Language, Cognition and Neuroscience, 2014
Vol. 29, No. 4, 424–434, http://dx.doi.org/10.1080/01690965.2013.791702

Routledge
Taylor & Francis Group

Shared predictive decision-making mechanisms in action and language

Edita Poljac[a,b]*, Kristoffer Dahlslätt[b] and Harold Bekkering[b]

[a]*Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom;* [b]*Department of Cognitive Psychology, Radboud University Nijmegen, Nijmegen, The Netherlands*

(*Received 31 May 2012; final version received 25 March 2013*)

Anticipatory eye movements are reported in studies on motor control as well as on language comprehension, implying that this major orienting system is involved in generating goal-directed behaviour within the action and the language domain. The cognitive contribution of these anticipatory eye movements to language and motor control, however, is still not well understood. This study investigated whether anticipatory eye movements reflect the working of a predictive mechanism that is shared between action and language and if so, whether the predictions are based primarily on an anticipation of the next discrete event (movement or word), or rather represent a semantic understanding of the end goal of the whole event (action or sentence). To this end, we designed two highly comparable paradigms with complex action sequences – one relying more strongly on the action and the other on the language system. The data demonstrated a pattern of predictive looks in our action observation paradigm that was similar to that observed in the visual world paradigm. These findings provide empirical evidence for the idea of a shared predictive mechanism that allows for fluent behaviour in action and language. Moreover, the pattern in both paradigms was such that it demonstrated an increase in predictive looks in the final action step. This finding implies that the predictive mechanism accumulates semantic information relevant for our overall (motor or linguistic) behavioural goals, rather than just predicting discrete events when making decisions about complex action sequences. Such a predictive mechanism facilitates understanding of complex situations, allowing for efficient and adaptive interaction with our environment.

Keywords: action observation; language comprehension; anticipatory eye movements and predictions; semantics

People interact with the world motivated and guided by their internal goals, demonstrating a preference for goal-directed behaviour from birth on (Craighero, Leo, Umiltà, & Simion, 2011). In a rapidly changing environment, the success of this goal-directed interaction with the physical world strongly depends on our ability to act in advance to environmental demands. From an evolutionary perspective, such a predictive mechanism would be beneficial for our survival, as it would allow us to optimise our goal-directed behaviour and to be adaptive (Butz, Sigaud, & Pezzulo, 2007; Imamizu & Kawato, 2009). Being able to optimise our behaviour has recently been suggested to be one of the core mechanisms of our brain (Friston, 2010). Understanding predictive processing in human goal-directed behaviour becomes hence essential if we are to understand the principles of human cognition. The current study aims to investigate how predictions are generated in human goal-directed behaviour within the action and the language system.

Anticipatory behaviour has often been studied experimentally by registering eye movements during various cognitive tasks. The idea behind using this method is that our internal goals guide our eyes to move and actively gather the sensory information we need to complete a given task. We know already that the human visual system involves more than just passive reception of sensory information entering our neurocognitive system in a bottom-up fashion (cf. Cavanagh, 2011; Tatler, Hayhoe, Land, & Ballard, 2011). In fact, studies within the research field of action (e.g. Flanagan & Johansson, 2003) and language (e.g. Altmann & Kamide, 1999) have shown that depending on the task and the provided context, our eyes move in anticipation of upcoming task-relevant information.

Different studies of sensorimotor control have demonstrated anticipatory eye movements in tasks involving simple actions. For instance, while performing actions involving object manipulation, participants' eye movements typically arrive at the target before the arm (e.g. Ballard, Hayhoe, & Pelz, 1995; Epelboim et al., 1995; Johansson, Westling, Backstrom, & Flanagan, 2001; Land, Mennie, & Rusted, 1999; Sailer, Flanagan, & Johansson, 2005). It is important to note here that although we know from different studies that

*Corresponding author. E-mail: e.poljac@donders.ru.nl

this action-related target selection is strongly coupled to visual attention (e.g. Deubel & Schneider, 1996), perceptual processing behind anticipatory eye movements seems to be guided by our conceptual expectations about action goals (cf. Ondobaka & Bekkering, 2012). Interestingly, when observing others performing actions, similar eye-movement patterns are elicited to those when people execute actions themselves both in adults (Flanagan & Johansson, 2003; Rotman, Troje, Johansson, & Flanagan, 2006) and in infants (Rosander & von Hofsten, 2011; Van Elk, van Schie, Hunnius, Vesper, & Bekkering, 2008). These findings suggest that our visual system gathers accurate task-specific information available in the environment and makes predictions about upcoming actions, both of which allow us to exhibit fluent motor control during goal-directed action execution and action observation (cf. Kawato, 1999; Miall & Wolpert, 1996).

Similar to these findings in action control, studies of language comprehension have provided ample evidence for the idea that anticipatory eye movements reflect dynamical updating of mental representations that people construct while processing provided linguistic and visual information over time (e.g. Altmann & Kamide, 1999, 2009; Knoeferle & Crocker, 2006, 2007; Knoeferle, Crocker, Scheepers, & Pickering, 2005). In these studies, participants would typically listen to spoken sentences while looking at displays showing common objects, some of which are referred to in the spoken text. The participants are instructed to look anywhere they want, while their eye movements are simultaneously being recorded (e.g. Cooper, 1974, Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The findings of this paradigm – known as the visual world paradigm – would usually demonstrate anticipatory eye movements towards the objects that are mentioned or are in some way associated with the spoken text (for a recent review on this paradigm, see Huettig, Rommers, & Meyer [2011]). It seems hence that anticipatory eye movements as detected when testing language comprehension reflect an interaction between the linguistic and the visual input being processed and utilised in a proactive way.

These recent studies in action control and psycholinguistics strongly indicate that people anticipate visual information that is relevant for their task goals. Considering that feedback processing is slow in general (Elliott et al., 2010; Keele & Posner, 1968), actions and language would clearly lack their fluency without such a predictive system. Interestingly, a comparison between these two research fields seems also to suggest that people proactively combine the task-relevant visual stream with the motor information (Flanagan & Johansson, 2003) in a similar way as they do with the linguistic information available (Huettig et al., 2011). It seems hence possible that our brain relies on a predictive mechanism that is employed when our behaviour is guided by our motor, but also when it is guided by our linguistic goals. The purpose of such a mechanism would be to make semantic predictions over time about upcoming behavioural goals, helping us to understand a given situation (cf. Kilner, Friston, & Frith, 2007). Our study was therefore developed to test the idea that the anticipatory eye movements as observed in action observation and language comprehension reflect the working of a shared predictive mechanism that allows us to interact efficiently with the surrounding world.

To test this idea of a shared predictive mechanism in action and language that allows us to make semantic predictions of actions over time, we applied two paradigms – one relying more strongly on the action and the other on the language system. Both were a modified version of the paradigms already existing in the literature, but were designed to be maximally comparable, allowing for the most direct achievable contrast of anticipatory eye movements observed in the action and in the language domain. Accordingly, we operationalised anticipatory eye movements in the same way in both paradigms, with predictions being enabled though object-oriented movements (action) and verbs (language). Our dependent measures were more elaborate than those typically used to investigate the association strength between language and perception. Usually, in the visual world paradigm, a region of interest would be specified for the visual input – mostly depicting semi-realistic scenes including different cartoon-like objects – based on its relation to the presented sentences (e.g. Altmann & Kamide, 1999; Huettig et al., 2011). The speed of the initiation of a saccadic eye movement to the region of interest (e.g. cake) would then be registered, while the participant is hearing the utterance. Faster saccade onsets are typically demonstrated when comparing the saccade onset for sentences including the critical verb (The boy will eat . . .) with those including a control verb (The boy will take . . .). In our study, however, we aimed to move beyond the mere perception-language associations, and measured therefore the dynamics of saccadic eye-movement patterns during a sequence of actions or words presented. In both paradigms, we used action sequences consisting of three steps, with the last step being the final goal of the presented sequence. For each of the action steps, we used the onset of the hand movement or the onset of the verb related to the object being manipulated during that action step to specify whether the saccade towards the region of interest was predictive (i.e. occurred before the actual – motor or verbal – action took place) or reactive (i.e. occurred during or after the actual action). We then specified for

each of the action steps of the action sequence whether the corresponding eye movement originated from the predictive or the reactive phase and defined the specific eye movement as either predictive or reactive accordingly. Since the eye movements from the reactive phase cannot include any predictions related to that specific action step, the average of its reactive eye movements served in our study as a baseline when specifying its corresponding anticipatory value. Accordingly, the predictive eye movements (calculated as looking times, gaze onsets and count ratios; see Methods for details) were defined as a relative measure comparing the events in which the eye movements came from the predictive phase with those events in which the eye movements came from the reactive phase and subsequently determining if this relative amount was larger than zero for each of the action steps. We furthermore compared the amount of predictive eye movements between the action steps.

In addition, the two paradigms included meaningful actions and dynamical stimulus material (auditory stream and video material; cf. Carmi & Itti, 2006). This was introduced to reflect the real-life situations as much as possible. Importantly, applying highly comparable stimulus material in our action observation paradigm (tackling action-related processes) and in our visual world paradigm (tackling language-related processes) was in line with our aim to test the assumption of a shared predictive mechanism for the action and the language domain. This assumption predicts that the patterns of anticipatory eye movements captured with the action observation paradigm should be similar to those captured with the visual world paradigm. Moreover, we aimed to test whether the predictions in a complex action sequence with a clear end outcome are based primarily on an anticipation of the next discrete event (movement or word), or that predictions serve a semantic understanding of the whole event (action or sentence). If the eyes just precede the upcoming discrete event (movement or word), we expected predictive looks to demonstrate a constant pattern across the elements of the whole event. However, if knowledge is integrated over the whole event (action or sentence) oriented towards the goal of the event, we expected predictive looks to increase as the action unfolds.

Method

Participants

A total of 34 participants, all native Dutch speakers, were recruited from the participant pool of the Radboud University Nijmegen. Seven were excluded from further analyses due to loss of eye-tracking data exceeding 60%. Out of the remaining 27 participants, 19 were females and 8 were males (mean age = 23.2). Participants received course credits or monetary compensation for taking part in the study.

Stimuli and tasks

Action observation paradigm. The stimuli consisted of 17 movies depicting multiple steps of object-related actions, in which each upcoming step of the action was dependent on the previous one. In addition, five catch movies were used, in which the final step (i.e. the outcome) of the whole action sequence was incompatible with the intent of the actor. All action sequences used in the movies consisted of three object-related action steps. Participants were required to watch the presented videos and to indicate by a button press when they thought that the intended outcome of the actor failed, as in the case of the catch movies. In all movies, the head of the actor was visible, whereas the eyes were covered by the brim of a hat eliminating intentional cues from actor's eye movements. The used movies had a set width of 1280 pixels (up-scaled from 720 p), with varying height. They all started with a 2-second freeze-frame containing no movement cues, and lasted for a duration ranging from 11 to 24 seconds. Direction of the first action step was counterbalanced (9 right, 8 left).

Visual world paradigm. Twenty-one images taken from the initial frames of all movies – five of which corresponded to the catch movies – were all paired with the corresponding spoken sentences, such that the actions described in the sentences matched the actions performed by the actor in the movies. In this way, comparable to the action observation paradigm, all action sequences used here also consisted of three object-related action steps. Figure 1 presents an example image and its corresponding sentence. The sentences were recorded by a female native speaker of Dutch. Comparable to the action observation paradigm, each trial started with displaying the image for 2 seconds, after which period the auditory stream started and lasted for a duration ranging from 5 to 11 seconds.

Procedure

During the experiment, the participant was seated approximately 60 cm from a Tobii 1750 eyetracker (Tobii 1750, Tobii Technology AB, Stockholm, Sweden). The presentation of stimuli and the recording of eye movements were controlled by Presentation 13.1 (Neurobehavioural Systems Inc., Albany, CA, USA). Videos and images were displayed on the 17'' integrated monitor of the eyetracker, using a resolution of 1280 × 1024, and sounds were played through a stereo speaker setup, while eye movements were sampled at 50 Hz. A

nine-point calibration procedure was used prior to recording, which started after successful calibration only. On each trial, the participant was first presented with a centred crosshair for 500 ms, followed by either a video (action observation paradigm), or a picture and sound combination (visual world paradigm), after which duration a blank screen of 2 seconds occurred. Stimuli pertaining to one of the two paradigms were presented in a blocked design. Each experimental block contained 17 experimental and 5 catch trials presented in a random order and with no stimulus repetition within blocks. Participants completed a total of four blocks – two blocks for each paradigm. For each participant, paradigm presentation varied between blocks according to an ABBA order, with the paradigm presented in the first block being counterbalanced across participants. Between each block there was a pause, and the experiment proceeded when the participant reported being ready to continue. Collection of the eye-movement data took approximately 25 minutes per participant.

Eye-movement data analysis

To prepare the eye-movement data collected in both paradigms for further analyses of their predictive values, we first defined square-shaped areas of interest (AOI). This was done in all movies and images separately by determining the AOIs based on the objects that were manipulated during a specific action sequence. Since each action sequence consisted of three object-related action steps, each individual movie or image included three AOIs in total (see Figure 1 for an example). Furthermore, each of these three AOIs related to just one of the three object-related action steps and was defined closely around the corresponding object. After having defined the AOIs, we then determined their corresponding time windows. Just as each AOI was related to the object being manipulated during the corresponding action step, its corresponding time window was tightly related to the movement of the hand that manipulated the relevant object (video) or the related verb (sentence). Importantly, within each of the time windows, we furthermore specified intervals we considered to relate to either predictive or reactive phase of an action step. This was done in a similar way in both paradigms used in our study. Specifically, in each time window, we first specified three time points: a starting, a middle, and an ending point. In the action observation paradigm, the starting points for all time windows were marked by the onset of the hand movement towards the upcoming (object related) target AOI in the action sequence, while middle points were created when the hand (and/or held object) entered the upcoming AOI. The ending points were usually created 1000 ms after the middle point. In the case where the time between the starting and the middle points was less than 1000 ms, however, the time windows were created symmetrically around the middle point, having same distance from middle point to both the ending and starting points. In the visual world paradigm, similar time windows were created, with the starting and the middle points being determined by the onset of the verb and the noun representing the upcoming target, respectively.

Critically, for both paradigms holds that, if the gaze entered an AOI before the middle point of its corresponding time window (i.e. before reaching the corresponding target object), it was considered a *predictive* eye movement, whereas if the gaze arrived after the middle point (i.e. after the corresponding object has been reached), it was considered a *reactive*

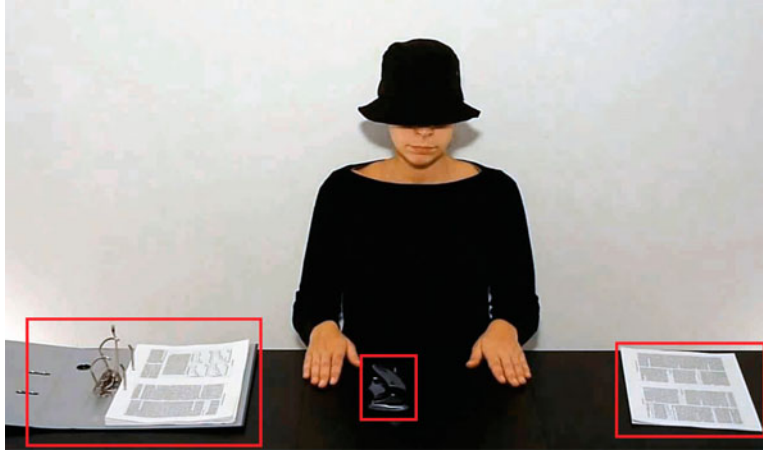

Figure 1. An example image used in the visual world paradigm. The accompanying sentence for this image was: 'Het meisje **pakt** netjes een hele stapel *papier* en **schuift** deze in de *perforator*. Vervolgens **maakt** ze er gaatjes in en **doet** het in de *map*' (verbs are marked in bold and targets are italicised). 'The girl **takes** carefully a bunch of *paper* and **slides** it in the *hole-puncher*. Then she **punches** holes in it and **puts** it in the *ring-binder*'. The squares in the image indicate the regions of interests used for the eye-movement analyses.

eye movement. Since the middle point within a time window was the moment at which the hand reaches the AOI of the target object of that action step, we considered this point in time as the one separating the predictive from the reactive eye movements.

To further illustrate how the three time points were determined for each of the time windows, let us consider the example given in Figure 1. Also here three AOIs were first defined – depicted as red squares in the figure – corresponding to each of the three individual action steps the whole action sequence consists of. Time windows were then created for all AOIs, with the three time points being subsequently determined for each of the time windows as follows: The first point of the first time window starts directly after the 2 seconds of freeze-frame each video started with, as this is when the actor's hand leaves the table (2000 ms), with the middle point occurring when the hand enters the AOI of the paper (2700 ms) and ending at 3400 ms. The second time window starts at the onset of the paper being moved towards the puncher (6580 ms), with the middle point occurring when the paper and the hands enter the AOI demarcating the hole-puncher (8180 ms) and ending at 9180 ms. The third time window starts when the paper is moved from the hole-puncher (12,500 ms), with the middle point starting when the paper and the hands enter the AOI of the ring-binder (13,900 ms) and ending at 14,900 ms.

After having classified the eye movements collected during individual action steps of each of the used action sequences (i.e. each trial) as either reflecting a predictive or a reactive look, we then applied this categorisation within three different measures of predictive looks in each of the paradigms – participants' looking times, gaze-onset times and counts of their looks. Looking times referred to the amount of time participants spent looking at an AOI within either the predictive or reactive portion of its corresponding time window. This amount of time was expressed as a percentage of the total duration of either predictive or reactive part of that time window, dependent on whether the eye movement related to the corresponding action step was classified as predictive or reactive. Furthermore, participants' gaze-onset times referred to the latency of the first gaze entering the corresponding AOI during a time window. Finally, participants' count of predictive and reactive looks for all action steps across trials were specified by calculating the ratios between the number of predictive and reactive looks, rather than their absolute values, aiming to account for possible differences in the amount of excluded trials between action steps. The trials that were excluded for all analyses were those in which gaze failed to enter a given AOI during the time window defined by the action directed towards that AOI. Analysing these three different measures of predictive looks allowed us to investigate different aspects of our eye-tracking data, which is in particular informative for the following two reasons: First, no real consensus has been reached in the literature so far on what would be the best measure of predictive looks. Second, although similar paradigms have been used before, this was to our knowledge the first study that used real-world images in the visual world paradigm and multi-step action sequences in action observation.

Importantly, all of our measures considered a relative score between predictive and reactive looks, based on the idea that if participants were really predicting the upcoming individual action steps, then the aggregate difference score (predictive minus reactive in looking times and gaze onsets) or their ratio (predictive divided by reactive in the count ratio) should significantly be larger than zero. On the contrary, if participants did not predict the upcoming action steps but rather reacted to the actor's movements or to the nouns in the sentences, then the opposite should be the case. The rationale for using this form of relative scores, with reactive looks being applied as a control measure, was that this allowed for an inherent form of control. In the case of the balance of predictive looks significantly deviating from zero as well as any significant differences between the individual action steps, we could be sure that any differences we detected were primarily generated by experimentally induced variance as intended, rather than by some randomly induced biases, such as for instance biases arising due to differences in object salience. This type of baseline is somewhat different from the typically applied baseline in studies investigating anticipatory eye movements: in the visual world paradigm, for instance, eye movements would typically be compared between conditions that involve disambiguation (i.e. the sentence involves the critical verb that is specifically related to one of the presented objects) and those without disambiguation (i.e. the sentence involves a neutral verb that can be related to most of the objects presented).

The three measures of predictive looks we applied were analysed separately using two-tailed one-sample *t*-tests ($\mu = 0$) to test whether participants were predicting upcoming action steps, and using repeated measures analyses of variance (ANOVAs), with paradigm (action observation/visual world) and action step (first/second/third) as within-subject factors, to investigate whether the patterns of predictive eye movements differed between the two paradigms applied in our study. For the purpose of these analyses, our data were normalised by arcsine transformation performed on proportional data of looking times and by square root transformation performed on count ratio data. Furthermore, for

the analyses of gaze onset and count ratio data, a linger correction was performed to ensure that the data were not derived from fleeting looks or random saccades. In order for a data point to be included in the analysis, gaze had to stay within the AOI for at least 100 ms after first entering it. Only data points that fulfilled this criterion were included in subsequent gaze onset and count ratio analyses. Next, for the action observation paradigm, in order to include gazes that exclusively resulted from participants observing the actor's actions, only looks originating from either the previous AOI or from the head region were included in subsequent analyses. These offset criteria were not necessary in the visual world paradigm, since the images used there were static and hence participants' gaze was solely guided by the sentences in the auditory stream.

Results

We first present data analyses of looking time, followed by gaze-onset data analyses before finally presenting the count ratio analyses. For each of these three measures of predictive eye movements, we first present (a) the analysis of individual predictive values for each of the three action steps that action sequences consisted of, followed by (b) the analysis of the patterns of these individual predictive values across the three action steps within each of the paradigms, and finally presenting (c) the analysis comparing these patterns between the action observation and the visual world paradigm.

Predictive looking time

Figure 2 (panel A) presents percentages of predictive looking times, which were calculated as aggregate difference scores (predictive minus reactive looking times) for each of the three action steps in both paradigms. In the action observation paradigm, predictive looking times were negative, that is, significantly smaller than zero in action step 1, $t(26) = -2.68$, $p = 0.013$, did not significantly differ from zero in action step 2, $t(26) = 1.60$, $p = 0.12$, and were positive, that is, larger than zero in action step 3 only, $t(26) = 2.38$, $p = 0.025$. In the visual world paradigm, predictive looking times for all action steps were significantly larger than zero, with step 1, $t(26) = 4.49$, $p < 0.001$; step 2, $t(26) = 5.24$, $p < 0.001$ and step 3, $t(26) = 11.10$, $p < 0.001$. A one-way ANOVA revealed a main effect of action step for both action observation, $F(2,26) = 9.84$, $p = 0.001$, and the visual world paradigm, $F(2,26) = 14.51$, $p < 0.001$. Further analyses on predictive looking times within the action observation paradigm demonstrated larger percentage of predictive looks in both step 2, $F(1,27) = 11.23$, $p = 0.002$, and in step 3, $F(1,27) = 15.12$, $p < 0.001$ compared to action step 1, whereas no significant difference was found between action steps 2 and 3, $F(1,27) = 1.25$, $p = 0.27$. For the visual world paradigm, while no significant difference was observed between action steps 1 and 2, $F(1,27) = 2.34$, $p = 0.14$, action step 3 demonstrated larger percentage of predictive looking times than either action step 1, $F(1,27) = 29.19$, $p < 0.001$ or action step 2, $F(1,27) = 18.81$, $p < 0.001$. The pattern of percentage of predictive looking times across the three action steps differed marginally between the two paradigms, $F(1,26) = 3.09$, $p = 0.06$. This marginal interaction was generated by difference between the paradigms detected when comparing percentages of predictive looking times in action steps 2 and 3, $F(1,27) = 6.31$, $p = 0.018$. Specifically, while significantly larger in action step 3 than in action step 2 in the visual world paradigm, predictive looking times were similar for these two action steps in the action observation paradigm.

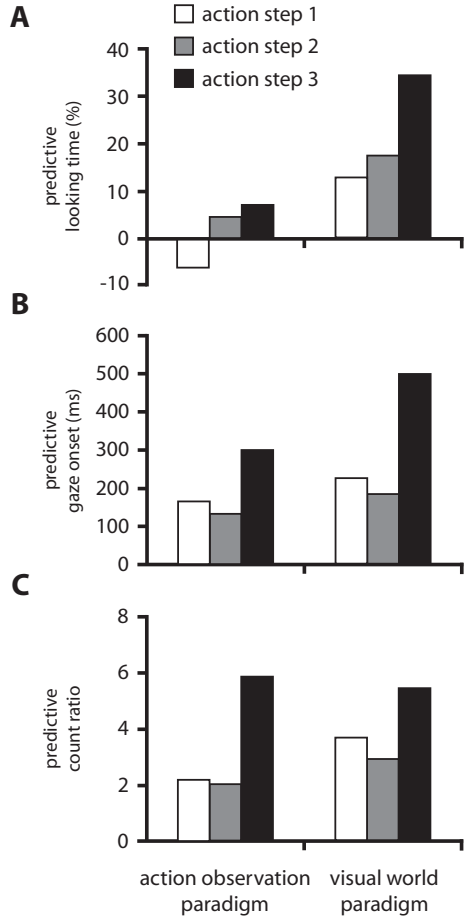

Figure 2. Three measures of anticipatory eye movements averaged across action sequences, which all consisted of three action steps. Panel A depicts percentage predictive looking times, panel B predictive gaze-onset latencies and panel C the predictive count ratio data. All of the panels depict data for the three action steps separately in both paradigms.

Predictive gaze onset

Predictive gaze-onset times – calculated as aggregate difference scores (predictive minus reactive gaze-onset times) – are presented for each action step for both paradigms in panel B of Figure 2. In the action observation paradigm, predictive gaze onsets for all action steps were significantly larger than zero, with $t(26) = 8.02$, $p < 0.001$; $t(26) = 6.06$, $p < 0.001$ and $t(26) = 10.39$, $p < 0.001$ for action steps 1, 2, and 3, respectively. Similarly, predictive gaze onsets for all action steps were also significantly larger than zero in the visual world paradigm, with $t(26) = 5.67$, $p < 0.001$; $t(26) = 6.61$, $p < 0.001$ and $t(26) = 8.00$, $p < 0.001$ for action steps 1, 2, and 3, respectively. A one-way ANOVA revealed a main effect of action step for both action observation, $F(2,26) = 22.45$, $p < 0.001$ and the visual world paradigm, $F(2,26) = 8.82$, $p < 0.005$. For action observation paradigm, the predictive gaze-onset times did not significantly differ between action steps 1 and 2, $F < 1$, whereas the gaze onset in action step 3 was significantly more predictive than either in action step 1, $F(1,27) = 25.41$, $p < 0.001$ or in action step 2, $F(1,27) = 43.34$, $p < 0.001$. A similar gaze-onset pattern was observed for the visual world paradigm, in which also no significant difference was observed between action steps 1 and 2, $F < 1$, whereas the gaze onset in action step 3 was significantly more predictive than either in action step 1, $F(1,27) = 15.17$, $p < 0.005$ or in action step 2, $F(1,27) = 18.10$, $p < 0.001$. Importantly, the pattern in predictive gaze onset across the three action steps did not significantly differ between the two paradigms, $F(1,27) = 1.68$, $p = 0.21$.

Predictive count ratio

The ratio between predictive and reactive looks for each action step in both paradigms is depicted in panel C of Figure 2. For action observation, the predictive count ratio was significantly larger than zero for in action steps, with $t(26) = 14.49$, $p < 0.001$; $t(26) = 13.54$, $p < 0.001$ and $t(26) = 10.92$, $p < 0.001$ for action steps 1, 2, and 3, respectively. In a similar way, predictive ratio scores were significantly larger than zero in all action steps also in the visual world paradigm, with $t(26) = 10.25$, $p < 0.001$; $t(26) = 11.58$, $p < 0.001$ and $t(26) = 10.09$, $p < 0.001$, for action steps 1, 2, and 3, respectively. A one-way ANOVA revealed a main effect of action step for action observation, $F(2,26) = 10.64$, $p < 0.001$, and a marginal effect in the visual world paradigm, $F(2,26) = 2.93$, $p = 0.07$. Further analyses of the data in the action observation paradigm revealed no significant difference in predictive count ratio between action steps 1 and 2, $F < 1$, whereas predictive count ratio was significantly higher in action step 3 than either in the step 1, $F(1,27) = 18.55$, $p < 0.001$ or in action step 2, $F(1,27) = 19.42$, $p < 0.001$. Also in the visual world paradigm, predictive count ratio did not significantly differ between action steps 1 and 2, $F < 1$, while it was again significantly higher in action step 3 than in action step 2, $F(1,27) = 5.41$, $p = 0.044$. In this paradigm, the difference in predictive count ratio between action steps 3 and 1 did not reach significance, $F(1,27) = 1.97$, $p = 0.17$. Importantly, the pattern in predictive count ratio across the three action steps did not significantly differ between the two paradigms, $F(2,26) = 1.10$, $p = 0.38$.

The visual world paradigm in our study was constructed such that participants could perform the task correctly without using the visual input – they would then simply rely on the auditory stream. Nevertheless, the participants viewed the scene by looking at the upcoming action targets, using the available visual information for predicting the events in the sentence. It is hence possible that this pattern of anticipatory eye movements in the visual world paradigm was reflecting some kind of a strategy that the participants developed over the course of the experiment (cf. Altmann & Kamide, 1999), possibly biased by the action observation paradigm. To further investigate this possibility, we tested whether the predictive looks in this paradigm were affected by action observation paradigm by testing how stable the pattern was across the experiment. For this purpose, we split the collected eye-movement data in four parts in the visual world paradigm, such that the four parts corresponded to the time progress along the course of the experiment. More specifically, the first and the second parts represented the first and the second half of the first experimental block and the third and the forth parts corresponded to the first and the second half of the second experimental block of the visual world paradigm. The analysis revealed that anticipatory eye movements were present in all measures of predictive looks present for all four different quarters of trials, averaged over the three action steps. Predictive looking times (12.38, 13.71, 18.92 and 15.15%; with $ts(26) > 5.49$, $ps < 0.001$); predictive gaze onsets (287, 216, 293 and 275 ms; with $ts(26) > 5.55$, $ps < 0.001$) and predictive counts ratios (0.14, 0.12, 0.16 and 0.14; $ts(26) > 4.67$, $ps < 0.001$) were observed in the first, second, third, and the forth-quarter of the visual world paradigm, respectively. This analysis demonstrates that anticipatory eye movements in the visual world paradigm were present from the start of the experiment and remained throughout, implying a stable predictive processing of information, rather than patterns generated by biases for the action observation paradigm.

Discussion

This study investigated whether anticipatory eye movements reflect the working of a shared predictive mechanism in action and language, which allows us to make semantic predictions of actions over time. Previous studies have already demonstrated that people move their eyes in anticipation of upcoming events that are relevant for their behavioural goals when tested with tasks that require action control or those assessing language comprehension. The present data demonstrate a pattern of anticipatory eye movements increasing over time when making decisions about complex action sequences: Different measures of predictive looks used in our study all demonstrated a peak in the final action step of action sequences. Critically, our data demonstrate a clear similarity in the pattern of anticipatory eye movements for action and language goals: Similar increase in predictive looks in the final action step was observed for both the action observation and the visual world paradigm. These findings seem to indicate that anticipatory eye movements reflect cognitive processes serving a shared predictive mechanism that allows for fluent behaviour within the action and the language system. The findings furthermore imply that anticipatory visuomotor behaviour depends on semantic processing of information available in the environment that is relevant to the overall (motor or linguistic) behavioural goal.

In general, our findings demonstrate that people exhibit similar patterns of anticipatory eye movements when making decisions based on action observation or on language comprehension. Specifically, in both of our paradigms, anticipatory eye movements were longer lasting, of quicker onset, and more frequent in the final goal of the action. The end outcome of the complex action sequences used in our study influenced the predictions of action goals made during action observation and language comprehension in a similar way. These similarities in anticipatory visuomotor behaviour imply a shared predictive mechanism that is employed to facilitate our behaviour guided by our motor or by our linguistic goals. Clarifying whether action and language utilise a single predictive cognitive and neural resource would, however, need further investigation, as the design used in our study does not allow us to say anything about this possibility. What we can conclude, though, is that both modalities rely on predictive mechanism, underpinned by either a single or by two separate prediction systems.

Moreover, findings seem to suggest that anticipatory eye movements actively gather and accumulate available semantic information relevant for our behavioural goals: All the measures we used to investigate predictive looks – looking times, gaze onsets and count ratio – demonstrated a significantly more prevalent anticipatory eye movements in the last and final step of the multi-step action sequences we employed. Given that each sequence consisted of three consecutive and mutually related action steps, we followed the pattern of anticipatory movements as the action sequence unfolded and observed that participants anticipated its final step much stronger than they anticipated the preceding action steps. This finding is important as it suggests that people not only anticipate discrete events in multi-step action sequences but also actively accumulate and combine the task-relevant visual information present in the external world while planning and executing their overall behavioural goal. Different studies on action control have shown already that when manipulating objects in multiple steps, people's eyes lead their hands each time a movement is made either in a more controlled (e.g. Johansson et al., 2001) or in a more natural setting (e.g. Land et al., 1999), or even when they just watch others executing the actions (e.g. Flanagan & Johansson, 2003). In a comparable way, some studies have also provided evidence for visual anticipation in language comprehension (Altmann & Kamide, 1999). Our study not only replicates but also extends these findings to show that people indeed anticipate future events but also actively gather the predicted external information relevant for their eventual behavioural goal during the course of complex actions (cf. Altmann & Mirković, 2009; Cuijpers, van Schie, Koppen, Erlhagen, & Bekkering, 2006). It is important to mention that it is not our intention here to make any kind of specific claims about the importance of the visuomotor or the language system for the predictive mechanism, as the design used in this study does not allow for any further specification on this matter. Our claim is, however, that this predictive mechanism is involved in situations requiring understanding of a given situation involving motor or linguistic behavioural goals.

Such a predictive mechanism would be generally beneficial, as it would aid and hence facilitate the process of understanding of a given situation. Accordingly, this predictive mechanism would allow for a fluent and optimal behavioural control while we interact with the environment motivated by our internal goals. Clearly, understanding of the situation at hand builds upon the knowledge people have already acquired through their past experiences, which is further integrated with the immediate external and episodic information actively gathered through our major orienting system – our goal-directed saccades. This explanation seems to suggest a clear link between a predictive mechanism and the process of understanding.

Different studies have tried to address this relation between human ability to make predictions of upcoming events and the process of understanding in action and in language. For instance, it has been suggested that the key requirement for action understanding is our ability to predict upcoming action steps (Prinz, 2006). The idea is that we have a tendency to simulate actions we observe being executed by others, reflected either in overt behaviour as imitation of the observed action (e.g. De Maeght & Prinz, 2004) or in increased brain activity of neuron population that fires both when watching actions in others and when performing the action ourselves (e.g. Rizzolatti, Fogassi, & Gallese, 2001). This process of re-enactment is suggested by Prinz to require the identification of the on-going action, but to even more so require the prediction of the action that will unfold from the current one. Accordingly, through this process of simulation, we make predictions and hence understand actions.

The re-enactment idea of perceptual-motor experiences has also recently been introduced to explain neurocognitive mechanisms behind language comprehension. Pickering and Garrod (2007) have, for instance, suggested that the language production system works as a forward model that simulates and predicts the speech stream. Such an emulator would then allow us to use these linguistic predictions to aid language comprehension in a similar way as this process occurs in action understanding. Interestingly, it has been found that for a listener to optimally understand a speaker, the listener's eye movements need to mirror the speaker's eye movements (Griffin & Bock, 2000; Richardson & Dale, 2005). Perhaps the most direct connection between the action and the language systems has recently been proposed in terms of neural exploitation (Gallese, 2007, 2008). This view suggests that our brain has utilised the existing and evolutionarily older neural networks involved in goal-related actions to serve the newly acquired function of language. Following on this idea, Glenberg and Gallese (2012) have provided an exhaustive description on how a mechanism of motor control – through for instance forward models – can be used to explain language learning, comprehension, and production.

It is interesting to note here that most of the studies investigating anticipatory behaviour and predictive processing in action and language use paradigms that are strongly relying on the visual system. This is not surprising since we know that vision is an important sense to humans in many aspects of our lives. Moreover, strong evolutionary indications stress its importance for survival. We know for instance that vision was the dominant sense in early primates, with its cortical areas in temporal and occipital cortex expanding significantly throughout its evolution towards the anatomy and physiology of the visual system in the present day humans (Kaas, 2008). The importance became even more evident by a recent identification of the *Pax6* gene as a possible master control gene for the eye development due to sharing common descent (Gehring, 2011; Halder, Callaerts, & Gehring, 1995), implying that our eyes have probably evolved only once early in the evolution. Also in our study, the paradigms used rely strongly on the visual input. It would be interesting to see if people make similar goal-related predictions that enter our brain through other senses, like for instance through our auditory or tactile streams. The most parsimonious approach would suggest that the same predictive mechanism is employed for all senses in tasks requiring goal-directed behaviour.

A couple of paradigm-related points from our study need further elaboration. First, although our visual world paradigm was constructed such that participants could perform the task correctly without using the visual input, this paradigm successfully detected anticipatory eye movements: The participants viewed the scene by looking at the upcoming action targets, using the available visual information for predicting the events in the sentence. These patterns of anticipatory eye movements in our visual world paradigm were present from the start of the experiment and remained throughout, implying a stable predictive processing of information, rather than patterns generated by some kind of a strategy that the participants developed over the course of the experiment. Stability of the predictive mechanism has already been empirically supported in studies on action control: Although we know that people strongly rely on the visual system while learning novel motor tasks (e.g. Sailer et al., 2005), with their eye-movement patterns changing during learning, they demonstrate anticipatory looks through the whole learning process (e.g. Förster, Carbone, Koesling, & Schneider, 2011). Findings from our study suggest a similar stability in language comprehension.

Second, our paradigms involved a more naturalistic and dynamic stimuli, and multi-step actions and multiple objects – involving experimental conditions that resemble more the real-life situations than the paradigms used in most of the previous studies on action and language. Our measures of predictive looks – looking times, gaze onsets and count ratio – have proven to be sensitive to detect similar patterns of anticipatory eye movements. This suggests that in principle the three measures could track the predictive mechanism as manifest through behaviour. It needs to be mentioned here, however, that the analysis of looking times produced slightly different patterns of anticipatory eye movements. Specifically, in the action observation paradigm, the first action step showed a

reactive pattern of eye movements, whereas the analyses of the gaze onsets and the count ratios showed anticipation during the same interval. It seems hence that looking time data capture processing of factors other than anticipation. Perhaps, these data reflect factors related to processing speed, attention and engagement in the task. When a participant's gaze lingers on a target for a long amount of time, this is possibly reflecting participant's interest in the target object, rather than their anticipatory capacity. In contrast, in the gaze-onset analysis, the total time of fixating a target is not taken into consideration. Accordingly, the overall processing speed of participants is likely to be the only additional factor modulating the results next to the anticipatory processing. For the count analysis, the influence of processing speed will still have an influence on the results, but less so than in the gaze-onset analysis. Our data seem to suggest that the gaze onset and count analyses are the most appropriate ways of operationalising anticipatory eye movements, since they best eliminate possible confounding factors.

These differences between the measures of anticipation for the first action step were not observed in the visual world paradigm. Interestingly, when inspecting the pattern of anticipatory eye movements in the first step and comparing it between the two paradigms, it becomes clear that the visual world paradigm demonstrates more anticipatory eye movements in this first action step for all of the three measures. This suggests that the first action step is easier to predict in the visual world paradigm than in the action observation paradigm. This is perhaps not that surprising if we take into account how much there is to predict in this first step during the two paradigms: whereas the onset of the verb occurs later in time and after some information in the sentence has already been provided, the onset of the actor's hands is the first event that happens in the action observation paradigm, providing scarce information to build one's expectations at that point in time.

Altogether, our study provides evidence for the idea that anticipatory eye movements are guided by a predictive mechanism necessary for fluent and timely control of decision making in both action and language. We suggest that this predictive mechanism is a fundamental part of human brain that allows for collecting of the necessary and available information that is meaningful to us based on our prior knowledge to understand the overall situation at hand. As such, this predictive mechanism allows us to optimise future behaviour that is a critical part of adaptive human cognition. An interesting way to follow would be to investigate if action and language rely on a single predictive neurocognitive mechanism or that these modalities utilise two separate prediction systems.

Acknowledgements

This work was supported by a Rubicon Grant (446-09-024) to the first author and a VICI Grant (453-05-001) to the last author from the Netherlands Organisation for Scientific Research (NWO).

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye-movements and mental representation. *Cognition*, *111*, 55–71. doi:10.1016/j.cognition.2008.12.005

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583–609. doi:10.1111/j.1551-6709.2009.01022.x

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks. *Cognitive Neuroscience*, *7*, 66–80. doi:10.1162/jocn.1995.7.1.66

Butz, M. V., Sigaud, O., & Pezzulo, G. (2007). Brains, anticipations, individual and social behaviour: An introduction to anticipatory behaviour systems. *Lecture Notes in Computer Science*, *4520*, 1–18. doi:10.1007/978-3-540-74262-3_1

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*, 4333–4345. doi:10.1016/j.visres.2006.08.019

Cavanagh, P. (2011). Visual cognition. *Vision Research*, *51*, 1538–1551. doi:10.1016/j.visres.2011.01.015

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *107*, 84–107. doi:10.1016/0010-0285(74)90005-X

Craighero, L., Leo, I., Umiltà, C., & Simion, F. (2011). Newborns' preference for goal-directed actions. *Cognition*, *120*, 26–32. doi:10.1016/j.cognition.2011.02.011

Cuijpers, R. H., van Schie, H. T., Koppen, M., Erlhagen, W., & Bekkering, H. (2006). Goals and means in action observation: A computational approach. *Neural Networks*, *19*, 311–322. doi:10.1016/j.neunet.2006.02.004

De Maeght, S., & Prinz, W. (2004). Action induction through action observation. *Psychological Research*, *68*, 97–114. doi:10.1007/s00426-003-0148-3

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827–1837. doi:10.1016/0042-6989(95)00294-4

Elliott, D., Hansen, S., Grierson, L. E. M., Lyons, J., Bennett, S. J., & Hayes, S. J. (2010). Goal-directed aiming: Two components but multiple processes. *Psychological Bulletin*, *136*, 1023–1044. doi:10.1037/a0020958

Epelboim, J. L., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., ...Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*, 3401–3422. doi:10.1016/0042-6989(95)00080-X

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*, 769–771. doi:10.1038/nature01861

Förster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11*, 1–16. doi:10.1167/11.7.9

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. doi:10.1038/nrn2787

Gallese, V. (2007). Before and below theory of mind: Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 659–669. doi:10.1098/rstb.2006.2002

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, *3*, 317–333. doi:10.1080/17470910701563608

Gehring, W. J. (2011). Chance and necessity in eye evolution. *Genome Biology and Evolution*, *3*, 1053–1066. doi:10.1093/gbe/evr061

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*, 905-922. doi:10.1016/j.cortex.2011.04.010

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279. doi:10.1111/1467-9280.00255

Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science*, *267*, 1788–1792. doi:10.1126/science.7892602

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151–171. doi:10.1016/j.actpsy.2010.11.003

Imamizu, H., & Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: Implications for higher-order cognitive functions. *Psychological Research*, *73*, 527–544. doi:10.1007/s00426-009-0235-1

Johansson, R. S., Westling, G., Backstrom, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *Journal of Neuroscience*, *21*, 6917–6932. Retrieved from http://www.jneurosci.org/content/21/17/6917.full

Kaas, J. H. (2008). The evolution of the complex sensory and motor systems of the human brain. *Brain Research Bulletin*, *75*, 384–390. doi:10.1016/j.brainresbull.2007.10.009

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*, 718–727. doi:10.1016/S0959-4388(99)00028-8

Keele, S. W., & Posner, M. I. (1968). Processing of visual feedback in rapid movements. *Journal of Experimental Psychology*, *77*, 155–158.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*, 159–166. doi:10.1007/s10339-007-0170-2

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*, 481–529. doi:10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye-movements. *Journal of Memory and Language*, *57*, 519–543. doi:10.1016/j.jml.2007.01.003

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*, 95–127. doi:10.1016/j.cognition.2004.03.002

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328. doi:10.1068/p2935

Miall, R. C., & Wolpert, D. (1996). Forward models for physiological motor control. *Neural Networks*, *9*, 1265–1279. doi:10.1016/S0893-6080(96)00035-4

Ondobaka, S., & Bekkering, H. (2012). Hierarchy of idea-guided action and perception-guided movement. *Frontiers in Cognition*, *3*, 579. doi:10.3389/fpsyg.2012.00579

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*, 105–110. doi:10.1016/j.tics.2006.12.002

Prinz, W. (2006). What re-enactment earns us. *Cortex*, *42*, 515–517. doi:10.1016/S0010-9452(08)70389-7

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*, 1045–1060. doi:10.1207/s15516709cog0000_29

Rizzolatti, G., Fogassi, G., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661–670. doi:10.1038/35090060

Rosander, K., & von Hofsten, C. (2011). Predictive gaze shifts elicited during observed and performed actions in 10-month-old infants and adults. *Neuropsychologia*, *49*, 2911–2917. doi:10.1016/j.neuropsychologia.2011.06.018

Rotman, G., Troje, N. F., Johansson, R. S., & Flanagan, J. R. (2006). Eye movements when observing predictable and unpredictable actions. *Journal of Neurophysiology*, *96*, 1358–1369. doi:10.1152/jn.00227.2006

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, *25*, 8833–8842. doi:10.1523/JNEUROSCI.2658-05.2005

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. doi:10.1126/science.7777863

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*, 1–5. doi:10.1167/11.5.5

van Elk, M., van Schie, H. T., Hunnius, H., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: Neurophysiological evidence for motor resonance in infancy. *Neuroimage*, *43*, 808–814. doi:10.1016/j.neuroimage.2008.07.057