Saccadic reaction times in infants and adults: spatiotemporal factors, gender, and inter-laboratory variation

Accepted in Developmental Psychology 14/03/17

Ben Kenward
Department of Psychology, Oxford Brookes University
Additional affiliation: Department of Psychology, Uppsala University
Corresponding author. ben.kenward@wolfson.oxon.org; +44 (0) 7857 445653; Oxford Brookes University, Headington Rd, Gipsy Ln, Oxford OX3 0BP, UK

Felix-Sebastian Koch
Department of Behavioural Sciences and Learning, Linköping University

Linda Forssman
School of Medicine, University of Tampere
Additional affiliation: Department of Psychology, Uppsala University

Julia Brehm
Department of Psychology, Uppsala University

Ida Tidemann
Department of Psychology, University of Oslo

Annette Sundqvist
Department of Behavioural Sciences and Learning, Linköping University

Carin Marciszko
Department of Psychology, Uppsala University

Tone Kristine Hermansen
Department of Psychology, University of Oslo

Mikael Heimann
Department of Behavioural Sciences and Learning, Linköping University

Gustaf Gredebäck
Department of Psychology, Uppsala University

Abstract

Saccade latency is widely used across infant psychology to investigate infants' understanding of events. Interpreting particular latency values requires knowledge of standard saccadic reaction times, but there is no consensus as to typical values. This study provides standard estimates of infants' (n=194, ages 9 to 15 months) saccadic reaction times under a range of different spatiotemporal conditions. To investigate the reliability of such standard estimates, data is collected at four laboratories in three countries. Results indicate that reactions to the appearance of a new object are much faster than reactions to the deflection of a currently fixated moving object; upward saccades are slower than downward or horizontal saccades; reactions to more peripheral stimuli are much slower; and this slowdown is greater for boys than girls. There was little decrease in saccadic reaction times between 9 and 15 month, indicating that the period of slow development which is protracted into adolescence begins in late infancy. Except for appearance and deflection differences, infant effects were weak or absent in adults (n=40). Latency estimates and spatiotemporal effects on latency were generally consistent across laboratories, but a number of lab differences in factors such as individual variation were found. Some but not all differences were attributed to minor procedural differences, highlighting the importance of replication. Confidence intervals (95%) for infants' median reaction latencies for appearance stimuli were 242 – 250 ms and for deflection stimuli 350 – 367 ms.

Keywords: Saccade; reaction time; latency; infants; replication; open science

Introduction

A saccade is an abrupt and rapid eye-movement serving to direct the photoreceptor-dense centre of the visual field – the fovea – at a target (Liversedge, Gilchrist, & Everling, 2011). New-borns can target objects using saccades. However, although the speed of eye-movements during saccades in young infants is no slower than in adults (Garbutt, Harwood, & Harris, 2006), the latency to react to a stimulus by beginning a saccade varies greatly according to infant age and situation type, as we discuss below. This variation is one focus of this study.

Saccade latency is one of the most frequently used measures in many areas of infant psychology research. Eye tracking studies rely on this measure directly to assess predictive abilities (e.g. Canfield, Smith, Brezsnyak, & Snow, 1997; Gredebäck & Falck-Ytter, 2015; Kenward, 2010), social cognition (e.g. Gredebäck & Melinder, 2010; Peltola, Leppanen, Palokangas, & Hietanen, 2008), priming (e.g. M. H. Johnson, Posner, & Rothbart, 1994), scanning of naturalistic scenes (e.g. Wass & Smith, 2014), object permanence (e.g. Bremner, Slater, & Johnson, 2015; Gredebäck & von Hofsten, 2007) and cognitive development (e.g. S. P. Johnson, 2003). Reaction times also impact looking time patterns during habituation (e.g. Spelke & Kinzler, 2007) and preferential looking paradigms (e.g. Atkinson, 2000) that have long been at the heart of infancy research.

In addition to being used as a dependent measure to assess other cognitive abilities, the development of the oculomotor system is its own field with a large range of studies (Luna, Velanova, & Geier, 2008; Rosander, 2007). Individual differences in saccadic reaction time (SRT) in infancy are robust over several months (Canfield, Wilken, Schmerl, & Smith, 1995; Haith & McCarty, 1990), and predict later Stanford-Binet IQ (Benson, Cherny, Haith,

& Fulker, 1993), processing speed (Jacobson et al., 1992) as well as white matter changes and ASD diagnosis at 24 (Elison et al., 2013) and 36 months of age (Elsabbagh et al., 2013)

Given the great importance to infancy research of measuring SRT, it is necessary to gain a better understanding of typical values. For example, infants' predictive gaze is frequently used as a dependent measure, and typical minimum SRT estimates are crucial to allow predictive saccades to be distinguished from reactive saccades, on the basis that predictive saccades are faster. However, there is a lack of consensus around typical minimum SRT, with values used varying between 133 and 233 ms (Canfield et al., 1997; Gredebäck, Johnson, & von Hofsten, 2010; Reznick, Chawarska, & Betts, 2000; Rose, Feldman, Jankowski, & Caro, 2002). Rose et al. (2002) conducted a sensitivity analysis and demonstrated that their conclusions about the longitudinal development of expectation learning were influenced by the choice of minimum SRT value (see also Gredebäck, Stasiewicz, Falck-Ytter, von Hofsten, & Rosander, 2009).

The primary aim of this study is therefore to provide comprehensive information as to infant SRT distributions across a range of ages, using a variety of unpredictable stimuli with different spatiotemporal properties. In order to fulfil this goal, a large sample is desirable, and to facilitate this we collect data at four different laboratories in three different Nordic countries: Norway, Sweden, and Finland. In the spirit of recent calls for increased replication within psychological research in general (Open Science Collaboration, 2015) and within infant studies (Frank et al., under review), a further aim is to take advantage of the multiple samples to examine whether SRTs are consistent and whether spatiotemporal effects on SRT are replicated across samples. We furthermore include an adult sample from each lab for comparison purposes.

One reason that previous estimates of typical SRTs have varied greatly is that SRT depends on the spatiotemporal stimulus properties. Generally, studies of infants' reactions to changing visual stimuli have included two broad types of stimulus change. New stimulus elements can appear (e.g. Canfield et al., 1997); and existing stimuli can move or deflect their movement (e.g. Gredebäck, Örnkloo, & von Hofsten, 2006). From comparing existing studies featuring these two types of event, it appears that reactions to unpredictable deflection are generally much later than reactions to unpredictable appearances. With respect to appearing stimuli, Canfield et al. (1997) demonstrated a decline in SRT from 440 ms at 2 months to 285 ms at 12 months. With respect to deflecting stimuli, Gredebäck et al. (2006) demonstrated a decline in SRT from 595 ms at 4 months to 442 ms at 8 months. However, to our knowledge, SRTs for movement and appearance have not been investigated in the same study, meaning that explanations for differences based on extraneous study differences cannot be ruled out. By presenting both event types in the same study, while keeping constant across event types potentially important parameters such as delay and location of event, we aim to provide a more standardised comparison of these event types than was previously available.

To maximise relevance of our results to other studies, we give our stimuli similar properties to those commonly reported in the literature. Appearing stimuli appear in the periphery following display of a central fixation stimulus (e.g. Hunnius & Geuze, 2004; Peltola et al., 2008). Unpredictable movement occurs in the form of a moving stimulus with constant velocity that suddenly changes direction. Very similar such deflecting stimuli have been used in studies of learning (e.g. Kochukhova & Gredebäck, 2007) and oculomotor control (e.g. Gredebäck et al., 2006), but reactions to such deflecting stimuli can also be of relevance for studies of action understanding in which infants track moving hands (Gredebäck & Falck-Ytter, 2015).

Within deflection trials, we additionally investigate the effect of direction of deflection. It has previously been found that vertical saccades have a longer SRT than horizontal saccades (Gredebäck et al., 2006), in line with other observations that infants' horizontal eye movements appear more mature than vertical eye movements (Richards & Holley, 1999). This may be because due to environmental demands; infants have more experience with horizontal than vertical eye-movements (Gredebäck et al., 2006). However, to our knowledge no study has compared upwards, downwards, leftwards, and rightwards saccades. We do so here on an exploratory basis.

Within appearance trials, we additionally examine the effect of the distance of appearing stimuli from the central fixation point by presenting stimuli paracentrally (on the macula but not the fovea) and fully peripherally. In adults, reaction time in similar conditions has been found to increase with distance from the centre (Ando, Kida, & Oda, 2001; Haines, 1975; Slater-Hammel, 1955), so we hypothesised that a similar effect might be found in infants.

For both deflection and appearance trials, we include a variable delay from fixation stimulus onset to deflection/appearance event. This variable is primarily included to increase unpredictability of the stimuli, and we make no prediction concerning its effect on SRT, but analyse its effects for exploratory reasons. We also explore the effects of gender: gender differences in infant vision are known (Alexander & Wilcox, 2012), although none of the known differences lead to specific predictions concerning gender and SRT.

To summarise, in addition to providing detailed information concerning infant SRT distributions under a range of spatiotemporal conditions and in different labs, we test the following hypotheses: SRT is faster in response to appearing stimuli than deflecting moving stimuli; SRT is faster for appearing stimuli that appear nearer to the fixation point; the

direction of movement deflection will affect SRT, with slower vertical SRTs; and SRT will reduce with age. We also investigate whether these effects differ at different labs and at different ages by including relevant interaction terms in our statistical models of SRT. We furthermore include in our models gender and event delay, although we do not include interactions with those terms in our initial models due to lack of predictions and the desirability of minimising the number of unnecessary interaction terms. We focus on 9- to 15-month-olds because this is an often assessed age range in studies using SRT as a tool, but most previous studies providing infant SRT estimates have assessed a younger age range, and there is therefore currently a paucity of standardised SRT data for older infants (Alahyane et al., 2016).

Methods

Internal replication and Open Science

Data was collected at four different labs from three different Nordic countries using three different models of Tobii eye-tracker. The measures employed to ensure standardization across labs were similar to what would normally be expected when one lab replicates another's procedure with their help. These measures therefore included exchange of stimuli, project files, written procedure descriptions, and questions and answers, but did not include visits between labs to ensure total standardization. These measures were adopted for practical reasons and because they were in line with the goal of investigating how well results from infant eye-tracking studies replicate across labs. Because the labs possessed eye-tracker screens with different physical sizes and native resolutions, this resulted in the stimuli being presented at slightly different sizes in the different labs (see Stimuli).

Our experiment and analysis can be replicated by downloading a method, data, and analysis package from an Open Science Framework repository (https://osf.io/hdngq). This

repository includes the E-prime experiment package, all raw data, a Perl script which extracts reactive saccades from gaze data files, an R script which conducts the statistical analyses and produces the visualisations, and additional documentation to facilitate replication and re-analysis.

Participants

APA ethical standards were complied with and every lab obtained ethical approval for the procedure from the appropriate local committee. Participants were healthy and from volunteer families recruited by mail sent to all local parents of babies of appropriate ages, with addresses taken from population registers and pre-existing volunteer pools in three medium sized Nordic cities and one Nordic capital city, with no special selection criteria except for the exclusion of pre-term birth infants (gestational age < 38 weeks). As such, participants' families were predominantly middle-class, of white European ethnicity, and well-educated. All parents or adult participants gave informed written consent.

The number of participants (Table 1) was determined by what was practical for each individual lab. The mean ages of the four age groups were 9.2 months (*SD* = .4), 12.1 months (*SD* = .3), 15.1 months (*SD* = .3), and 30.4 years (*SD* = 7.2). Data from all participants from Labs A and C was included in analysis, from Lab B one participant was excluded because of fussing at the procedure start, and from Lab D one participant was excluded because of calibration failure.

Stimuli

The four labs used three different types of Tobii eye-tracker with different maximum frame rates, native screen resolutions, and physical sizes, and used different approaches (centring, stretching, or perfect fit due to match with native resolution) to display the 1280 x

1024 pixel stimuli. As a consequence, there were slight differences in apparent size (in visual degrees) of stimulus elements. For stimulus size parameters, we therefore report mean values in the text but specific values for each lab in Table 2, which also describes hardware. All stimulus films were displayed at 60 frames per second. The E-prime software package (Psychology Software Tools, Inc.) was used to present the stimuli and record the data.

Table 1. Numbers of participants included in analysis (female numbers in parentheses)

Age	Lab A	Lab B	Lab C	Lab D	All
9 months	15 (7)	10 (4)	23 (15)	20 (11)	68 (37)
12 months	11 (6)	11 (3)	17 (9)	22 (10)	61 (28)
15 months	12 (6)	14 (6)	22 (11)	17 (8)	65 (31)
All infants	38 (19)	35 (13)	62 (35)	59 (29)	194 (96)
Adults	8 (5)	8 (3)	12 (8)	12 (10)	40 (26)

Table 2. Hardware and stimulus parameters in the different labs

	Lab A	Lab B	Lab C	Lab D
Tobii eye-tracker model	T60	T120	TX300	TX300
Recording frame rate (Hz)	60	120	300 then 60	300 then 60
Physical screen size (cm)	33.7 x 27.0	33.7 x 27.0	51.0 x 28.5	51.0 x 28.5
Native resolution (pixels)	1280 x 1024	1280 x 1024	1920 x 1080	1920 x 1080
Adaption of stimuli to screen	Perfect fit	Perfect fit	Centred, not enlarged	Stretched to fit
Screen area used for display (cm)	33.7 x 27.0	33.7 x 27.0	34.0 x 27.0	51.0 x 28.5
Screen area used for display (°)	29 x 24	29 x 24	30 x 24	40 x 25
Distance from centre to near events (°)	5.9	5.9	6.0	7.5
Distance from centre to far events (°)	13.3	13.3	13.4	16.7
Appearance fixation circle diameter at maximum size (°)	1.9	1.9	1.9	2.3
Appearance rectangle size (°)	4.5 x 7.1	4.5 x 7.1	4.6 x 7.1	6.2 x 7.4
Deflection circle diameter (°)	1.2	1.2	1.2	1.5
Deflection circle speed (°/s)	7.3	7.3	7.3	9.1

Note: visual degree (°) parameters are estimated based on a distance of 60 cm between screen and eyes.

Delay and distance from the central point for appearance and deflection events were standardised for both stimulus types. Short delay was 2650 ms and long delay was 3650 ms, with short delay stimulus clips lasting 4167 ms and long 5167 ms. Near and far events were

centred on points on the lines joining diagonally opposite corners, 6.3° and 14.2° from the screen centre respectively (mean values across labs).

Appearance stimuli (e.g. Supplementary Videos 1 & 2) began with a screen-centred red fixation circle slowly pulsing in size (pulse period 1667 ms) with a maximum diameter of 2.0° (mean across labs). After a random long or short delay, simultaneously the fixation circle disappeared and an appearance rectangle appeared in a random screen corner, randomly either near to (paracentrally, 6.3°) or far (peripherally, 14.2°) from the centre (means across labs). The rectangle measured 5.0° x 7.1° (mean across labs) and consisted of a white background containing either an emotional or neutral adult face or an ovoid face silhouette filled with noise from the same colour spectrum.

Deflection stimuli (e.g. Supplementary Videos 3 & 4) consisted only of a moving red circle of diameter 1.3° (mean across labs), initially travelling from the far location in one random corner towards a location (near or far at random) in the diagonally opposite corner. On reaching this opposite location, the circle deflected to move either horizontally or vertically at random. For example, when the circle began in the bottom left, it could deflect downwards or leftwards at the near or far point. Movement speed throughout was a constant 7.5°/s (mean across labs) meaning that near or far deflection location was confounded with delay (it took longer to reach the further location). This was acceptable because neither variable was of interest: for deflection these variables' purpose was to create unpredictability.

For both deflection and appearance trials, manipulations were included which varied the social nature of the stimuli. However, the focus of this manuscript is solely on spatiotemporal determinants of and inter-laboratory variation in SRT – the results of these manipulations will be reported elsewhere (Kenward et al., in prep.). For deflection, half the participants saw additional familiarisation stimuli intended to establish the red ball as an

animate agent (it moved in a goal-directed biological manner between objects) and half saw stimuli intended to establish the ball as inanimate (it bounced off objects mechanically). For appearance, three-quarters (within-subjects) of the appearing objects were faces, and one-quarter were perceptually similar non-face stimuli (ovoid face silhouettes filled with noise). Note that these variables, although not analysed here, were counter-balanced with the reported variables.

Procedure and display sequence

After explaining the procedure to the parent or adult participant and obtaining consent, the participant was seated in their parent's lap with their eyes approximately 60 cm from the screen, and the standard Tobii calibration procedure was run using five- or nine-points according to each lab's experience of what worked best for them (Gredebäck et al., 2010). The stimulus sequence was then displayed until the end or until the participant became too fussy to continue viewing.

Sixteen appearance stimuli were created by fully counterbalancing appearance corner, distance, and delay. Sixteen deflection stimuli were created by fully counterbalancing corner, distance, and deflection direction. Each participant viewed one of four different pseudo-random presentation orders in which no more than two stimuli in a row were appearance or deflection. Each stimulus was presented together with a short sound chosen from a collection of 16 short sounds such as bells and horns (stimulus video and sound pairings were different for the four different stimulus orders). In addition, at the start (twice) and after every eighth stimulus (once), familiarisation stimuli were presented, each lasting 7 seconds. Half the participants saw a red ball moving in a goal-directed biological manner between objects; the other half saw the ball bouncing off the same objects mechanically. The entire stimulus set including 32 test stimuli and 9 familiarisation stimuli was presented twice, leading to a total

presentation time of approximately six minutes if the procedure was continued to the end of the stimulus set.

Initially all labs collected data at the maximum rate for their eye-tracker (Table 2). However, when data collection was already underway, it was discovered that the highest rate of 300 Hz, used by two labs, was resulting in many missed frames, and these labs therefore reduced their data rate to 60 Hz for the remaining participants. Before the eye-tracking procedure, infants also participated in a behaviour task lasting approximately one minute. Parents also filled out questionnaires. These measures are not analysed here but full details are available in the Open Science Framework repository.

Reactive saccade identification

Raw gaze data was obtained directly from the eye-tracker TET server using E-Prime Extensions for Tobii. The gaze point was the average for the two eyes if both were tracked, except that if the validity score was lower for one eye or if only one was tracked, only one eye was used. Gaze data at 300Hz was smoothed with a five-point moving average to remove high frequency noise, making it more standardised with respect to the lower frequency data. Raw gaze data was otherwise unprocessed prior to saccade detection. For example, there was no interpolation of missing data, although due to smoothing, a period containing missing frames at 300 Hz might have no missing frames after smoothing. Although stimulus parameters differed slightly between labs, analysis parameters were identical between labs. A saccade was defined to begin when gaze movement speed exceeded 30°/s, as long as movement slower than 30°/s was detected within 0.1 s of the saccade start, at which point the saccade was defined to end. In other words, a saccade is detected when a period of fixation with sub-30°/s movement contains a period of faster movement lasting less than 0.1 s. The threshold value of 30°/s was chosen because visual examination of velocity profiles indicated

it produced few false positives in distinguishing saccades from other velocity spikes, and was in line with previous infant saccade analysis (Gredebäck et al., 2006). False positives, due to occasional measurement error producing apparently artefactual movements with high speed but low amplitude (jitter), were minimised by a requirement that saccades be at least 0.5° in amplitude. For a saccade to be valid, the eye-tracker had to have registered valid coordinates throughout the time of the saccade.

Reactive saccades were defined as beginning within 0.1 and 1.0 seconds of the appearance or deflection event. This lower cut-off was chosen to be well below any minimum SRT previously known in infants. For appearance, valid reactive saccades began in a circular area of radius 2.0° centred on the central fixation point and ended within an area encompassing the appearing stimulus rectangle and all points within 0.33° of it. For deflection, valid reactive saccades began within a circle of radius 2.5° centred on the moving circle at point of deflection, and ended at any point which was in the right direction relative to the starting point. The right direction was defined as being within 45° of the direction of post-deflection movement.

Analysis strategy and statistics

To test our hypotheses, we use general linear mixed models implemented using the nlme package in the R programming environment (version 3.2.2, R Core Team, 2015). To account for the within-subject design, participant is included as a random factor; all other model variables are categorical fixed factors. SRT was right-skewed, but after square root transformation (previously used for infant SRT data, Hunnius & Geuze, 2004), models were found to have acceptable fit, as assessed through inspection of diagnostic scatter plots of the residuals.

Because we use mixed models which yield separate variances for random and fixed effects, most standard effect sizes are unavailable. We utilise Bartoń's (2015) implementation in R of Nakagawa and Schielzeth's (2013) $R^2_{GLMM(m)}$, which measures fit of the fixed components of the model, and $R^2_{GLMM(c)}$, which measures fit for fixed and random components together. Because $R^2_{GLMM(c)}$ is analogous and interpretable similarly to standard R^2, we state it as a measure of overall model fit. Effect sizes for individual fixed factors are stated as ΔR^2, defined as the reduction in $R^2_{GLMM(m)}$ when that factor and its interactions are removed from the model, but all other factors remain.

For visualisation of SRT distributions, we pool all saccades in the relevant category and display violin style kernel density plots, using Scott's (1992) rule of thumb for bandwidth estimation, but with density estimate clipped at the extremes of the data. On the same figures we plot group means and 95% confidence intervals for individual means.

Results

Data quality

Initially the data was inspected to confirm that infants had maintained attention to the stimuli and that eye-tracking had functioned well. As expected, reduced numbers of trials for infants compared to adults was due to substantially reduced attention over the course of the session (Table 3, see Table S1 for this information additionally broken down by infant age). Two labs had begun sampling at 300 Hz but reduced to 60 Hz mid-way through data collection because 300Hz sampling apparently led to poor data quality (Table 3). Because the number of valid reactive saccades was similar across labs when these labs sampled at 60 Hz (Table 3), the higher level of data loss prior to this adjustment is mainly attributable to intermittent eye-tracking failure at 300 Hz. An analysis of the sensitivity of mean SRT to inclusion of individuals contributing few data points demonstrated very little effect (see

Supplementary Analysis). All tracked saccades from all labs are therefore included in analysis (2577 for infants and 1580 for adults). However, we note that data quality was not identical across labs even when all sampled at a lower frequency, which could be accounted for by differences in session length due to differing tolerance for fussiness between labs (see supplementary analysis, including table S2).

Table 3. Description of reactive saccade samples from different labs

		N trials with a tracked reactive saccade				Proportion of frames with tracked gaze			
		Appearance		Deflection		Session minute 1		Session minute 5	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
					Infants				
All	194	6.6	7.2	6.7	6.1	.57	.33	.27	.30
Lab A	38	11.0	7.3	8.5	5.3	.80	.24	.29	.31
Lab B	35	10.1	7.5	8.4	5.9	.65	.32	.24	.29
Lab C 300 Hz	21	2.6	3.5	4.1	4.3	.34	.23	.10	.18
Lab C 60 Hz	41	6.7	7.2	8.3	7.8	.63	.33	.46	.33
Lab D 300 Hz	52	2.6	4.6	4.1	4.5	.37	.27	.16	.23
Lab D 60 Hz	7	5.9	7.0	6.3	6.2	.86	.09	.37	.36
					Adults				
All	40	19.6	10.7	19.9	8.6	.93	.09	.89	.18
Lab A	8	26.5	6.7	27.0	4.7	.96	.03	.95	.04
Lab B	8	19.5	7.3	23.6	9.4	.88	.14	.76	.35
Lab C 60 Hz	12	17.6	12.6	19.6	8.2	.95	.05	.94	.05
Lab D 300 Hz	12	17.0	11.7	13.1	5.1	.92	.09	.88	.11

Note: the maximum possible number of trials with a tracked reactive saccade is 32 for both stimulus types.

Summary of spatiotemporal effects and inter-lab differences

Table 4 shows summary statistics for SRTs, separated by all factors found to have significant effects, except for the effects of laboratory. Statistical models comparing infants and adults are presented only as supplementary information – these comparisons are obvious from the graphical summaries.

SRT is slower for infants than for adults for all types of investigated events (Models S1, S2, and S3, Figure 1). Contrary to expectations, there were no main effects of infant age

(Models 1, 2, and 3), but infant age interacted with lab for both appearance (Model 2) and deflection (Model 3) stimuli. We return to the issue of development below. SRT for appearance is faster than for deflection for all ages (Models 1 and S1, Figure 1).

Location of appearance stimuli influenced SRT, with slower responses to stimuli appearing further from the fixation point, but this effect was much stronger for infants than for adults (Models 2 and S2, Figure 2). For appearance stimuli only there was an unexpected effect of delay time, with a longer delay resulting in a very slightly slower SRT in infants but not adults (Models 2 and S2, Figure 3). There was also an unexpected infant gender effect, with girls slightly faster than boys in response to appearance, for far stimuli only (Model 2, Figure 4).

Direction of deflection influenced SRT, with responses to upwards movement slower than all other directions for infants but no differences between other directions (Model 3, Figure 5). This effect differed between labs. There were less clear indications of a similar effect in adults (Model S3).

Table 4. SRT descriptive statistics separated by conditions causing significant differences.

Age	Event type	Appearance location	Deflection direction	Gender	Delay	n	M (ms)	SD (ms)	Mdn (ms)
Infant	Appearance	All	-	All	All	148	277	93	258
Infant	Appearance	All	-	All	Long	129	278	83	267
Infant	Appearance	All	-	All	Short	128	266	101	247
Infant	Appearance	Near	-	All	All	140	252	99	233
Infant	Appearance	Far	-	All	All	115	306	61	295
Infant	Appearance	Far	-	Female	All	61	293	49	291
Infant	Appearance	Far	-	Male	All	54	321	69	315
Adult	Appearance	All	-	All	All	38	164	24	164
Adult	Appearance	Far	-	All	All	37	171	27	167
Adult	Appearance	Near	-	All	All	36	160	24	161
Infant	Deflection	-	All	All	All	165	375	109	363
Infant	Deflection	-	Up	All	All	136	395	131	382
Infant	Deflection	-	Down	All	All	126	357	126	340
Infant	Deflection	-	Horizontal	All	All	146	363	97	363
Adult	Deflection	-	All	All	All	40	265	60	248
Adult	Deflection	-	Up	All	All	40	273	76	257
Adult	Deflection	-	Down	All	All	37	258	71	250
Adult	Deflection	-	Horizontal	All	All	39	254	59	229

Note: These group summaries are of individuals' mean values within each condition combination. Some individuals contribute single data points to their individual mean, but excluding these individuals had almost no appreciable effect (see Supplementary Analyses).

There was a main effect of lab on SRT (Model S1), but this was due only to differences in response to deflection stimuli when adult data was included (Model S3, Figure 6), and was not found in response to appearance stimuli (Model S2) or when only infants were analysed (Model 1, 2, & 3). For deflection stimuli, one lab in particular (Lab D) had longer deflection SRTs for adults. This lab happened to be the one with the largest screen display, and therefore had faster moving stimuli (see below). A replication check of all major effects found they were all replicated in at least two labs, and all but one replicated in at least three labs (see below).

Minimum likely SRTs

Table 5 shows confidence intervals for estimates of some lower percentiles of the population SRT distributions. This information is informative as regards the likely lowest latencies for different types of reactive saccade.

Table 5. 95% confidence intervals for percentiles of the population SRT distributions

			95% CI for percentile (ms)		
Stimulus type	Age	*N* saccades	5%	(lower quartile) 25%	(median) 50%
Appearance	Infant	1278	167 – 183	200 – 208	242 – 250
Appearance	Adult	783	117 – 123	133 – 147	150 – 167
Deflection	Infant	1299	150 – 177	300 – 317	350 – 367
Deflection	Adult	797	133 – 150	200 – 200	225 – 233

Note: Confidence intervals are calculated using the binomial method (Conover, 1999, p.145). Saccades from all individuals are pooled within a category.

Statistical models of factors influencing SRT

Model 1: Appearance versus deflection

The model ($R^2_{GLMM(c)}$ = .35) was constructed with all infant data and with the fixed factors shown in Table 6.

Table 6. Determinants of SRT for appearance and deflection in infants (Model 1)

Factor	*df*	*F*	*p*	ΔR^2
Type (appearance vs. deflection)	1,2394	752.9	.000***	.202
Age (9 vs. 12 vs. 15 months)	2,157	0.2	.802	.018
Lab	3,157	1.9	.136	.023
Gender	1,157	0.4	.544	.000
Delay (long vs. short)	1,2394	3.2	.072	.001
Age x Type	2,2394	1.0	.354	.006
Age x Lab	6,157	2.1	.056	.018
Lab x Type	3,2394	1.0	.393	.006
Lab x Age x Type	6,2394	3.6	.002***	.006

Note: * $p < .05$; ** $p < .01$; ***: $p < .001$

Model 2: Factors influencing SRT for appearance stimuli

The model ($R^2_{GLMM(c)}$ = .65) was constructed with all infant appearance data and with the fixed factors shown in Table 7. Contrary to predictions, there was a gender effect, with girls having shorter SRTs than boys. Because of this unexpected effect, a follow-up model was constructed using the same original factors, plus the interactions of gender with distance, age, and lab, in order to determine whether the effect of gender depended on those variables. The gender interactions with age and lab were not significant (ps > .5, ΔR^2 values ≤ .005), but the interaction between gender and distance was significant (p = .006, ΔR^2 = .005). Follow-up models separated by distance demonstrated that the effect of gender held for far appearances (p = .009, ΔR^2 = .063), but not near appearances (p = .233, ΔR^2 = .013), as illustrated by Figure 4.

Due to an unexpected (and very small) but significant delay effect, a follow-up model was constructed using the same original factors, plus the interactions of delay with type, age, and lab, in order to determine whether the effect of delay depended upon those variables. None of these interactions were significant (ps > .2).

The focus of this manuscript is not on the social aspects of the displayed stimuli, but in this context it is important to know whether the gender effect was because three-quarters of the appearing stimuli were faces, or whether it also held for the non-face noise stimuli. To test this, we repeated the model with gender interactions, also including the interaction between gender and appearance type (face vs. noise). This interaction was not significant (p = .670, ΔR^2 = .045). Furthermore, an additional follow-up model, including only far appearances which were noise, indicated a near significant effect of gender (p = .057, ΔR^2 = .050). Note that power is seriously reduced when only this subset (one-quarter of the appearance trials) is included.

Table 7. Determinants of SRT for appearance stimuli in infants (Model 2)

Factor	*df*	*F*	*p*	ΔR^2
Distance (near vs. far)	1,1117	470.6	.000***	.139
Age (9 vs. 12 vs. 15 months)	2,135	0.3	.746	.042
Lab	3,135	2.0	.114	.059
Gender	1,135	4.1	.046*	.015
Delay (long vs. short)	1,1117	13.8	.000***	.003
Age x Distance	2,1117	1.1	.325	.003
Age x Lab	6,135	2.4	.028*	.043
Lab x Distance	3,1117	1.9	.128	.005
Lab x Age x Distance	6,1117	1.7	.107	.003

Note: * $p < .05$; ** $p < .01$; ***: $p < .001$

Model 3: Factors influencing SRT for deflection stimuli

The model ($R^2_{GLMM(c)} = .18$) was constructed with all infant deflection data and with the fixed factors shown in Table 8. Because of the significant deflection direction effect (Figure 5), we ran follow-up models with the same factors, but each including data from only two deflection directions, in order to make each specific pairwise direction comparison. SRT was significantly slower for upwards deflection compared to all other directions (p-values ≤ .017 and ≥ .001, ΔR^2 values ≤ .044 and ≥ .032) , but there were no other differences.

Table 8. Determinants of SRT for deflection stimuli in infants (Model 3)

Factor	*df*	*F*	*P*	ΔR^2
Direction (up vs. down vs. left vs. right)	3,1097	5.9	.001***	.045
Age (9 vs. 12 vs. 15 months)	2,152	0.7	.502	.038
Lab	3,152	1.9	.127	.058
Gender	1,152	0.1	.823	.000
Delay (long vs. short)	1,1097	0.0	.836	.000
Age x Direction	6,1097	1.4	.211	.015
Age x Lab	6,152	2.3	.036*	.030
Lab x Direction	9,1097	3.3	.000***	.029
Lab x Age x Direction	18,1097	0.9	.601	.009

Note: * $p < .05$; ** $p < .01$; ***: $p < .001$

Replicability of the effects

For each of the stimulus property effects summarised above, we examined whether the effect was replicated across different labs by recreating the relevant models for each lab's data separately. All effects were replicated across at least two labs, and the stronger effects were replicated by all labs (Table 9). We note that even the effects not initially hypothesised did replicate. The gender effect for far appearing stimuli, although clear in the pooled sample, was significant in only two labs.

Table 9. Effect size (ΔR^2) and statistical significance of effects for all data and by lab

Effect	All	Lab A	Lab B	Lab C	Lab D	*N* replications
Deflection vs. appearance	.175***	.176***	.246***	.125***	.193***	4
Infants vs. adults	.233***	.341***	.325***	.132***	.128***	4
Appearance distance in infants	.139***	.202***	.200***	.066***	.201***	4
Appearance delay in infants	.003***	.004*	.008**	.003**	-.001	3
Far appearance by infant gender	.063**	.147*	.106*	.036	.069	2
Deflection direction	.019***	.018***	.013**	.008†	.044***	3

Note: † $p < .1$; * $p < .05$; ** $p < .01$; ***: $p < .001$

Focus on inter-lab differences

Sizeable lab differences in SRT were found in response to deflection events. Visual inspection of the data revealed that this effect was driven by one lab (Lab D) having considerably higher SRT for deflection in adults (Figure 6). Although this figure hints at a similar but weaker effect in infants, no such effect was detected, consistent with Model S3's detection of a difference between infants and adults with respect to the lab effect. Note that this lab difference was not present for appearance stimuli, which is why Model S1 indicated lab differences in the effect of appearance versus deflection in adults.

Figure 6 also suggests that Labs C and D produced more variable SRTs for deflection in infants than Labs A and B (the density plots have longer tails). This effect is also apparent

in a plot of individual mean values (Figure 7). Levene's tests confirmed that the labs differed in the amount of variation between individual infants' SRTs for deflection stimuli, $F(3,161) = 4.0$, $p = .008$, although not for appearance stimuli, $F(3,144) = 1.7$, $p = .166$. One possible reason for this difference is the presence of poorer quality 300 Hz data from Labs C and D, but after removing this data, the difference in SRT variability for deflection stimuli remained, $F(3,103) = 3.7$, $p = .014$. Because greater variability could affect estimates of minimum likely SRT, the estimates presented in Table 5 were recalculated without the data from Labs C and D. With the exception of the lower 5th percentile for infant deflection stimuli, which had a central estimate 20 ms later, differences were negligible (Table S6).

Focus on infant development

The other inter-lab differences were in the form of interactions between lab and age. Visual inspection of age regression plots for each lab (Figure 7) indicated that for both appearance and deflection, although there were no significant effects of age, three of four labs evidenced a trend for reduction of SRT with age. The fact that the trend-violating labs were different for the different stimulus types is in line with the Model 1 interaction between age, stimulus type, and lab. The lack of obvious non-linearity justifies the inclusion of age as a covariate rather than categorical factor, and versions of Models 2 and 3 with this modification were created. Because SRT is square root transformed in our models, age regression coefficients are not directly interpretable. However, back-transformation allows a gradient to be calculated at specific SRT values. An infant with the mean value of 277 ms for appearance stimuli is predicted to experience a change of -3 ms after one month, 95% CI [-13,7]. For deflection stimuli an infant with the mean value of 375 ms is predicted to experience a change in SRT of -10 ms after one month, 95% CI [-23,3]. Linear extrapolations of these values result in adult mean values (Table 4) being reached in early childhood. However, even

after the exclusion of Lab C from the appearance model, there is no significant age effect, $F(1,94) = 1.5$, $p = .220$, $\Delta R^2 = .006$, although exclusion of Lab D from the deflection model produces a significant age effect, $F(1,111) = 7.6$, $p = .007$, $\Delta R^2 = .023$.

Discussion

Testing the saccadic reaction times (SRTs) of almost two-hundred infants from four labs in three countries under a variety of spatiotemporal conditions revealed a number of expected and unexpected effects. We now discuss the implications of the results, beginning by focussing on the consequences for attempts to distinguish between reactive and predictive saccades by establishing minimum likely infant SRTs.

Minimum likely infant SRTs

The shapes of the infant SRT distributions we obtained indicate minimum likely SRTs in the sampled population. For both appearance and deflection stimuli, only 5% of reactive saccades would be earlier than around 170 ms, which is within the reasonably narrow 95% CI for the lower 5th percentiles for both stimulus types (Table 5). As reviewed earlier, the cut-off thresholds which have been used to define the lower limit of purely reactive saccades in previous studies have ranged between 133 and 233 ms (Canfield et al., 1997; Gredebäck et al., 2010; Reznick et al., 2000; Rose et al., 2002). Based on our data, 133 ms is unnecessarily conservative, especially for appearance stimuli, but the commonly used threshold of 200 ms is too liberal when considering individual saccades. This value falls within our 95% CI for the lower quartile for appearance stimuli. It is therefore not generally justified to assume that an infant saccade faster than 200 ms is predictive – we expect around a quarter of reactive saccades to be this fast in this appearance paradigm.

However, in many studies the important issue is not what proportion of saccades should be considered too early to be reactive. Rather, the issue is whether an average SRT for an entire sample is too early for the whole sample to be reactive. Our estimates of population central tendencies are considerably later than 200 ms (the 95% CI for the median is 242 to 250 ms for appearance, and later still for deflection). It is therefore reasonable to assume that SRT samples which are on average earlier than 200 ms constitute evidence of expectation in similar paradigms. Given the lower confidence limits for the medians, the commonly used comparison value of 200 ms can in fact be regarded as unnecessarily conservative, and samples from similar paradigms with medians lower than 242 ms are likely to include predictive saccades.

Generally, our appearance stimuli were of a type likely to produce fast SRTs – the stimuli were visually salient, included near (paracentral) appearances, and there was no overlap between the fixation stimulus and the appearance stimulus (which can produce "sticky fixation", Hunnius & Geuze, 2004). However, one caveat is that SRTs might have been slightly earlier if the fixation stimulus had disappeared before the appearance stimulus (the "gap/overlap" paradigm, Peltola et al., 2008). We did not vary the offset between the fixation stimulus disappearance and the subsequent appearance because it was not feasible to manipulate further variables, given the already complex design. We note, however that the difference in SRT between gap trials and no-gap trials is typically not great – for example, one study found a mean difference for 11-month-olds of 14 ms (Wass, Porayska-Pomsta, & Johnson, 2011).

Appearance versus deflection stimuli

Previous work suggested that infants have considerably shorter SRTs in response to suddenly appearing stimuli (Canfield et al., 1997) than in response to direction change of

tracked moving stimuli (Gredebäck et al., 2006). However, as these event types had not been included in the same study it was previously possible that this was due to extraneous factors. The current study indicates that this effect is real and strong, with responses to appearing stimuli almost 100 ms faster. This result highlights the fact that details of the specific task will have large effects on infant SRTs, a practical issue that needs to be taken into account in any study using infant SRT as a response measure.

There might be several potential sources for these differences. As noted above, when an initial fixation image is maintained on the screen, appearance SRT increases (sticky fixation). Perhaps the same difficulty disengaging occurs when the attended object does not disappear but rapidly change its direction. Another factor that impacts deflection but not sudden appearance tasks is the presence of a visual buffer representing how a moving object will travel over time (Grönqvist, Gredebäck, & Hofsten, 2006). In the deflection paradigm the predictive buffer assumes that the object will continue on the same path and it might take time to overcome this expectation. No such visual buffer is assumed to exist in the appearance paradigm since images appear in consecutive locations without visible movement between the two. In other words, it is likely that differences in SRT between the two paradigms are caused by differences in the processes that guide tracking of smooth and continuous trajectories and suddenly reappearing images.

The effect of appearance distance from the fixation point

There was a strong effect of appearance distance in infants: SRTs for paracentral stimuli were much earlier than for peripheral stimuli (54 ms). For adults, this difference was small, only 9 ms, although still significant. The weakness of this effect in adults is consistent with previous research showing large effects only at greater eccentricities than investigated here (Haines, 1975; Slater-Hammel, 1955). The reason why this effect is so much greater in

infants is not currently clear, although neuro-imaging research indicating different processing speeds for central and peripheral stimuli (Stephen et al., 2002) may provide clues for future work. However, we again note that this result highlights the fact that small differences in stimulus properties (in this case, an eight visual degree difference in eccentricity) can have profound differences on infant SRT. This fact needs to be taken into account in any study interpreting infant SRT.

Gender differences

For far appearance stimuli only, girls had shorter SRTs: the mean difference between boys and girls was 28 ms. There was no such difference in adults. The effect was replicated in only two labs, but an absence of relevant significant interaction effects indicates that this was due to lack of power. This effect was not expected, but gender differences in infant visual perception and motor control are known to exist (Alexander & Wilcox, 2012). There are several possible explanations for this effect. It is possible that girls have superior visual perception in this respect, but it is also possible that their orienting responses once the stimulus is perceived are faster. However, for children as young as four years, boys tend to have faster reaction times (Dykiert, Der, Starr, & Deary, 2012), although we are not aware of any studies which have tested younger children. On the other hand, there are studies showing that infant girls are superior with respect to some aspects of visual perception (Alexander & Wilcox, 2012). For example, infant girls have more mature visual-pattern-evoked event-related potentials (Malcolm, McCulloch, & Shepherd, 2002). Furthermore, adult women have superior peripheral colour vision, whereas the evidence is less clear for such gender differences in the central visual field (Murray, Parry, McKeefry, & Panorgias, 2012). Together this prior evidence suggests that the current result is likely to reflect gender

differences in infants' peripheral vision rather than gender differences in their orienting responses.

The effect of deflection direction

The finding of later SRTs in response to vertical movements is in line with previous work showing that infants have superior oculomotor control with respect to horizontal movement (Gredebäck et al., 2006; Richards & Holley, 1999). This was suggested to depend upon the fact that the environment provides more opportunities for infants to train horizontal than vertical movement. However, previous studies of SRT have not separated upward from downward saccades. Here, we demonstrated that it is only upward and not downward saccades that are later than horizontal saccades. While we are not aware of previous adult work separating upward and downward SRT, there are differences in the velocity profiles of the saccadic movements themselves (Collewijn, Erkelens, & Steinman, 1988). Also, infants have more experience with downwards than upwards optic-flow, suggesting a possible experience dependent influence (Gilmore, Raudies, & Jayaraman, 2015). Explanations for these differences, and the possible connection between the current finding and this previous work, must await further investigation, although we note that the effect is rather small and may not have a great deal of practical significance.

The effect of delay

For appearance stimuli, we found that SRTs for long delay trials were slightly longer. This effect was small – the mean difference was only 12 ms. Although there was no significant interaction between delay length and stimulus type, no such effect was found for deflection trials, so the lack of interaction was probably due to low power to detect an interaction for this weak effect. We had not predicted this effect, but given its weakness and the inability of the current design to distinguish between possible explanations, we do not

discuss it further except to note that the result illustrates how high-powered studies are able to detect effects of strong statistical but limited practical significance.

Inter-laboratory differences unrelated to age

The main inter-laboratory difference of note was the finding that one lab had much later SRTs for deflection stimuli for adults, although the effect was not detected in infants or with respect to appearance stimuli. We can identify one likely cause of this effect. Of the two labs with larger screens, one lab (Lab D) stretched the stimuli out across the whole screen, whereas the other lab (Lab C) centred the stimuli. Because the deflection stimuli moved at a rate which was constant in terms of pixels per second, this lab therefore had deflection stimuli which moved at 9.1°/s rather than 7.3°/s for the other labs (Table 2). This difference might have resulted in later SRTs for participants who were tracking a faster moving object.

Additionally, an interaction between lab and deflection direction indicated that the effects of deflection direction were different between labs. This might also be explained by this difference in stimulus presentation – in stretching a non-widescreen stimulus to fit a widescreen, a slight distortion in horizontal/vertical aspect ratio occurred, meaning that in contrast to the other labs, movements following horizontal and vertical deflections had slightly different speeds.

The difference in stimulus presentation was not intentional, but was a consequence of replicating the method across labs without perfect cross-checking of all aspects of stimulus presentation. This result underlines how small, unintended, and potentially unnoticed procedural differences can cause unexpected differences in results when replicating a study.

A further laboratory difference relates to differences in sampling frequency. Before those labs which began sampling at 300 Hz decreased to 60 Hz, a much greater number of tracking frames were missed. Although infant eye-tracking results can be affected by data

quality (Wass, Forssman, & Leppänen, 2014), here there were no main effects of lab on infant SRT. This does however raises the question of whether 300 Hz is an appropriate sampling frequency for infants when using the Tobii TX300. The current study was not designed to systematically investigate this issue and firm conclusions are therefore unwarranted. We further note that other infant studies conducted by participating labs (e.g. Leppänen, Forssman, Kaatiala, Yrttiaho, & Wass, 2015; Peltola, Forssman, Puura, van IJzendoorn, & Leppänen, 2015) have obtained satisfactory data quality when sampling at 300 Hz. We also note that a number of parameters such as background illumination and head position affect tracking quality (Tobii Technology AB, 2013), and that sampling frequency might interact with these and also with stimulus-specific factors in determining tracking quality. We therefore do not recommend avoiding tracking infant gaze at 300 Hz, but do recommend caution and consideration of these factors when choosing sample frequency.

Labs C and D obtained infant SRTs with greater variability than Labs A and B, for deflection stimuli but not for appearance stimuli. The apparent specificity of this effect to one stimulus type and its independence from sampling frequency suggests that this is not a result simply of differences in apparatus between the labs. It could result from minor procedural differences that differently affected the two conditions. For example, eye-tracking accuracy when moving stimuli (deflection) are fixated might be more severely affected by luminance levels than when stationary stimuli are fixated (appearance). According to such accounts, the larger number of unusually short and long latency saccades seen in Labs C and D are artefacts of noisier tracking data. However, conclusive discussion of this unexpected result is not possible given the current data. Regarding our estimates of minimum likely infant SRT, only the lower 5th percentile for deflection stimuli was more than trivially influenced by the

greater variability in data from Labs C and D. This therefore has few implications for our conclusions regarding minimum likely infant SRT.

Development and inter-lab differences

In early infancy, SRT reduces rapidly with age. Canfield et al. (1997) observed that between 2 and 12 months, appearance SRT reduced by a mean of 16 ms per month, and Gredebäck et al. (2006) observed that between 4 and 8 months, deflection SRT reduced by a mean of 38 ms per month. In contrast, for the current sample, the equivalent reductions were 3 and 10 ms respectively, with the previously reported values for younger infants outside the current 95% confidence intervals. Although this much slower development in late infancy contrasts with early infancy, it is consistent with development in older children. There is a paucity of relevant studies of children between the ages of one and four years, but one recent study of this age range examining SRT in response to appearing stimuli demonstrated a reduction of 2.4 ms per month (Alahyane et al., 2016). Indeed, reduction of SRT continues (but continues to slow) into adolescence, indicating that development relates to general and protracted brain development such as axon myelination (Luna et al., 2008). The current results are therefore compatible with existing data, but by assessing the little investigated age range of late infancy, we demonstrate that the slowdown in SRT development begins already towards the end of the first year of life.

The reason why Lab D did not follow the same developmental trend for deflection stimuli is likely to be the same reason that it differed in other respects particular to deflection – the larger display area meant the stimuli moved faster. If the developmental curve in relation to faster stimuli is more protracted, it must also be flatter. The reason why Lab C showed a different development pattern for appearance stimuli is a mystery, and we suggest that type I error is plausible.

Conclusion

This study had two main aims. Firstly, we set out to quantify typical infant SRTs under a range of spatiotemporal conditions. Secondly, we aimed to examine whether the effects would replicate across four different infant labs. We found that the commonly used cut-off value of 200 ms SRT (with shorter latencies regarded as predictive) is probably unnecessarily conservative. Mean and median SRTs in the conditions with shortest SRT were around 250 ms, so under most conditions samples on average faster than this are likely to contain predictive saccades. However, the variation was large: roughly a quarter of reactive saccades in the appearance condition are expected to be faster than 200 ms.

We demonstrated that several spatiotemporal factors (appearing stimuli versus deflecting stimuli, and distance of appearance) have strong effects on infant SRT which could be of practical significance in any of the many studies using infant SRT as a dependant measure. We demonstrated a number of interesting unexpected effects (gender differences in response to appearing stimuli and the effects of upwards deflection versus other directions for moving objects) which deserve further study. Finally, we demonstrated that the detected effects generally could be replicated across labs, but also that comparatively strong inter-lab differences can easily be created by unintended minor differences in procedure implementation. Replication across labs produced some unexpected differences, some of which (such as different levels of between-individual variation) were difficult to explain. This "messy" aspect of our data highlights the reality of the context dependence of data collection in a way which cannot be fully addressed by single-lab studies and thus is frequently ignored. These results highlight the importance for infant psychology of continuing to increase the number of replication studies.

References

675 Alahyane, N., Lemoine- Lardennois, C., Tailhefer, C., Collins, T., Fagard, J., & Doré-Mazars, K. (2016). Development and learning of saccadic eye movements in 7- to 42-month-old children. *Journal of Vision, 16*(1), 6-6. doi:10.1167/16.1.6

Alexander, G. M., & Wilcox, T. (2012). Sex differences in early infancy. *Child Development Perspectives, 6*(4), 400-406. doi:10.1111/j.1750-8606.2012.00247.x

680 Ando, S., Kida, N., & Oda, S. (2001). Central and peripheral visual reaction time of soccer players and nonathletes. *Perceptual and Motor Skills, 92*(3), 786-794. doi:10.2466/pms.2001.92.3.786

Atkinson, J. (2000). *The developing visual brain*. Oxford: Oxford University Press.

Bartoń, K. (2015). MuMIn: Multi-model inference (Version 1.15.1). Retrieved from

685 http://cran.r-project.org/package=MuMIn

Benson, J. B., Cherny, S., Haith, M. M., & Fulker, D. W. (1993). Rapid assessment of infant predictors of adult IQ: Midtwin-midparent analyses. *Developmental Psychology, 29*(3), 434-447.

Bremner, J. G., Slater, A. M., & Johnson, S. P. (2015). Perception of object persistence: The

690 origins of object permanence in infancy. *Child Development Perspectives, 9*(1), 7-13. doi:10.1111/cdep.12098

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development,*

695 *62*(2), 1-145.

Canfield, R. L., Wilken, J., Schmerl, L., & Smith, E. G. (1995). Age-related change and stability of individual differences in infant saccade reaction time. *Infant Behavior and Development, 18*(3), 351-358. doi:10.1016/0163-6383(95)90023-3

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular co-ordination of human horizontal saccadic eye movements. *The Journal of Physiology, 404*, 157-182.

Conover, W. J. (1999). *Practical Nonparametric Statistics (3rd edition)*: Wiley.

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Sex differences in reaction time mean and intraindividual variability across the life span. *Developmental Psychology, 48*(5), 1262-1276. doi:10.1037/a0027550

Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., . . . Piven, J. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *American Journal of Psychiatry, 170*(8), 899-908. doi:10.1176/appi.ajp.2012.12091150

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry, 74*(3), 189-194. doi:10.1016/j.biopsych.2012.11.030

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., . . . Yurovsky, D. (under review). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. Retrieved from https://osf.io/preprints/psyarxiv/27b43/

Garbutt, S., Harwood, M. R., & Harris, C. M. (2006). Infant saccades are not slow. *Developmental Medicine and Child Neurology, 48*(8), 662-667. doi:10.1017/s0012162206001393

Gilmore, R. O., Raudies, F., & Jayaraman, S. (2015). What accounts for developmental shifts in optic flow sensitivity? *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 19-25. doi:10.1109/DEVLRN.2015.7345450

Gredebäck, G., & Falck-Ytter, T. (2015). Eye movements during action observation. *Perspectives on Psychological Science, 10*(5), 591-598. doi:10.1177/1745691615589103

Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology, 35*(1), 1-19. doi:10.1080/87565640903325758

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition, 114*(2), 197-206.

Gredebäck, G., Örnkloo, H., & von Hofsten, C. (2006). The development of reactive saccade latencies. *Experimental Brain Research, 173*(1), 159-164. doi:10.1007/s00221-006-0376-z

Gredebäck, G., Stasiewicz, D., Falck-Ytter, T., von Hofsten, C., & Rosander, K. (2009). Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Developmental Psychology, 45*(4), 1190-1194. doi:10.1037/a0015667

Gredebäck, G., & von Hofsten, C. (2007). Taking an action perspective on infant's object representations. In C. v. Hofsten & K. Rosander (Eds.), *Progress in Brain Research* (Vol. 164, pp. 265-282): Elsevier.

Grönqvist, H., Gredebäck, G., & Hofsten, C. v. (2006). Developmental asymmetries between horizontal and vertical tracking. *Vision Research, 46*(11), 1754-1761. doi:10.1016/j.visres.2005.11.007

Haines, R. F. (1975). Peripheral visual response time and retinal luminance-area relations. *Optometry & Vision Science, 52*(2), 85-95.

Haith, M. M., & McCarty, M. E. (1990). Stability of visual expectations at 3.0 months of age. *Developmental Psychology, 26*(1), 68-74.

Hunnius, S., & Geuze, R. H. (2004). Gaze shifting in infancy: a longitudinal study using dynamic faces and abstract stimuli. *Infant Behavior and Development, 27*(3), 397-416. doi:10.1016/j.infbeh.2004.02.003

Jacobson, S. W., Jacobson, J. L., O'Neill, J. M., Padgett, R. J., Frankowski, J. J., & Bihun, J. T. (1992). Visual expectation and dimensions of infant information-processing. *Child Development, 63*(3), 711-724.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1994). Facilitation of saccades toward a covertly attended location in early infancy. *Psychological Science, 5*(2), 90-93. doi:10.1111/j.1467-9280.1994.tb00636.x

Johnson, S. P. (2003). The nature of cognitive development. *Trends in Cognitive Sciences, 7*(3), 102-104. doi:10.1016/S1364-6613(03)00030-5

Kenward, B. (2010). 10-month-olds visually anticipate an outcome contingent on their own action. *Infancy, 15*(4), 337-361. doi:10.1111/j.1532-7078.2009.00018.x

Kenward, B., Koch, F., Brehm, J., Forssman, L., Hermansen, T. K., Marciszko, C., . . . Gredebäck, G. (in prep.). Infants' sub-300 ms recognition of emotional faces in the visual periphery.

Kochukhova, O., & Gredebäck, G. (2007). Learning about occlusion: Initial assumptions and rapid adjustments. *Cognition, 105*(1), 26-46.

Leppänen, J. M., Forssman, L., Kaatiala, J., Yrttiaho, S., & Wass, S. (2015). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods, 47*(2), 538-548. doi:10.3758/s13428-014-0473-z

Liversedge, S., Gilchrist, I., & Everling, S. (Eds.). (2011). *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press.

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition, 68*(3), 293-308. doi:10.1016/j.bandc.2008.08.019

Malcolm, C., McCulloch, D., & Shepherd, A. (2002). Pattern-reversal visual evoked potentials in infants: gender differences during early visual maturation. *Developmental Medicine & Child Neurology, 44*(05), 345-351. doi:10.1017/S0012162201002183

Murray, I. J., Parry, N. R., McKeefry, D. J., & Panorgias, A. (2012). Sex-related differences in peripheral human color vision: a color matching study. *Journal of Vision, 12*(1). doi:10.1167/12.1.18

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142. doi:10.1111/j.2041-210x.2012.00261.x

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). doi:10.1126/science.aac4716

Peltola, M. J., Forssman, L., Puura, K., van IJzendoorn, M. H., & Leppänen, J. M. (2015). Attention to faces expressing negative emotion at 7 months predicts attachment security at 14 months. *Child Development*. doi:doi: 10.1111/cdev.12380

Peltola, M. J., Leppanen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Developmental Science, 11*(1), 60-68. doi:10.1111/j.1467-7687.2007.00659.x

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reznick, J. S., Chawarska, K., & Betts, S. (2000). The development of visual expectations in the first year. *Child Development, 71*(5), 1191-1204.

Richards, J. E., & Holley, F. B. (1999). Infant attention and the development of smooth pursuit tracking. *Developmental Psychology, 35*(3), 856-867.

Rosander, K. (2007). Visual tracking and its relationship to cortical development. In C. v. Hofsten & K. Rosander (Eds.), *Progress in Brain Research* (Vol. Volume 164, pp. 105-122): Elsevier.

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Caro, D. M. (2002). A longitudinal study of visual expectation and reaction time in the first year of life. *Child Development, 73*(1), 47-61.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*: Wiley.

Slater-Hammel, A. T. (1955). Reaction time to light stimuli in the peripheral visual field. *Research Quarterly. American Association for Health, Physical Education and Recreation, 26*(1), 82-87. doi:10.1080/10671188.1955.10612805

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89-96. doi:10.1111/j.1467-7687.2007.00569.x

Stephen, J. M., Aine, C. J., Christner, R. F., Ranken, D., Huang, M., & Best, E. (2002). Central versus peripheral visual field stimulation results in timing differences in dorsal stream sources as measured with MEG. *Vision Research, 42*(28), 3059-3074. doi:10.1016/S0042-6989(02)00415-7

Tobii Technology AB. (2013). Accuracy and precision test report: TX300 fw 1.1.1 RC Bright Light Illumination Mode. Retrieved from http://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-tx300-eye-tracker-fw-1.1.1-accuracy-and-precision-test-report.pdf

Wass, S. V., Forssman, L., & Leppänen, J. M. (2014). Robustness and Precision: How Data Quality May Influence Key Dependent Variables in Infant Eye-Tracker Analyses. *Infancy, 19*(5), 427-460. doi:10.1111/infa.12055

Wass, S. V., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology, 21*(18), 1543-1547. doi:10.1016/j.cub.2011.08.004

Wass, S. V., & Smith, T. J. (2014). Individual Differences in Infant Oculomotor Behavior During the Viewing of Complex Naturalistic Scenes. *Infancy, 19*(4), 352-384. doi:10.1111/infa.12049

Figures

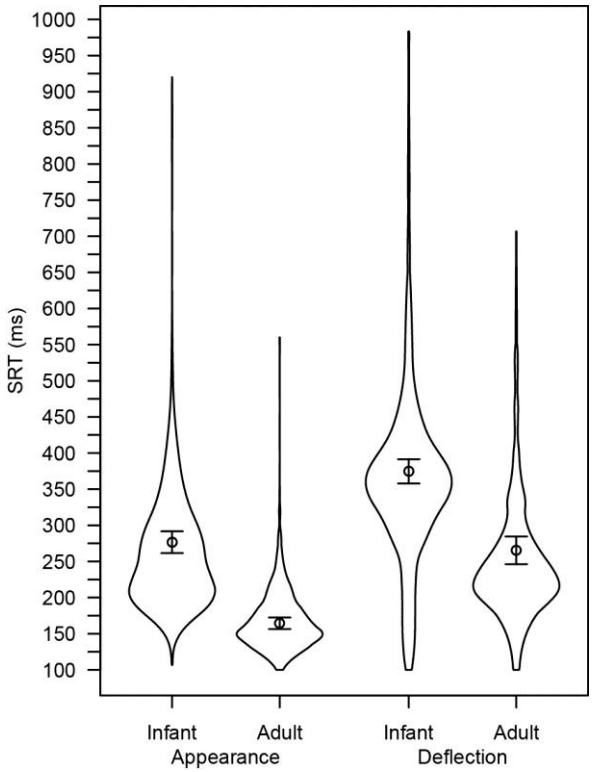


Figure 1. Infant and adult SRTs for appearance and deflection stimuli. Density plots of all saccades and 95% CIs for individual means.

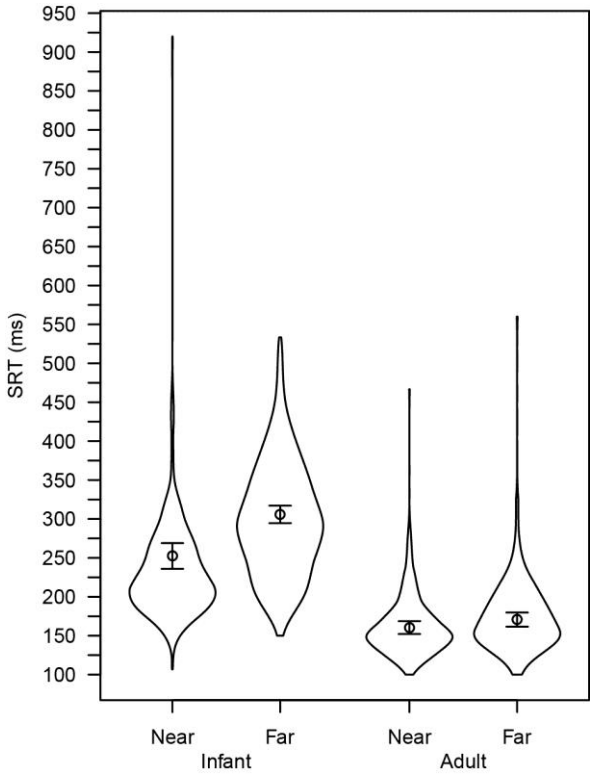


Figure 2. Infant and adult SRTs for near and far appearance stimuli. Density plots of all saccades and 95% CIs for individual means.

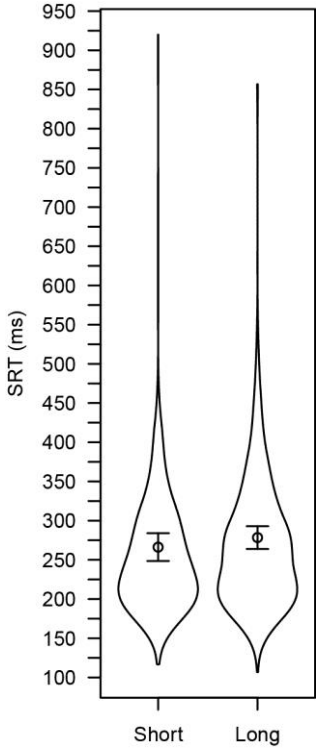


Figure 3. Infant SRTs for short and long delayed appearance stimuli. Density plots of all saccades and 95% CIs for individual means.

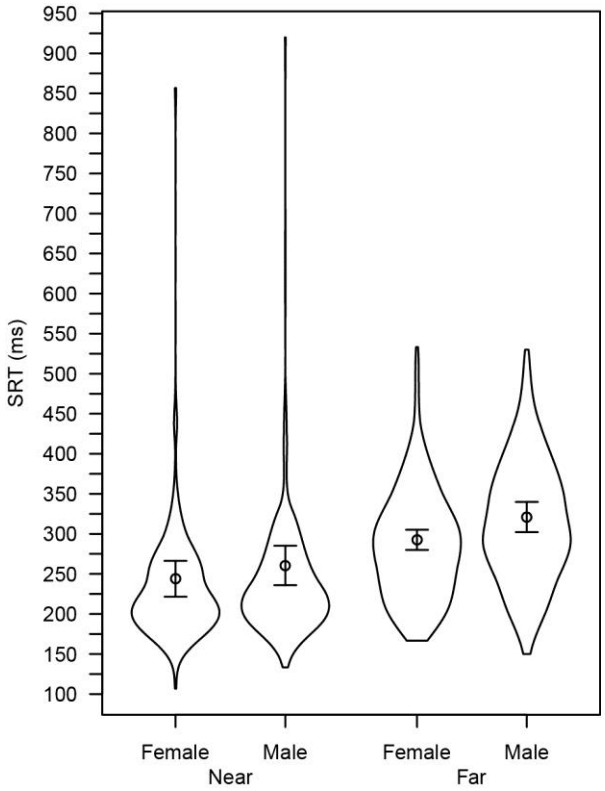


Figure 4. Infant SRTs for near and far appearance stimuli, by infant gender. Density plots of all saccades and 95% CIs for individual means.

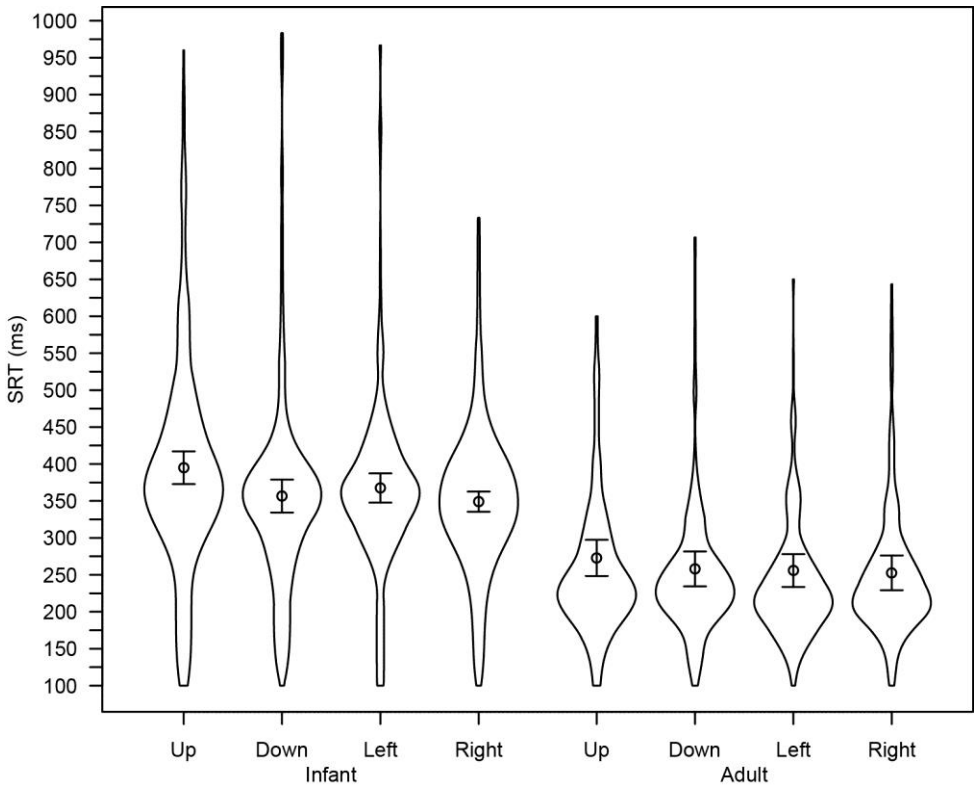


Figure 5. Infant and adult SRTs for deflections in different directions. Density plots of all saccades and 95% CIs for individual means.

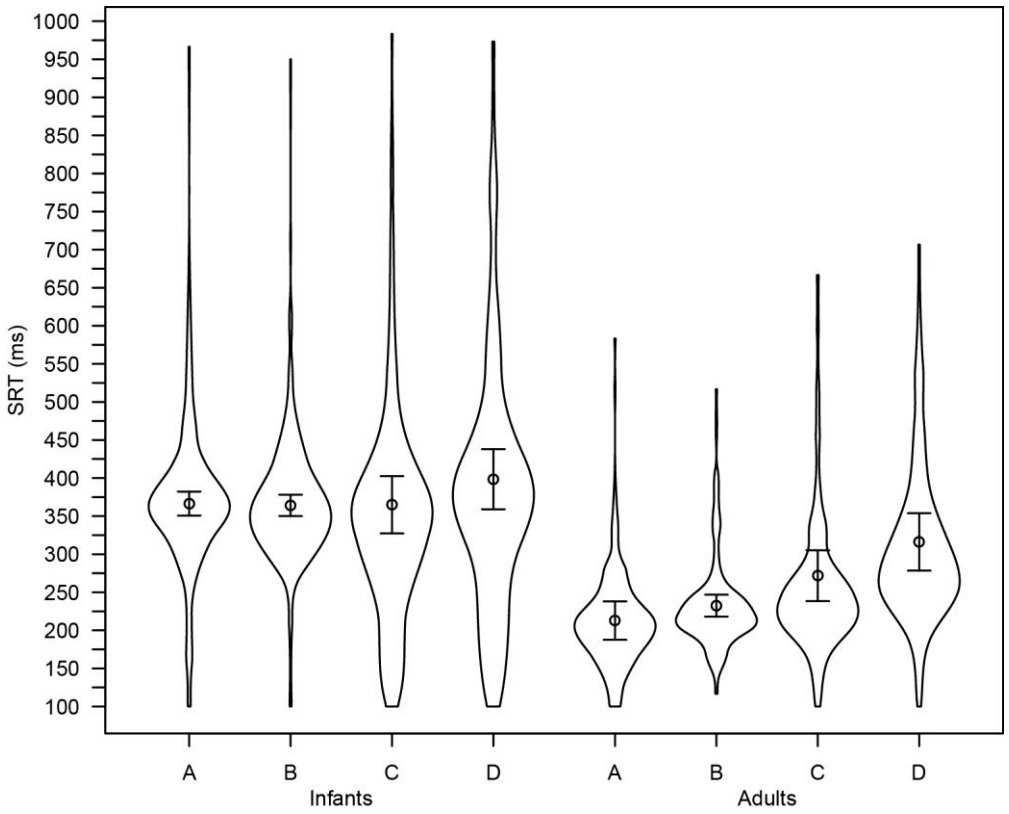
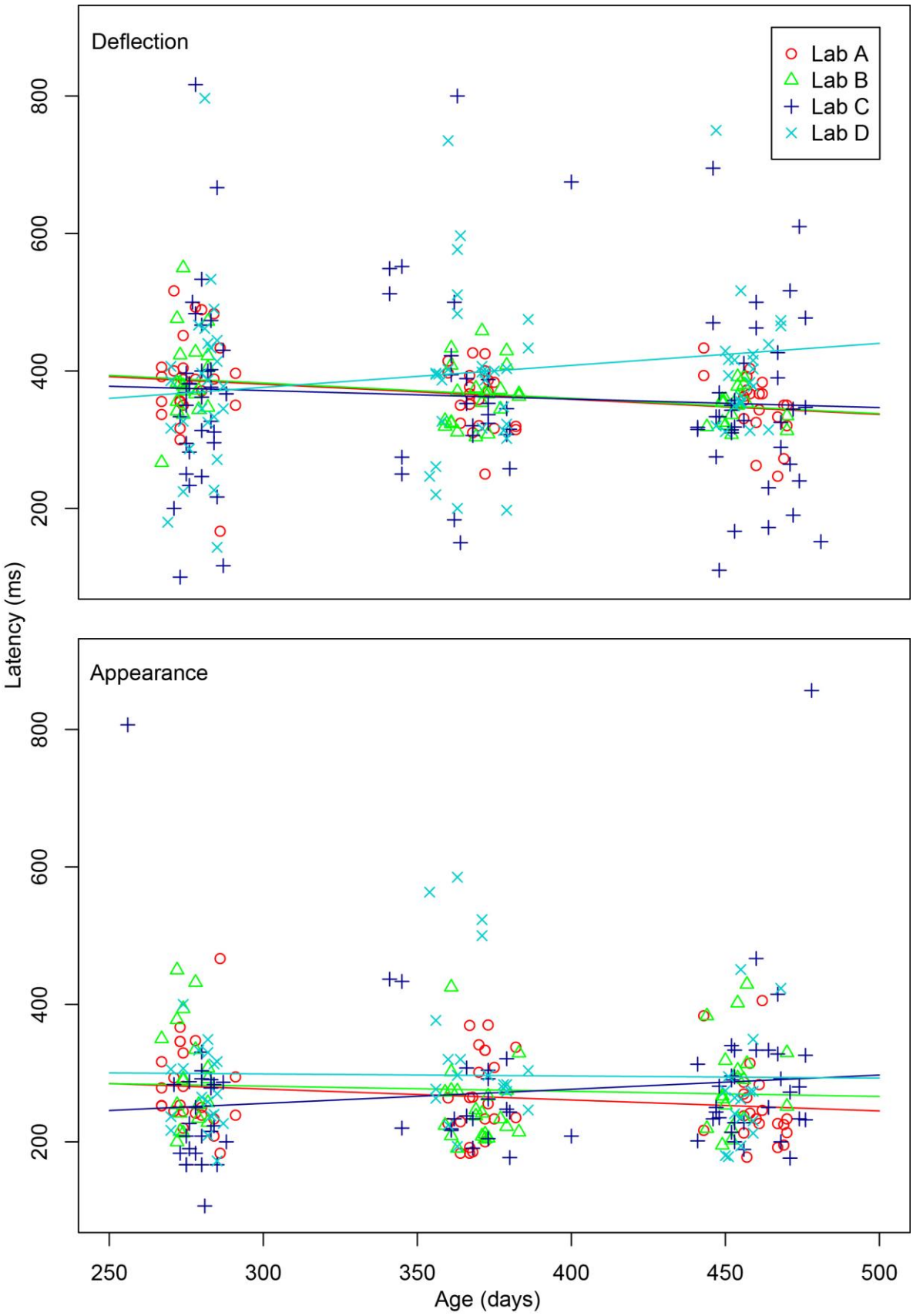


Figure 6. Infant and adult SRTs for deflection stimuli, by lab. Density plots of all saccades and 95% CIs for individual means.



845 Figure 7. Infant developmental trends for deflection and appearance stimuli, by lab. Data points are individual mean values. Linear regressions of these values are plotted for each lab.