be something of an average between the two plans, and will therefore deviate towards the distractor. Only after some time is knowledge of the task brought to bear, with the result that the plan for an eye movement to the distractor is gradually inhibited. So the later the saccade is executed, the less it will deviate towards the distractor (McSorley et al., 2006, Van der Stigchel, 2010).

It is particularly important to establish which measure is most sensitive to this basic effect of saccade latency. This is because latency is often investigated as a modulating factor in studies involving additional variables of interest, and in many studies the principal finding is an interaction of saccade latency with this additional variable. For example, elderly people show a more shallow slope relating deviation and latency than do younger people (Campbell et al., 2009), and some manipulations, such as the physical salience of the distractor, are only apparent at short saccade latencies (van Zoest, Donk, & Van der Stigchel, 2012), whereas others, such as the social relevance of the distractor, are only apparent at longer saccade latencies (Laidlaw et al., 2015).

The second paradigm in which we measure effect sizes is one that is designed to investigate the effect of distractor salience on saccade trajectory deviation. In this paradigm, the target appears within an array of vertical lines. One line is oriented slightly differently from the others, and this line serves as the distractor. By varying the extent to which the orientation of the distractor differs from that of the surrounding vertical lines, it can be investigated how this contrast, or ‘salience’, affects the trajectory of the saccade. As noted above, this paradigm reveals that the more salient distractors (i.e. those whose orientation contrasts more starkly with that of the surrounding lines) elicit greater deviation towards them, but only for short latency saccades (van Zoest et al., 2012). This finding has been explained as the result of more salient distractors eliciting more oculomotor activity during planning of the saccade (White et al., 2012). However, this activity is transient, which results in salience effects on saccade trajectories disappearing at longer latencies (Donk & van Zoest, 2008). Similar findings have been made for other sources of salience, such as the luminance of the distractor (Jonikaitis & Belopolsky, 2014).

We consider it important to investigate effect sizes for the effect of a basic feature of the distractor because it may be the case that the measures most sensitive to the basic effect of saccade latency may not be the same measures that are most sensitive to changes in the distractor. In view of the fact that many studies vary the type of distractor (e.g. Jonikaitis & Belopolsky, 2014; Laidlaw et al., 2015; McSorley & van Reekum, 2013; McSorley & Morriss, 2015; van Zoest et al., 2012; Weaver et al., 2011) we wish to be able to recommend optimal measures specifically for this type of study.

Study 1: McSorley et al. (2006)

In Study 1, in order to investigate measures of saccade trajectory deviation in one of the simplest situations possible, we analyzed data from the basic target-distractor paradigm described above, in which the target and the distractor are two shapes that appear simultaneously at random locations and are not varied in any way. We extracted the eight measures described in the introduction above, and used PCA to identify clusters of related measures. We also calculated the effect sizes for the basic effect of saccade latency on trajectory deviation, to identify measures that have the most power to detect this effect.

Methods

Data

Data were taken from a previously-published eye movement study (McSorley et al., 2006) with the authors’ permission. Readers are referred to the original article for a detailed description of the methods. Briefly, seven participants completed 420 trials each of a saccade task in which the goal was to make an eye movement to a target shape that could appear randomly in one of four possible locations, while ignoring a simultaneously-appearing distractor shape, which appeared nearby. Eye movements were recorded using an Eyelink with a sampling rate of 250 Hz. **Figure 2** gives a schematic of the stimulus display.

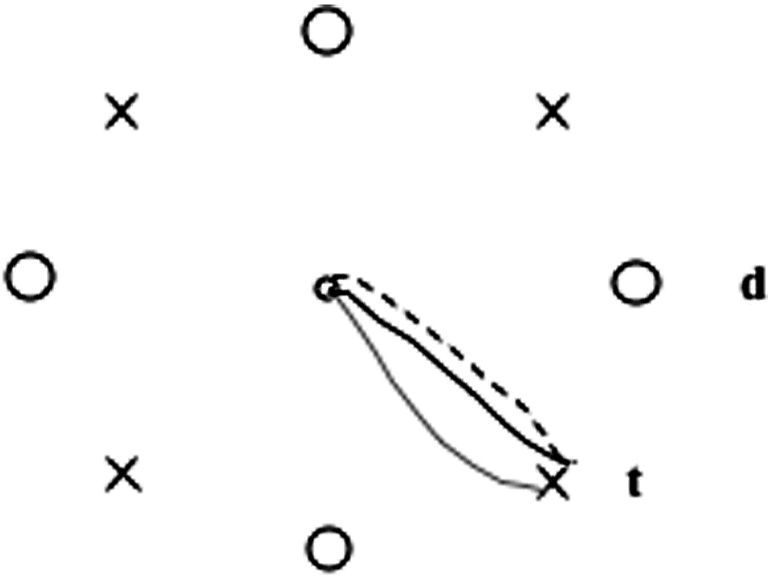

Figure 2: Example stimulus display for the target-distractor task. Figure shows all possible target positions and distractor positions, though only one target (**t**) and one distractor (**d**) were displayed on any given trial. Bold line shows an example saccade trajectory from a trial without a distractor. Dashed line shows an example of a saccade deviating towards the distractor. Grey line shows an example of deviation away from the distractor. Reproduced from McSorley et al. (2006).

Data processing

All gaze samples falling outside the dimensions of the stimulus monitor were discarded. Gaze samples that did fall within the dimensions of the monitor were smoothed, in order to average out small-scale sampling noise. This was achieved by replacing the x and y coordinates of each sample with the mean of coordinates from all samples within 2.5 ms of the current sample (i.e. smoothing with a ‘rectangular sliding window’).

For each trial, gaze samples were re-centered on the fixation spot to correct for drift in the eye tracking system. This was accomplished by assuming that the participant was fixating the fixation spot as instructed during the 60 ms prior to the onset of the task display. The median gaze position during this time window was then assumed to be the center of the screen, and all samples for the trial were re-centered on this point by rigid body translation.

To extract the first saccade from the processed samples, we used a ‘velocity peak method’ (e.g. Smeets & Hooge, 2003). This method avoids erroneously categorizing small fluctuations in gaze velocity as saccades, as may occur with a fixed saccade velocity criterion (Nyström & Holmqvist, 2010). The first velocity peak was identified as the first set of contiguous samples with a velocity greater than 100 deg/s. The start and end points of the saccade were identified by searching from this peak backwards and forwards in time respectively until finding a sample with a velocity below 35 deg/s and an acceleration below 0 deg/s^2.

The eight measures of saccade trajectory deviation described above were calculated for each extracted saccade. Each measure was calculated in a clockwise direction. An implementation of all saccade trajectory calculations for the MatLab programming environment is available from the corresponding author’s website[1]. A baseline measure of deviation was calculated as the mean deviation in trials with no distractor, separately for each target position that appeared in the experiment. This was subtracted from the deviations in distractor trials to correct for any tendency to make slightly leftward or rightward saccades even in the absence of a distractor (Walker & McSorley, 2008). If on a given trial the distractor was located anticlockwise of the target, the sign of the measures was reversed, so that positive values indicate deviation towards the distractor and

---

[1] sites.google.com/site/luketudge/home/resources

negative values deviation away. In addition to the eight measures of saccade trajectory, saccade latency was also calculated. Latency is defined as the duration in ms of the period between the onset of the target and the participant's initiation of a saccade.

Trials were excluded from further analysis if saccade latency was less than 80 ms (suggesting an anticipatory saccade) or greater than 600 ms (suggesting a saccade that was not an immediate reaction to the onset of the stimuli), if saccade landing point was more than 30 angular degrees either side of the target, or if the participant was not fixating the screen within 2 degrees of visual angle of the fixation point at the time the saccade was initiated.

This data analysis procedure is slightly different from the published data processing procedure applied in the original study (McSorley et al., 2006). These differences were undertaken in order to ensure compatibility with the analysis of the data from our own experiment. To check that this harmonization of data processing procedures did not alter the conclusions drawn, we repeated all analyses described below but after processing the raw data according to the procedures described in the original article rather than the procedure described above. This version of the analysis entailed no qualitative differences in any of the conclusions drawn.

To identify groups of measures that may reflect the same underlying phenomenon, a Principal Components Analysis (PCA) was conducted. For each principal component, the loadings of each measure onto that component were calculated. Groups of measures that may reflect the same underlying phenomenon will load maximally onto the same component. To prepare data for PCA, data were combined across all participants by standardizing values within each participant. For each measure, each participant's mean was subtracted from their values, then values were divided by their standard deviation. Using all standardized values together, eight principal components were extracted. Results are reported for PCA using only those components with eigenvalues greater than 1, indicating that they accounted for more variance than did the measures themselves on average (Kaiser, 1960). The component loadings were calculated using the oblimin rotation so as to allow for correlations among the components themselves.

It is possible that some relevant between-participant differences remain after the standardization procedure, and that the results of the PCA reflect these differences and not a structure of relationships among the eight measures that is common to all participants. To check for this possibility, PCA was therefore also carried out separately for each participant using only their data.

For the analysis of effect sizes, the standardized effect size (η^2_G) for the effect of saccade latency was calculated for each measure. To prepare data for analysis of effect sizes, four 'latency bins' were created for each participant. This was achieved by grouping each participant's trials into four quarters, from lowest to highest latency, and then calculating the mean latency and mean saccade trajectory deviation within that latency bin for each of the eight measures of deviation. For each measure, the participant means were then entered into a one-way analysis of variance, with latency bin as a four-level factor. Effect sizes were based on the main effect of the latency bin factor. In the original study (McSorley et al., 2006), eight latency bins were used, and not four. However, we use four in order to preserve comparability with other studies that also used four (e.g. van Zoest et al., 2012; Tudge & Schubert, 2016).

Results

Principal Components Analysis

Three principal components had eigenvalues greater than 1, and were therefore included in the final analysis. *Area curvature*, *maximum curvature*, and *quadratic curvature* all loaded maximally onto the first component. These are all measures that are neither early nor late, but measure the curved shape of the saccade trajectory, so we term this the 'mid-saccade' component. *Initial direction*, *overall initial direction*, and *initial average curvature* all loaded maximally onto the second principal component. Since these are all early measures, we term this the ‘early’ component. Finally, the two remaining measures, *saccade deviation* and *overall direction*, loaded maximally onto the third principal component. The interpretation of this third component is somewhat less clear (see Discussion, below), but since it includes the only measure of late deviation, we term this the ‘late’ component. **Table 2** gives the loadings of the eight measures onto the three components.

Component	mid				early				late			
Study	McS	McS (r)	vZ	vZ (r)	McS	McS (r)	vZ	vZ (r)	McS	McS (r)	vZ	vZ (r)
area curvature	**0.99**	**0.97**	**0.98**	**0.94**	0.01	0.03	0.02	0.08	0.00	0.00	-0.04	-0.02
quadratic curvature	**0.99**	**0.98**	**0.98**	**0.94**	0.01	-0.02	-0.01	-0.01	-0.05	-0.04	0.00	0.00
maximum curvature	**0.99**	**0.93**	**0.97**	**0.97**	-0.06	-0.04	-0.03	-0.07	0.05	0.04	0.03	0.00
initial direction	0.00	0.01	0.02	0.15	**0.98**	**0.99**	**0.97**	**0.87**	-0.12	-0.15	-0.11	-0.07
overall initial direction	-0.02	-0.01	0.02	0.11	**0.91**	**0.89**	**0.76**	**0.66**	0.21	0.20	0.37	0.47
initial average curvature	0.13	0.08	0.06	-0.02	**0.60**	**0.58**	**0.82**	**0.95**	-0.19	0.03	0.13	-0.12
overall direction	-0.05	-0.05	0.00	0.00	-0.10	-0.10	-0.09	-0.14	**1.01**	**1.00**	**1.01**	**1.00**
saccade deviation	0.23	0.18	0.10	0.04	0.35	0.30	0.49	0.49	**0.72**	**0.73**	**0.63**	**0.64**

Table 2: Loadings for the different measures on the first three components for all four data sets (excluding the down-sampled data from our replication of McSorley et al., 2006). McS: McSorley et al. (2006); vZ: van Zoest et al. (2012); (r): replication. Maximum loadings are shown in bold.

The three components were also positively correlated with each other. The early and mid components were most strongly correlated (r = .44). The late component was somewhat less strongly correlated with the early (r = .23) and mid components (r = .21). **Figure 3** shows the correlations among the individual measures themselves.

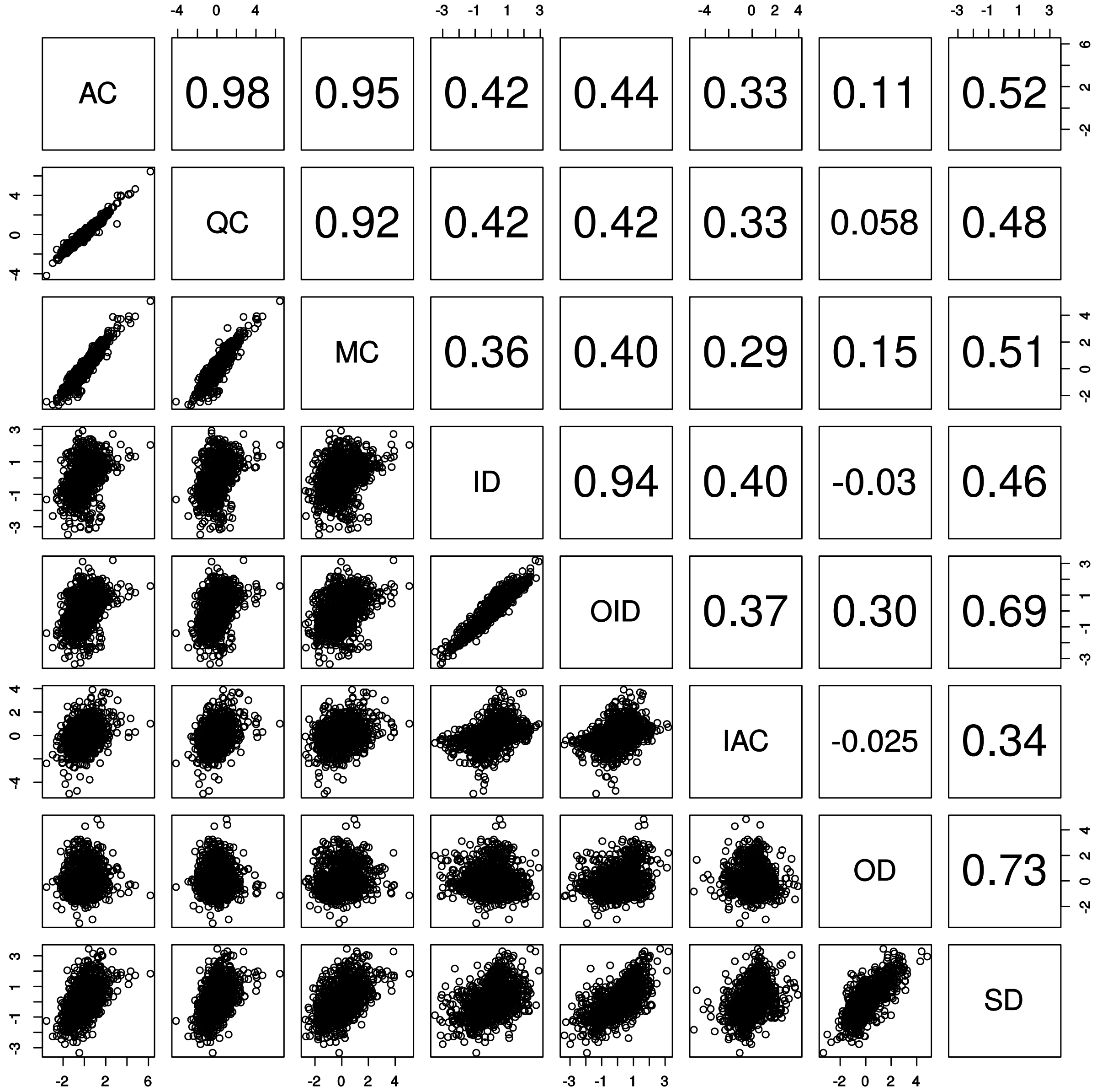

Figure 3: Scatterplot matrix showing relationships among the measures of saccade trajectory deviation, using standardized data from all participants, as described in the Methods section. The cells along the diagonal give the abbreviated names of the eight measures of saccade trajectory deviation (as given in the Introduction). Each of the cells below the diagonal shows a scatterplot of the association between the measure named in that column and the measure named in that row. Each point in each scatterplot represents one saccade. The values for each measure are standardized to *z* scores for ease of comparison, and are given in a scale at the very ends of each row. Each of the cells above the diagonal gives Pearson's correlation coefficient *r* for the correlation between the measure named in that column and the measure named in that row.

Effect sizes: Saccade latency

Effect sizes for the main effect of saccade latency were greatest for *overall direction* (.77) and *saccade deviation* (.75), the two measures that loaded maximally onto the late component. For the three measures that loaded maximally onto the mid-saccade component, effect sizes were somewhat smaller (between .30 and .35). For the remaining measures that loaded maximally onto the early component, effect sizes were very variable, ranging from .07 for *initial average curvature* to .52 for *overall initial direction*. All effect sizes are listed numerically in **Table 3**. **Figure 4** gives a visual comparison of the effect sizes. *Overall direction* and *saccade deviation* yielded the largest effect sizes, and *initial direction* and *initial average curvature* yielded the smallest.

	McSorley et al. (2006) (250 Hz)		Replication (1250 Hz)		Replication (downsampled to 250 Hz)	
	η^2_G	p	η^2_G	p	η^2_G	p
area curvature	.35	.002	.17	< .001	.17	< .001
quadratic curvature	.30	.005	.15	< .001	.15	< .001
maximum curvature	.32	.002	.13	< .001	.12	< .001
initial direction	.11	.327	.22	< .001	.14	< .001
overall initial direction	.52	< .001	.29	< .001	.21	< .001
initial average curvature	.07	.586	.19	< .001	.16	< .001
overall direction	.77	< .001	.31	< .001	.30	< .001
saccade deviation	.75	< .001	.29	< .001	.28	< .001

Table 3: Effect sizes (η^2_G) and *p*-values for the main effect of saccade latency for all eight measures for all three data sets based on the target-distractor paradigm in McSorley et al. (2006).

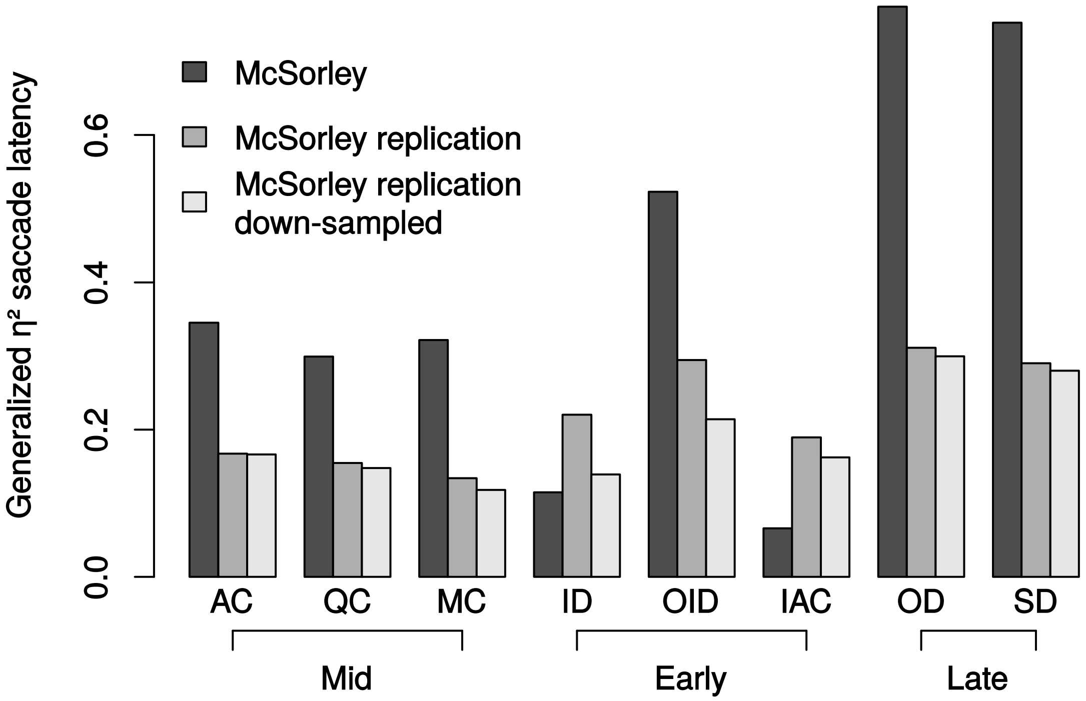

Figure 4: Effect sizes (η^2_G) for the effect of saccade latency on each of the eight measures. Measures are grouped by 'mid', 'early', and 'late' PCA component. The different colored bars shown side-by-side give effect sizes for each of the three data sets based on the target-distractor paradigm in McSorley et al. (2006).

Figure 5 gives an alternative visualization of the differences between a measure with a large effect size, *overall direction*, and a measure with a small effect size, *initial direction*. For each measure, mean saccade latency and deviation are plotted for the four latency quartiles. The established negative association of latency and deviation (McSorley et al., 2006) is clearly visible for *overall direction*, and is large relative to the variance in this measure, whereas the same trend is not clearly discernible for *initial direction*, and to the extent that the trend exists, it is slight relative to the variance in the measure.

The results of the analysis of variance also illustrate the advantage of a measure with a large effect size over a measure with a small effect size. Analysis of variance compares differences among groups, in this case latency quartiles, to differences within groups, which in this case are a reflection of the variance in the measure being used. As **Figure 5** shows, for *initial direction* the differences in deviation between latency quartiles are small relative to the variance in the measure, whereas for *overall direction* the opposite is the case. *Initial direction* should therefore have less power to detect the effect of saccade latency. The hypothesis test for the analysis of variance confirmed this conclusion. There was a significant main effect of saccade latency quartile on *overall direction*: $F(3,18) = 33.92$; $p < .001$, but not on *initial direction*: $F(3,18) = 1.23$; $p = .33$.

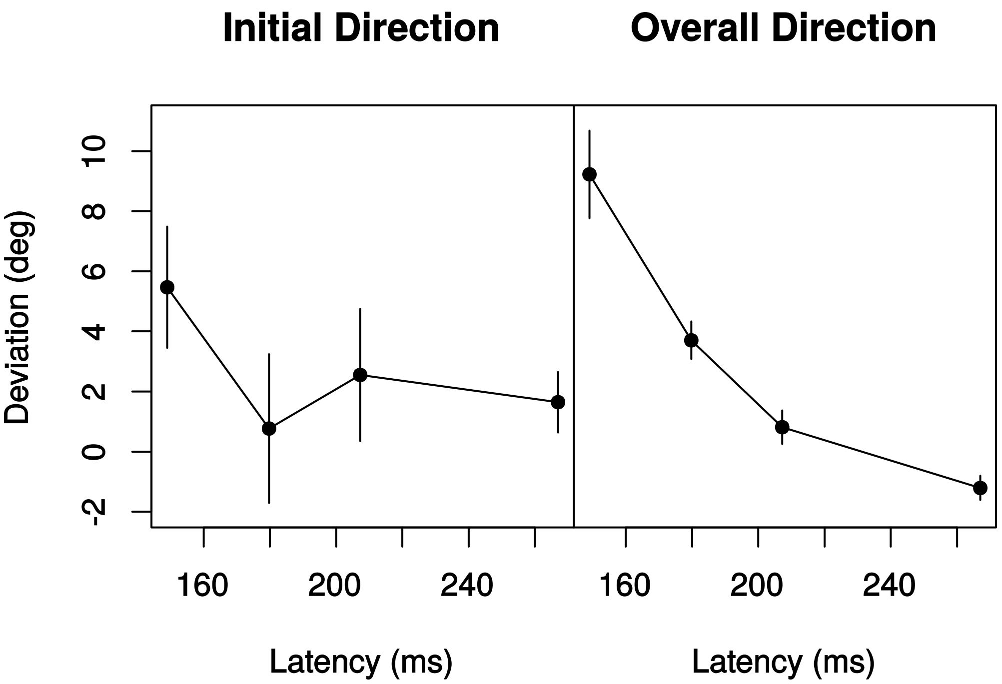

Figure 5: Mean latency and saccade trajectory deviation for four latency bins, shown for *initial direction* and *overall direction*. Error bars show ±1 Standard Error of the Mean (SEM).

Comparison of effects across saccade trajectory

As noted above, it appears to be the case that the overall direction measure affords a particularly clear reflection of the effect of saccade latency. This provides some initial support for the conclusion that gaze samples from later in the saccade are more informative. A reviewer suggested that we follow up on this conjecture by analyzing in more detail the change in effect size as the saccade progresses from start to end point.

In order to do this, we calculated separate measures of saccade trajectory deviation for different parts of the saccade. To create a set of comparable points along the trajectories of many different saccades of different amplitudes and durations, we applied a Vincentizing procedure (Vincent, 1912). Specifically, ten ‘virtual’ gaze samples were created for each saccade, evenly-spaced along the path of the saccade. The coordinates of each of these virtual gaze samples were estimated by linear interpolation between the two closest real samples in the saccade (see van Zoest, et al., 2012, for a similar use of linear interpolation to create evenly-spaced gaze samples). For each of these ten gaze samples, the angle between a straight line from saccade start to the gaze sample and a straight line from saccade start to the target was calculated, as for the *saccade deviation* measure. The first interpolated sample occurred at one tenth of the distance along the saccade, the second at one twentieth the distance, and so on; the final one occurred at saccade endpoint, and was therefore equivalent to the *overall direction* measure.

In the results of this additional analysis, the effect size for the main effect of saccade latency on the angular deviation of the saccade was greatest at the end of the saccade (i.e. for *overall direction*, 0.77), and lowest at the beginning of the saccade (0.68), with a monotonic increase in between. **Figure 6** illustrates this increase in effect sizes from saccade start to saccade end.

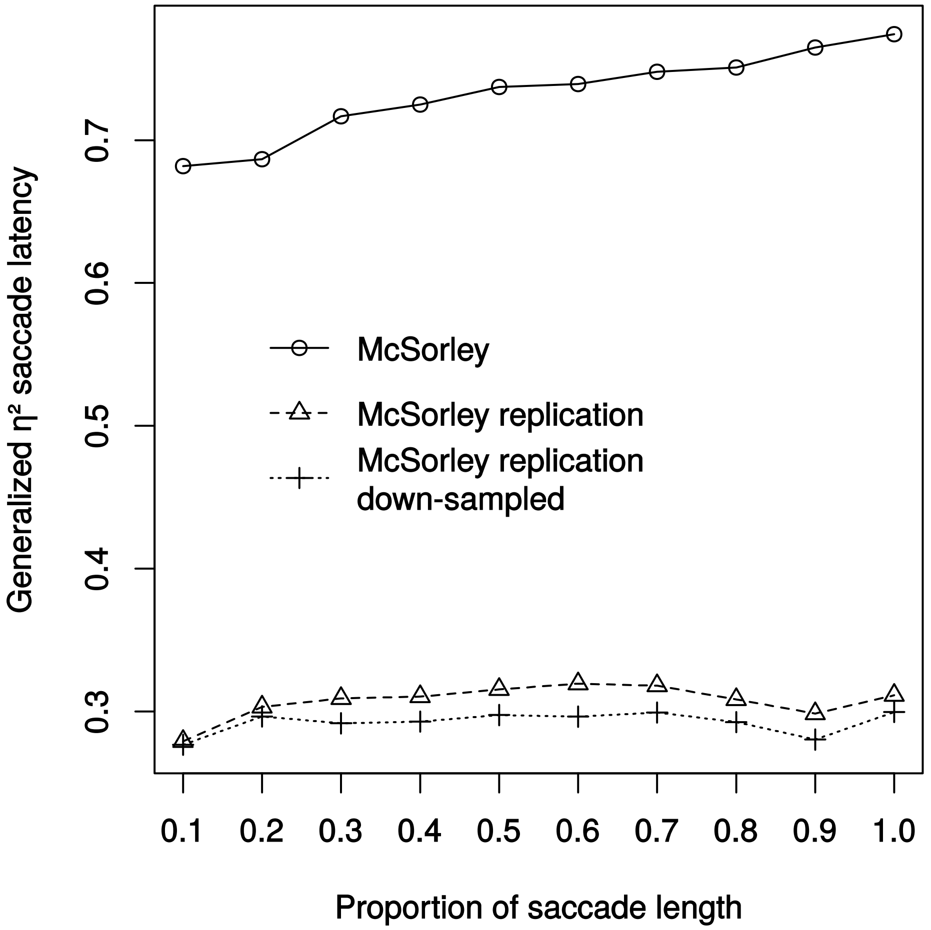

Figure 6: Effect size (η^2_{G}) for the effect of saccade latency on angular deviation of saccade from a straight line to the target, measured at ten different points along the saccade. The *y* axis gives effect sizes. The *x* axis gives the point at which deviation was measured, as a proportion of total saccade length. For example, 0.5 is halfway through the saccade, and 1 is at saccade endpoint (equivalent to *overall direction*). Separate lines show data from each of the three data sets based on the target-distractor paradigm in McSorley et al. (2006).

Discussion

Based on the results from Study 1, there appear to be three clusters of measures that reflect three underlying components of a saccade: its early deviation, its curved trajectory, and its later deviation. These components are themselves moderately positively correlated with each other. The later measures, *saccade deviation* and *overall direction*, appear to have the greatest power to measure the effect of saccade latency. This conclusion is further supported by the finding that, within the saccade, effect sizes increase for measures based on later gaze samples.

With the exception of *overall initial direction*, the early measures seem particularly poorly suited to measuring the effect of saccade latency, as they have low effect sizes compared to the other measures. However, this may in part be due to the fact that McSorley et al. (2006) used an eye tracker with a fairly low sampling rate of 250 Hz. Generally, the effect of a higher sampling rate is to help average out random variance in the eye tracker’s estimates of gaze position, particularly if spatial smoothing of the gaze samples is applied. With a low sampling rate, there may be a large amount of variance in the gaze samples, which probably leads to more variance in the measures themselves, which in turn means smaller effect sizes, all else being equal.

To see why spatial noise might disproportionately affect the early measures of saccade trajectory deviation, it helps to consider **Figure 1**. The gaze samples on which the early measures are based are located close to the start of the saccade, near the corner at which the angle of deviation is calculated. This means that these samples have high leverage on that angle. Small movements of these samples can lead to big changes in the angle. Movements of the same magnitude for later samples lead to much smaller changes in the angle of deviation.

Study 2: Replication of McSorley et al. (2006)

In order to check the generalizability of the results from Study 1 to a new group of participants and to different eye tracking system, we conducted our own experiment with the same paradigm, and repeated all the analyses described above. In addition, in order to check whether the sampling rate of the eye tracker is relevant for effect sizes, we conducted the experiment using an eye tracker with a high sampling rate (1250 Hz), and conducted the analysis once using all samples, and a second time after down-sampling the data to 250 Hz.

Methods

19 participants (12 female, 7 male, mean age 28.5, age range 18 to 49) completed the same target-distractor task as described in McSorley et al. (2006). All relevant parameters of the experiment, such as the size and shape of stimuli and the timing of display onsets were kept the same as reported in the original study. The only change was to double the number of trials that each participant completed, from 420 to 840.

The task display was programmed using MatLab with the Psychophysics Toolbox, and shown on a Samsung SyncMaster 2233 monitor with a refresh rate of 60Hz using the default manufacturer settings for brightness and contrast. Eye movements were recorded from the left eye only, using an SMI iView X Hi-Speed system with a sampling rate of 1250 Hz. The experiment was constructed in a blinded room with a diffuse, dim light source. The participant was seated at a desk facing the display monitor at a distance of approximately 70 cm, with chin resting on the eye tracking system's built-in chin rest. The eye tracking system was controlled from a separate PC at the experimenter's desk nearby.

The data processing and analysis procedures were the same as described above for Study 1. The only exception was that the analysis of effect sizes was carried out twice, once as normal, then a second time after down-sampling gaze samples to 250 Hz. Down-sampling was achieved by using only every fifth sample. We also applied the same additional Vincentized analysis described above for Study 1.

Results

Principal Components Analysis

The structure of component loadings for the first three principal components in the aggregate analysis was the same as for Study 1 (i.e. measures that loaded maximally onto a particular component in Study 1 also did so in Study 2). **Table 2** gives the loadings of the eight measures onto the three components. Again, the three components were positively correlated with each other. The pattern of correlations was similar to those in Study 1. The early and mid components were most strongly correlated (r = .69), and the late component was less strongly correlated with the early (r = .30) and mid components (r = .26).

Effect sizes: Saccade latency

Effect sizes were generally lower than in Study 1. The 1250 Hz data showed a similar overall pattern to Study 1, with *overall direction* and *saccade deviation* yielding relatively high effect sizes. The exception was that effect sizes for the early measures were no longer very low compared to the other measures (see **Figure 4**, above). For the data down-sampled to 250 Hz, down-sampling selectively reduced effect sizes for the early measures, while having almost no impact on the other measures. **Figure 4** shows the changes in effect size as a result of down-sampling. All effect sizes are also given in **Table 3**.

Comparison of effects across saccade trajectory

Effect size for the main effect of saccade latency was greater at the end of the saccade (i.e. for *overall direction*, $\eta^2_G = 0.31$), than at the beginning of the saccade ($\eta^2_G = 0.28$). However, this time the increase in between was not completely monotonic, with the greatest effect size being achieved for the gaze samples located at 60% of the total length of the saccade, very slightly higher that at the end of the saccade ($\eta^2_G = 0.32$). **Figure 6** illustrates the change in effect sizes from saccade start to saccade end.

Discussion

With a new experiment we confirmed the generalizability of the relationships among the measures as revealed in Study 1, namely the three groupings of early, mid-, and late measures of saccade deviation.

In the analysis of effect sizes, there were two discrepancies between the two studies. First, effect sizes in Study 2 were considerably smaller than in Study 1. However, we do not think that this difference is consequential for our conclusions. Effect sizes are a reflection of the the variance in the data as well as the experimental effects. There may be fairly trivial differences between the two studies that led to greater variance in Study 2, for example the use of slightly different participant groups who may have different levels of experience in experiment participation, or the use of a different eyetracking system (the Eyelink in McSorley et al., 2006, and the iView X in the present study). However, what is striking despite the difference in effect size values is that the relative profile of effect sizes over the different measures is the same in the two studies. We are concerned with the relative merits of the measures, rather than the specific values of the effect sizes.

Second, although most aspects of the relative profile of effect sizes generalize well from the first data set to the second, the early measures performed relatively better in Study 2. We were able to attribute at least some of this change to the fact that we used an eye tracking system with a sampling rate of 1250 Hz, whereas McSorley et al. (2006) only used 250 Hz. However, we should be somewhat cautious in attributing this discrepancy in its entirety to the sampling rate of the eye tracking system. Although down-sampling our data to the same sampling rate as in McSorley et al. (2006) reduced effect sizes selectively for early measures, as this explanation predicts, the early measures still showed relatively high effect sizes for our data. We may nonetheless conclude that an eye tracking system with a high sampling rate is better for obtaining reliable measures of early saccade trajectory deviation.

In both studies, the late measures *saccade deviation* and *overall direction* yielded the highest effect sizes, as did measures of saccade deviation based on gaze samples located later in the saccade. *Saccade deviation* and late measures may therefore be best suited to detecting the effects of experimental manipulations. However, we measured effect sizes based only on the effect of saccade latency. Future researchers may be interested in selecting a measure that is optimal for detecting other effects.

Study 3: van Zoest et al. (2012)

In Study 3, we aimed to test the power of the different measures to detect the effect of varying a feature of the distractor. For that purpose, we investigated a target-distractor paradigm in which the physical salience of the distractor varies. In this case, differences in salience are achieved by displaying the distractor in an array of vertical lines. The distractor is also a line, but is oriented either slightly differently (low salience) or very differently (high salience) from the other lines. It has been shown that if a distractor is more salient, i.e. contrasts more starkly with its surroundings, then it will produce greater saccade trajectory deviations (van Zoest et al., 2012). Van Zoest et al. (2012) also reported an effect size, but only for the *saccade deviation* measure. As in Studies 1 and 2, we calculated effect sizes for all eight measures to assess how well each of them reflects the effect of distractor salience on saccade trajectory deviation.

We also calculated the same PCA analysis as for the other data sets, as well as repeating the analysis of effect sizes for saccade latency, in order to test the generalizability of the earlier conclusions to a different experimental paradigm.

Methods

Data

Data were taken from a previously-published eye movement study (van Zoest et al., 2012), with the authors' permission. Readers are referred to the original article for a detailed description of the methods. Briefly, ten participants completed 624 trials each of a saccade task in which the goal was to make an eye movement to a target shape (a small circle) that could appear randomly in one of two possible locations, vertically either above or below the fixation point at the center of the screen. Simultaneously with the onset of the target, an array of vertical lines appeared on the screen. One of these lines served as the distractor, and could be of two types. Either the distractor was oriented slightly differently from the other lines, in which case it was a low-salience distractor, or it was oriented very differently from the other lines, in which case it was a high-salience distractor. Eye movements were recorded using an Eyelink II with a sampling rate of 500 Hz. The aim of the original study was to test whether distractors of high salience elicit greater saccade trajectory deviation than distractors of low salience. **Figure 7** shows an example stimulus display.

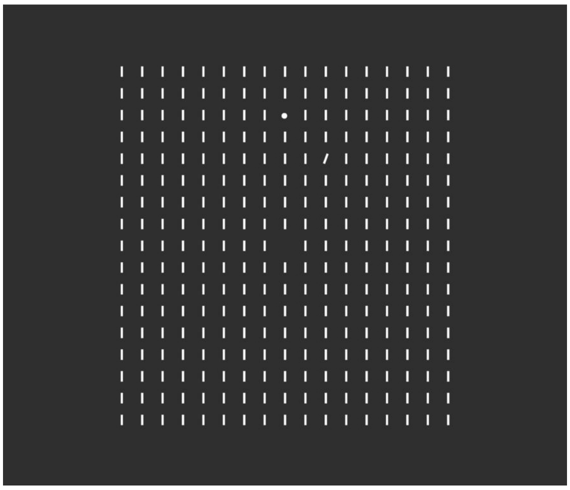

Figure 7: Example stimulus display for the distractor salience task, with a low-salience distractor. Reproduced from van Zoest et al. (2012).

Data processing

Data were processed in the same manner as described above for the basic target-distractor paradigm, with the exception of the analysis of variance procedure. As well as latency quartile, distractor salience was added as an additional factor with two levels, resulting in a 4 x 2 design (as in van Zoest et al., 2012). Effect sizes were then calculated for the main effect of distractor salience.

Results

Principal Components Analysis

The structure of component loadings for the first three principal components was the same as for the other data sets (i.e. measures that loaded maximally onto a particular component in the data from the McSorley et al., 2006, data sets also did so in Study 3). These results support the same groupings of measures into three underlying components as in the first two studies. **Table 2** gives the loadings of each measure onto each component. The three components were again positively correlated with each other, with the early and mid components most strongly correlated (*r* = .44), and the late component less strongly correlated with the early (*r* = .34) and mid components (*r* = .19).

Effect sizes: Saccade latency

The pattern of effect sizes for the effect of saccade latency was slightly different from that observed for the data sets based on McSorley et al. (2006). A mid-saccade measure, *quadratic curvature* showed the highest effect size (.22). The late measures *overall direction* (.21) and *saccade deviation* again showed high effect sizes, though the effect sizes for the other mid-saccade measures were almost as high (between .18 and .19). Again, the early measures with the exception of *overall initial direction* (.18) showed the smallest effect sizes (between .07 and .10). **Figure 8** displays the results for the effect of saccade latency, and the values are given in **Table 4**.

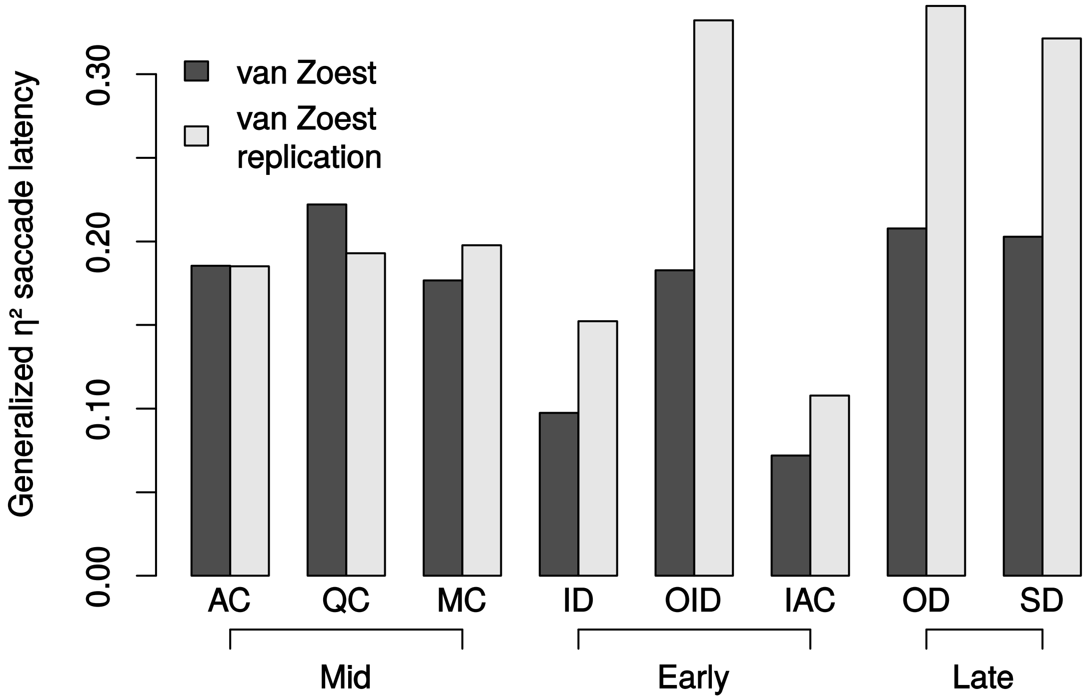

Figure 8: Effect sizes (η^2_G) for the effect of saccade latency on each of the eight measures. The different colored bars shown side-by-side give effect sizes for each of the two data sets based on the target-distractor paradigm in van Zoest et al. (2012).

	van Zoest et al. (2012) (saccade latency)		Replication (saccade latency)	
	η^2_G	*p*	η^2_G	*p*
area curvature	.19	.004	.19	< .001
quadratic curvature	.22	< .001	.19	< .001
maximum curvature	.18	.007	.20	< .001
initial direction	.10	.073	.15	< .001
overall initial direction	.18	< .001	.33	< .001
initial average curvature	.07	.155	.11	.003
overall direction	.21	< .001	.34	< .001
saccade deviation	.20	< .001	.32	< .001

Table 4: Effect sizes (η^2_G) and *p*-values for the main effect of saccade latency for all eight measures for both data sets based on the target-distractor paradigm in van Zoest et al. (2012).

Effect sizes: Distractor salience

The analysis replicated the main finding of the original study (van Zoest et al., 2012), namely that saccade trajectory deviation towards the distractor is greater when that distractor is of high salience, compared to when it is of low salience. The effect size for this main effect (i.e. for the difference in deviation between low and high salience distractors) was greatest for *overall direction* (.05) and *saccade deviation* (.05), slightly lower for *overall initial direction* (.04), and lowest for all other measures (between .00 and .02; see **Figure 9**). All values are given in **Table 5**.

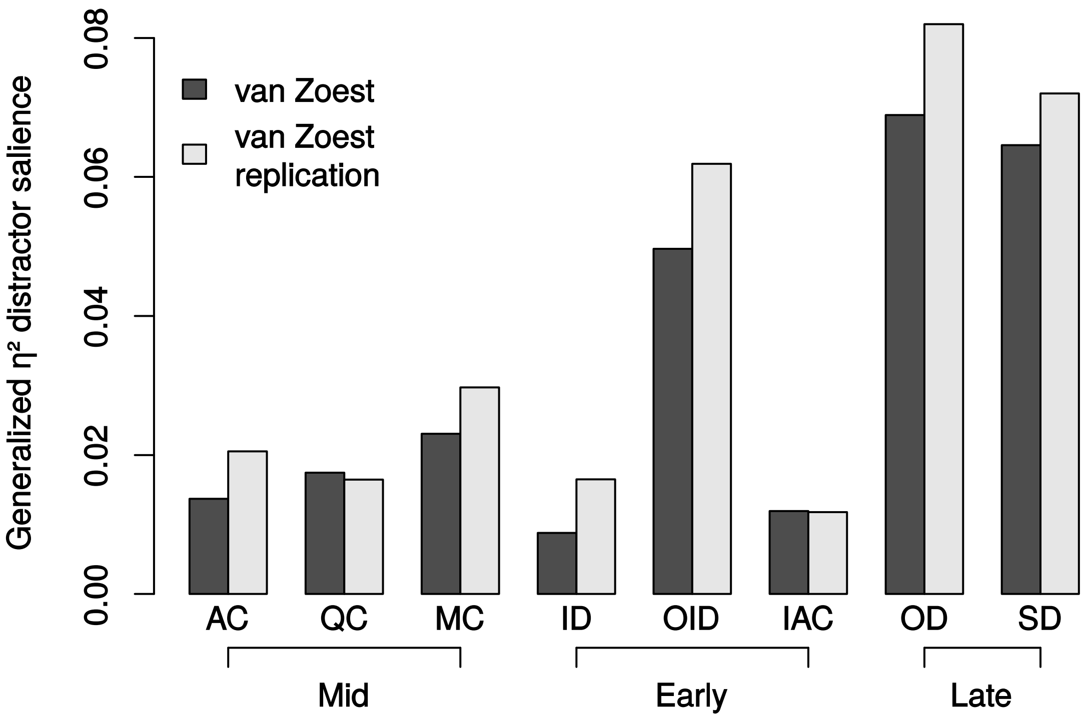

Figure 9: Effect sizes (η^2_G) for the effect of distractor salience on each of the eight measures. The different colored bars shown side-by-side give effect sizes for each of the two data sets based on the target-distractor paradigm in van Zoest et al. (2012).

	van Zoest et al. (2012) (distractor salience)		Replication (distractor salience)	
	η^2_G	p	η^2_G	p
area curvature	.014	.023	.021	< .001
quadratic curvature	.017	.015	.016	< .001
maximum curvature	.023	.016	.030	< .001
initial direction	.009	.067	.016	.003
overall initial direction	.050	< .001	.062	< .001
initial average curvature	.012	.122	.012	.016
overall direction	.069	.004	.082	< .001
saccade deviation	.065	< .001	.072	< .001

Table 5: Effect sizes (η^2_G) and p-values for the main effect of distractor salience for all eight measures for both data sets based on the target-distractor paradigm in van Zoest et al. (2012).

To use the same example as in Study 1, we may use hypothesis tests to illustrate the difference in power between *overall direction*, which yielded a large effect size, and *initial direction*, which yielded a small effect size. In this case, we are interested in power to detect the effect of distractor salience, so the relevant hypothesis test is for the difference in deviation between high and low salience distractors. With *overall direction* as a dependent measure, this difference was significant: $F(1,9) = 15.09$; $p < .01$, whereas the same effect for *initial direction* was not, or only marginally so: $F(1,9) = 4.35$; $p = .07$.

Comparison of effects across saccade trajectory

Effect size for the main effect of saccade latency was greatest in the middle of the saccade, for the gaze samples located at 50% of the total length of the saccade (0.23). Effect sizes were lower both at the beginning of the saccade (0.21) and at its end (0.21). **Figure 10** illustrates the change in effect sizes for saccade latency from saccade start to saccade end.

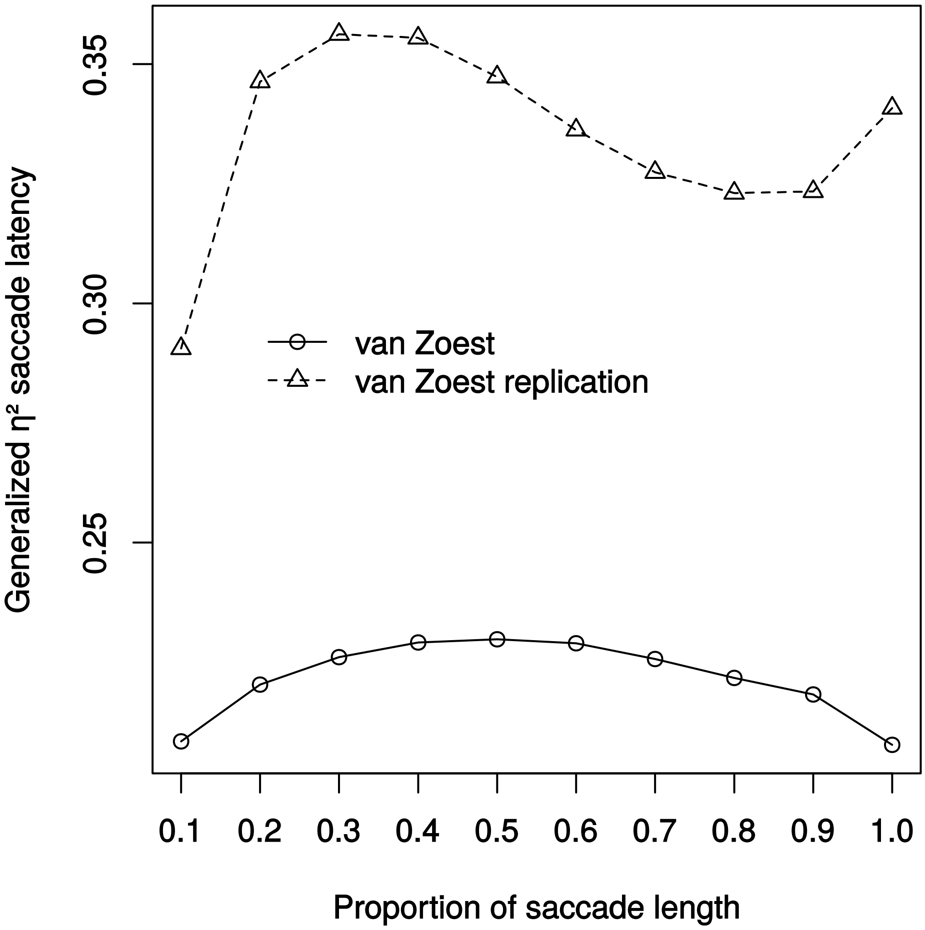

Figure 10: Effect size (η^2_G) for the effect of saccade latency on angular deviation of saccade from a straight line to the target, measured at ten different points along the saccade. Separate lines show data from each of the two data sets based on the target-distractor paradigm in van Zoest et al. (2012).

Effect size for the main effect of distractor salience was greater at the end of the saccade (i.e. for *overall direction,* 0.069), than at the beginning of the saccade (0.062). The increase in between was not completely monotonic, with an initial decrease in effect sizes fro the first few gaze samples, the lowest occurring for the gaze samples located at 50% of the total length of the saccade (0.057).

Figure 11 illustrates the change in effect sizes for distractor salience from saccade start to saccade end.

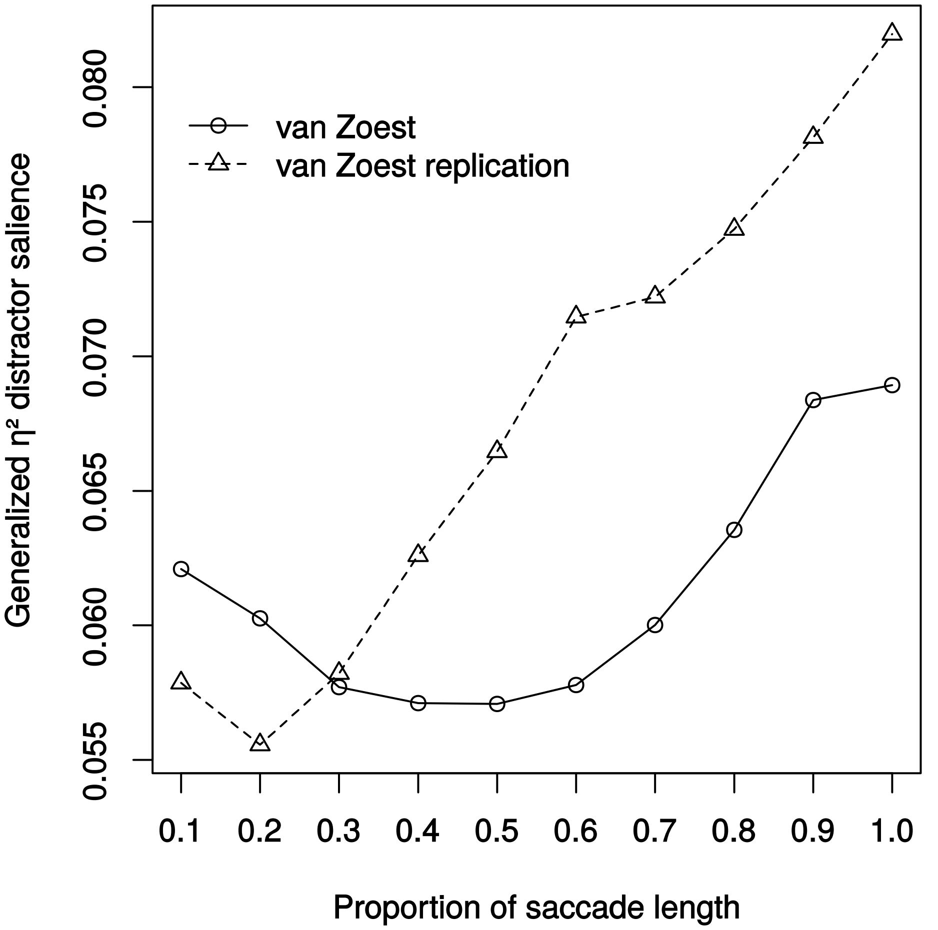

Figure 11: Effect size (η^2_G) for the effect of distractor salience on angular deviation of saccade from a straight line to the target, measured at ten different points along the saccade. Separate lines show data from each of the two data sets based on the target-distractor paradigm in van Zoest et al. (2012).

Discussion

The results replicate the main finding of van Zoest et al. (2012), that greater distractor salience produces greater saccade trajectory deviation. The effect sizes for the effect of distractor salience are considerably smaller than for the effect of saccade latency. This is a reflection of the fact that saccade latency has a much more pronounced effect on saccade trajectories than does distractor salience, and may also be due to the fact that the effect of distractor salience is only present at shorter latencies, so may be somewhat obscured in the data as a whole (van Zoest et al., 2012).

The original study found the effect of distractor salience to be significant using the *saccade deviation* measure. In our analysis, *saccade deviation* was one of the most powerful measures for detecting this difference, along with *overall direction*, which suggests that the authors used an optimal, or close to optimal, measure for detecting the effect of interest. For the effect of distractor salience, the superiority of *overall direction*, *saccade deviation*, and *overall initial direction* was even more pronounced than for the effect of saccade latency in the data sets based on McSorley et al. (2006). This suggests that the usefulness of these measures may not be limited to measuring the effects of saccade latency, but may be more general.

Study 4: Replication of van Zoest et al. (2012)

Again, in order to check the generalizability of the conclusions from Study to 3 new participants and a different eyetracking system, we conducted our own experiment using the same paradigm.

Methods

22 participants (17 female, 5 male, mean age 26.5, age range 19 to 36) completed 900 trials each of the same task as described in van Zoest et al. (2012). The technical set-up of the experiment was as described above for Study 2. All relevant parameters of the experiment, such as the size and shape of stimuli and the timing of display onsets were kept the same as reported in the original study. The only change was to increase the number of trials that each participant completed, from 624 to 900. The data processing and analysis procedures were the same as described above for Study 3.

Results

Principal Components Analysis

The structure of component loadings for the first three principal components was the same as in the other data sets (i.e. measures that loaded maximally onto a particular component in the first three studies also did so in Study 4). These results support the same groupings of measures into three underlying components as in the first three studies. **Table 2** gives the loadings of each measure onto each component. The three components were again positively correlated with each other, with the early and mid components most strongly correlated ($r = .55$), and the late component less strongly correlated with the early ($r = .24$) and mid components ($r = .20$).

Effect sizes: Saccade latency

The pattern of effect sizes for saccade latency was more closely similar to that observed for the data sets based on McSorley et al. (2006) than it was in the original data from van Zoest et al. (2012) analyzed in Study 3. In particular, *overall direction*, *saccade deviation*, and *overall initial direction* again showed higher effect sizes (between .32 and .34) than the other measures (between .11 and .20). **Figure 8** illustrates these differences, and all effect size values are given in **Table 4**.

Effect sizes: Distractor salience

The data revealed a very similar pattern to Study 3. *Overall direction* and *saccade deviation* yielded the largest effect sizes (.08 and .07, respectively), followed by *overall initial direction* (.06), then the other measures (between .01 and .03; see **Figure 9**). All values are given in **Table 5**.

Comparison of effects across saccade trajectory

Effect size for the main effect of saccade latency showed a non-linear trend across the length of the saccade. It was smallest at the beginning of the saccade (0.29) but increased rapidly thereafter, reaching its highest point at 30% of the saccade trajectory (0.36). It decreased afterwards, until 80% of the saccade trajectory (0.32), and then finally increased again somewhat until the end of the saccade, i.e. for *overall direction* (0.34). **Figure 10** illustrates the change in effect sizes for saccade latency from saccade start to saccade end.

Effect size for the main effect of distractor salience was greater at the end of the saccade (i.e.

for *overall direction*, 0.082), than at the beginning of the saccade (0.058). The increase in between was almost monotonic, excepting a slight initial decrease in effect sizes for the second gaze sample, located at 20% of the total length of the saccade (0.056). **Figure 11** illustrates the change in effect sizes for distractor salience from saccade start to saccade end.

General discussion

In the discussion of our results, we consider first the findings from PCA in all four studies. The aim of this analysis was to identify commonalities among the different measures and to organize them into related groups. This makes it clearer where findings from different experiments using different measures may be comparable and where not. We then consider the analysis of effect sizes for the decrease in saccade trajectory deviation with increasing saccade latency (based on all data sets), and for the increase in deviation with increasing distractor salience (based on the data sets for the van Zoest et al. 2012 paradigm). The aim of this analysis was to determine which measures have the greatest power to detect these effects. Since the pattern of effect sizes was similar for saccade latency and for distractor salience, many of the conclusions we offer are general to both effects.

It is important to note here that the results of the two analyses, PCA and effect sizes, are in principle independent of one another. Although the two approaches may appear similar, in the sense that they both aim to account for variance in the measures of saccade trajectory, the questions that the two methods address are quite different. The variance that PCA aims to account for is the covariance among the measures, and therefore in a sense their similarities with one another, and this is done without reference to saccade latency or distractor salience. The variance that the analysis of effect sizes aims to account for is the variance within each measure that is attributable to saccade latency and to distractor salience. A measure may in principle be only loosely related to the other variables yet highly sensitive to the effects of experimental manipulations, and vice versa.

We used correlation and PCA to explore the structure of relationships among the eight measures of saccade trajectory deviation. This analysis revealed a component structure that was consistent for four different data sets. Given the pattern of loadings, the first three components seem to reflect three separate aspects of saccade trajectory deviation. One aspect is the state of deviation at the very beginning of the saccade (early component), another is the curvature of the whole trajectory (mid-saccade component), and a third is the state of deviation at the end of the saccade (late component).

However, the status of *saccade deviation* is somewhat problematic for the interpretation of the late component. *Saccade deviation* is calculated as an average over all gaze samples within the saccade. As such, it is not a late measure. That it is nonetheless grouped on a common underlying component with *overall direction*, the only late measure, may be due simply to the distribution of gaze samples over the trajectory of the saccade. It is known that saccades tend to slow towards their end (Van Opstal & Van Ginsbergen, 1987). Because the eye tracking system records gaze position regularly over time but not necessarily over space, a slowing of the saccade towards its end will result in more samples being collected towards the end, so these will contribute more to a measure that averages over all samples, such as *saccade deviation*.

Another possibility is that the correlation between *overall direction* and *saccade deviation* reflects their common status as target-based measures. Since they are both measured relative to the position of the target, variation in how close the saccade lands to the target will affect both measures. This conjecture is somewhat strengthened by the fact that *saccade deviation* also correlates more highly with *overall initial direction*, the only other target-based measure, than it does with the endpoint-based measures.

Saccade deviation generally correlates highly with all the other measures (see **Figure 3**). It also loads to some extent onto the early and mid-saccade components, whereas other measures load predominantly onto only one component. These properties recommend *saccade deviation* as a good general measure for new investigations without any strong hypotheses about specific components of the saccade. Use of a measure that correlates with all others also has the advantage of preserving the comparability of new results with many different existing findings.

We turn now to the analysis of effect sizes. Little systematic work has been done to compare the power of different measures of saccade trajectory deviation. One previous study compared the power of measures informally, by observing whether statistically significant effects were obtained for each measure (Van der Stigchel & Theeuwes, 2006). However, this analysis only included four measures, and did not report effect sizes, only statistical significance at certain α-thresholds (.05 and .01). Another study performed a similar comparison of *overall direction* and *maximum curvature* (McSorley, Cruickshank, & Inman, 2009).

Our results suggest that *saccade deviation* and *overall direction* are the most appropriate measures, as they showed the largest effect sizes, both for the effect of saccade latency in Studies 1 and 2, and for the effect of distractor salience in Studies 3 and 4. The fact that *overall direction*, a measure based on only a single sample, showed clear effects relative to its variance also speaks against the assertion that full-sample measures are preferable because they average out noise in the eye tracking system’s measurements (Ludwig & Gilchrist, 2002). Indeed, the full-sample measures did not perform consistently well. Although *saccade deviation* showed relatively large effects for both saccade latency and distractor salience, as noted above, the other two full-sample measures, *area curvature* and *quadratic curvature*, showed intermediate-sized effects for saccade latency and relatively very small effects for distractor salience.

We found additional evidence to support the idea that measures made later in the saccade reflect more reliably the effects of experimental manipulations. In our analysis of angular deviations at different points along the length of the saccade, we found that later points tended to show larger effect sizes. However, we are cautious in recommending the use of *overall direction* for new studies in general. Although it showed relatively large effect sizes for the two variables of interest we investigated (saccade latency in Studies 1 and 2 and distractor salience in Studies 3 and 4), two previous studies found it to be less sensitive to the experimental manipulation than some other measures. McSorley et al. (2009) manipulated the distance of the distractor from the target. They found significant effects on *overall direction* only when the distractor was fairly close to the target, whereas this modulation was no longer observable among the greater target-distractor distances. *Maximum curvature*, on the other hand, could detect differences among a wider range of target-distractor distances. This modulation of *overall direction* specifically by distractors located close to the target is well-known, as the ‘global effect’ (Coren & Hoenig, 1972; Walker, Deubel, Schneider, & Findlay, 1997; Van der Stigchel & Nijboer, 2011). We therefore recommend overall direction as an optimal measure only for studies in which the target and distractor are located close to one another, at 45 angular degrees of separation or less.

Van der Stigchel and Theeuwes (2006) measured saccade trajectory deviation relative to a location where either nothing appeared, a distractor appeared, or the participant expected a distractor to appear, though it did not. In a comparison of the effect of this manipulation on four measures of saccade trajectory deviation, the authors found that *overall direction* was the only one that did not yield a significant hypothesis test.

Some important features of the experimental design in Van der Stigchel and Theeuwes (2006) may help explain this discrepancy. The position of the distractor, if it appeared, was completely predictable, and participants were also informed between 800 and 1300 ms in advance where the target would appear. Saccade trajectory deviation towards a distractor is known to be attenuated by foreknowledge of the target and distractor (Moher, Abrams, Egeth, Yantis, & Stuphorn, 2011; Walker et al., 2006) and by task preparation in general (Tudge & Schubert, 2016). In such cases, the attenuation can be such that an overcompensation occurs and the saccade deviates away from the distractor (Walker & McSorley, 2008). Informally, we have observed that *overall direction* does not tend to show significant deviation away from a distractor, only towards it. This lack of deviation away is visible in Figure 2a of Van der Stigchel and Theeuwes (2006), and in our own **Figure 5**, above. We therefore tentatively suggest that *overall direction* may not be a suitable measure for paradigms that involve task preparedness or top-down control, which are likely to produce deviation away from the distractor (Van der Stigchel, 2010).

To speculate a little further, there may even be a reasonable physiological explanation for this particular feature of *overall direction*. It has been hypothesized that the cerebellum monitors saccade trajectories while they are underway, and corrects them back towards the target (Quaia et al., 1999). *Overall direction* represents a moment at which such an ongoing correction has already been carried out to its maximum extent, at the end point of the saccade. It may therefore be the case that deviation away from the distractor has been ‘corrected away’ by the time *overall direction* is measured. That the same does not happen to deviation towards the distractor may simply reflect the fact that deviation towards is generally of a greater magnitude to begin with, so the cerebellum is not able to correct it all before the end of the saccade.

There may be instances in which we also have theoretical reasons to want to measure saccade trajectory deviation at an early stage, before much correction has taken place, for example if we are interested in the bottom-up attentional capture elicited by the distractor. In this case, we might prefer an early measure. Unfortunately, in the present study, the early measures showed relatively very small effect sizes, particularly for the effect of distractor salience in Studies 3 and 4, an effect that is likely to be of interest in investigations of bottom-up attentional capture. However, there was one clear exception to this trend. For the effect of distractor salience, *overall initial direction* showed effect sizes only slightly smaller than *saccade deviation* and *overall direction*. *Overall initial direction* may therefore be a good choice where an early measure is required. In addition, the results from Study 2 suggest that an eyetracking system with a high sampling rate is particularly beneficial when making early measures of deviation.

Overall initial direction, *saccade deviation*, and *overall direction* were the only target-based measures we investigated, and were also those that showed the largest effects, for both saccade latency and distractor salience. Our results therefore support the general recommendation that target-based measures be preferred. As well as the purely pragmatic consideration of statistical power, we argue that target-based measures are also preferable on theoretical grounds. If it is the case that saccade trajectory deviation reflects the extent to which a motor plan for a saccade to the distractor interferes with a saccade to the target (Van der Stigchel, 2010; Walker & McSorley, 2008), then to properly quantify this interference we ought to measure it relative to the eye movement to the target that would otherwise occur. Endpoint-based measures can in theory miss the phenomenon altogether, by quantifying a straight but very erroneous saccade as having zero deviation.

It is important to bear in mind the correct interpretation of the standardized effect sizes, η^2_G, that we report here. These reflect the difference in each measure of saccade trajectory deviation between levels of the explanatory variable, i.e. different saccade latencies or levels of distractor salience, relative to the variance in the measure (Olejnik & Algina, 2003). A low effect size therefore has two possible causes. On the one hand, the explanatory variable might have no effect on the measure, or an effect too small to be of any interest. On the other hand, the variance in the measure may simply be too great for the effect to be clearly discernible. Our results cannot distinguish between these two alternatives.

However, it is not the purpose of our investigation to determine whether saccade latency or distractor salience have theoretically interesting effects on different aspects of a saccade trajectory. Rather, we aim to determine which measures are likely to enable future researchers to best distinguish those experimental effects from noise.

Since we are concerned here with effect sizes as single summary measures acquired from one experiment, we do not present typical inferential statistics based on the participant as the unit of measurement. For the validation of our conclusions we instead rely on the arguably better alternative of replication (Cohen, 1994). The conclusions we present above are strengthened by the fact that they hold true both for a reanalysis of data from existing studies (McSorley et al., 2006; van Zoest et al., 2012) and for new data from our own experiments.

Finally, we should note one important limit to the scope of our conclusions regarding effect sizes and the usefulness of different measures. There is of course no guarantee that these conclusions will hold true for every new experimental manipulation that future researchers employ. We tried to broadly cover some of the most common manipulations by including saccade latency, which often features in interactions with other manipulations (e.g. Campbell et al., 2009; van Zoest et al., 2012; Tudge & Schubert, 2016) and a manipulation of the nature of the distractor, also a common type of manipulation (e.g. e.g. Jonikaitis & Belopolsky, 2014; Laidlaw et al., 2015; McSorley & van Reekum, 2013; McSorley & Morriss, 2015; van Zoest et al., 2012; Weaver et al., 2011). However, we omitted one broad type of manipulation, namely ‘top-down’ manipulations of the participant’s own allocation of attention (e.g. Van der Stigchel & Theeuwes, 2006; Tudge & Schubert, 2016). The fact that the broad pattern of our conclusions regarding effect sizes agree for the effects of both saccade latency and distractor salience is suggestive of a more general pattern applicable across all manipulations, but further investigations are required to establish whether this is really the case.

In summary, we conclude that the *saccade deviation* measure is a good default measure of saccade trajectory deviation, as it loads reasonably highly onto all of the first three principal components of the various measures, shows relatively high effect sizes for the effect of saccade latency and that of distractor salience, and there is some evidence from another study (Van der Stigchel & Theeuwes, 2006) that it can measure deviation away from a distractor more reliably than *overall direction*. We also conclude that target-based measures are generally preferable, and therefore that if a measure of early deviation is required, *overall initial direction* is recommended. We hope that this empirically-based advice will inform future researchers’ choices of dependent measure when working with a target-distractor paradigm.

Acknowledgements

We thank Wieske van Zoest for generously sharing data from her experiments. Luke Tudge is supported by the Berlin School of Mind and Brain PhD scholarship.

References

Campbell, K., Al-Aidroos, N., Pratt, J., & Hasher, L. (2009). Repelling the young and attracting the old: Examining age-related differences in saccade trajectory deviations. *Psychology and Aging*, *24*, 163-168.

Cohen, J. (1994). The Earth is round: p < 0.05. *American Psychologist, 49*, 997-1003.

Coren, S. & Hoenig, P. (1972). Effect of non-target stimuli upon length of voluntary saccades. *Perceptual and Motor Skills, 34*, 499-508.

Donk, M. & van Zoest, W. (2008). Effects of salience are short-lived. *Psychological Science, 19,* 733-739.

Doyle, M. & Walker, R., (2001). Curved saccade trajectories: voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research, 139*, 333-344.

Erkelens, C., & Sloot, O. (1995). Initial directions and landing positions of binocular saccades. *Vision Research, 35,* 3297–3303.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417-441.

Jonikaitis, D. & Belopolsky, A. (2014). Target-distractor competition in the oculomotor system is spatiotopic. *Journal of Neuroscience, 34*, 6687-6691.

Kaiser, H. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Laidlaw, K., Badiudeen, T., Zhu, M., & Kingstone, A. (2015). A fresh look at saccadic trajectories and task-irrelevant stimuli: Social relevance matters. *Vision Research, 111*, 82-90.

Ludwig, C., & Gilchrist, I. (2002). Measuring saccade curvature: A curve-fitting approach. *Behavioral Research Methods, 34*, 618-624.

McPeek, R., & Keller, E. (2001). Short-term priming, concurrent processing, and saccade curvature during a target selection task in the monkey. *Vision Research, 41*, 785-800.

McPeek, R., Han, J., & Keller, E. (2003). Competition between saccade goals in the superior colliculus produces saccade curvature. *Journal of Neurophysiology, 89*, 2577-2590.

McSorley, E., Haggard, P., & Walker, R. (2004). Distractor modulation of saccade trajectories: spatial separation and symmetry effects. *Experimental Brain Research, 155*, 320-333.

McSorley, E., Haggard, P., & Walker, R. (2006). Time course of oculomotor inhibition revealed by saccade trajectory modulation. *Journal of Neurophysiology, 96*. 1420-1424.

McSorley, E., Cruickshank, A., & Inman, L. (2009). The development of the spatial extent of oculomotor inhibition. *Brain Research, 1298*, 92-98.

McSorley, E. & van Reekum, C. (2013). The time-course of implicit affective picture processing: An eye movement study. *Emotion, 13*, 769-773.

McSorley, E. & Morriss, J. (2015). What you see is what you want to see: Motivationally relevant stimuli can interrupt current resource allocation. *Cognition and Emotion, 14*, 1-7.

Moher, J., Abrams, J., Egeth, H., Yantis, S., & Stuphorn, V. (2011). Trial-by-trial adjustments of top-down set modulate oculomotor capture. *Psychonomic Bulletin & Review, 18*, 897–903.

Nyström, M. & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods, 42*, 188-204.

Olejnik, S. & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*, 434-447.

Port, N., & Wurtz, R. (2003). Sequential activity of simultaneously recorded neurons in the superior colliculus during curved saccades. *Journal of Neurophysiology, 90*, 1887-1903.

Quaia, C., Lefèvre, P., & Optican, L. (1999). Model of the control of saccades by superior colliculus and cerebellum. *Journal of Neurophysiology, 82*, 999-1018.

Sheliga, B., Riggio, L., Craighero, L., & Rizzolatti, G. (1995). Spatial attention-determined modifications in saccade trajectories. *Neuroreport, 6*, 585-588.

Smeets, J., & Hooge, I. (2003). Nature of variability in saccades. *Journal of Neurophysiology, 90*, 12-20.

Tipper, S., Howard, L., & Paul, M. (2001). Reaching affects saccade trajectories. *Experimental Brain Research, 136*, 241-249.

Tudge, L. & Schubert, T. (2016). Accessory stimuli speed reaction times and reduce distraction in a target-distractor task. *Journal of Vision, 16*, 11.

Van der Stigchel, S. (2010). Recent advances in the study of saccade trajectory deviations. *Vision Research, 50*, 1619-1627.

Van der Stigchel, S. & Theeuwes, J., (2005). Relation between saccade trajectories and spatial distractor locations. *Cognitive Brain Research, 25*, 579-582.

Van der Stigchel, S. & Theeuwes, J. (2006). Our eyes deviate away from a location where a distractor is expected to appear. *Experimental Brain Research, 169*, 338-349.

Van der Stigchel, S., & Nijboer, T. (2011). The global effect: what determines where the eyes land? *Journal of Eye Movement Research, 4*, 1-13.

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience & Biobehavioral Reviews, 30*, 666-679.

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2007). The spatial coding of the inhibition evoked by distractors. *Vision Research, 47*, 210-218.

Van Opstal, A. & Van Ginsbergen, J. (1987). Skewness of saccadic velocity profiles: A unifying parameter for normal and slow saccades. *Vision Research, 27*, 731-745.

van Zoest, W., Donk, M., & Van der Stigchel, S. (2012). Stimulus salience and the time course of saccade trajectory deviations. *Journal of Vision, 12*, 1-13.

Vincent, S. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs, 1*, 5.

Viviani, P., Berthoz, A., & Tracey, D. (1977). The curvature of oblique saccades. *Vision Research, 17*, 661–664.

Walker, R., Deubel, H., Schneider, W., & Findlay, J. (1997). Effect of remote distractors on saccade programming: evidence for an extended fixation zone. *Journal of Neurophysiology, 78*, 1108-1119.

Walker, R., McSorley, E., & Haggard, P. (2006). The control of saccade trajectories: Direction of curvature depends upon prior knowledge of target location and saccade latency. *Perception & Psychophysics, 68*, 129–138.

Walker, R., & McSorley, E. (2008). The influence of distractors on saccade-target selection: Saccade trajectory effects. *Journal of Eye Movement Research, 2*, 1-13.

Weaver, M., Lauwereyns, J., & Theeuwes, J. (2011). The effect of semantic information on saccade trajectory deviations. *Vision Research, 51*, 1124-1128.

White, B., Theeuwes, J., & Munoz, D. (2012). Interaction between visual- and goal-related neuronal signals on the trajectories of saccadic eye movements. *Journal of Cognitive Neuroscience, 24*, 707-717.