ARTICLE IN PRESS

Acta Psychologica xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Eye movements during transitive action observation have sequential structure

Andrew Webb[a,*], Alistair Knott[a], Michael R. MacAskill[b,c]

[a] Dept. of Computer Science, University of Otago, Dunedin, New Zealand
[b] Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand
[c] Dept. of Medicine, University of Otago, Christchurch, New Zealand

ARTICLE INFO

Article history:
Received 1 August 2008
Received in revised form 8 September 2009
Accepted 23 September 2009
Available online xxxx

PsyINFO Classification:
2323
2330
2346

Keywords:
Action observation
Gaze behaviour
Proactive

ABSTRACT

In this paper, we explore how observers of a reach-to-grasp action can identify and distinguish between the agent and patient (i.e. target) of the action. We investigate the hypothesis that there is a characteristic sequential structure to the observer's pattern of saccades, with the agent being fixated first, and then the target. We report an experiment which indicates that this sequence of saccades, while not ubiquitous, is overwhelmingly more likely than chance. The experiment also sheds some light on the mechanisms which allow the observer to saccade from the agent of the action to the target.

1. Introduction

The topic of this paper is *visual action recognition*—the process whereby an observer watches an action and computes a mental representation of it. What is involved in this process depends on how one interprets the notion of an 'action representation'. Most narrowly, an action representation can be understood as a simple action category (e.g. *grab* or *push*). More broadly, it can be understood as a representation of 'what the agent did', including reference to any objects which were acted on (e.g. an agent might *grab a cup* or *push a button*). Most broadly, it can be understood as a representation of the complete event which the observer witnessed, including a reference to the agent as well as to the action performed. In this paper, we are interested in action representations in this broadest sense. We focus on a simple situation, in which an observer watches an agent reach and grasp a nearby target object. Our interest is in the perceptual mechanisms which allow the observer to build a representation of the grasp action as an event, including reference to two objects (the agent and the target), and indicating the role each object plays in the action.

Our particular interest is in how the observer identifies and represents the agent and the patient of the action. Two questions can be distinguished. One is about perceptual mechanisms: how does the observer perceptually identify which object in the viewed scene is the agent and which is the target? Another is about representations: how are thematic roles such as 'agent' and 'patient' defined, and what representational device binds particular objects to particular roles? These two questions are related, because action recognition mechanisms must ultimately deliver event representations. In this paper, our aim is to study the action recognition mechanism, to see how it might deliver distinct representations of agent and patient.

We will investigate a very simple hypothesis, which concerns the time course of visual attention during action recognition. The hypothesis is that an observer of a reach-to-grasp action canonically attends first to the agent, and then to the patient of the action. We will investigate this hypothesis by studying the eye movements of observers. There are some reasons to predict that observers will tend to saccade first to the agent. It is known that observers' attention is captured by objects which begin to move, especially if they are animate (Abrams & Christ, 2003). And it is also commonly assumed that the agent of an action exhibits more motion and/or animacy than the patient. (In fact, this is a criterion used by some to help *define* the notions of agent and patient—see e.g. Dowty, 1991.) There are also some good reasons to expect that observers of a reach-to-grasp action only attend to the target object after attending to the agent. Some recent studies (Flanagan & Johansson, 2003; Rotman, Troje, Johansson, & Flanagan, 2006) have found that observers of a reach-to-grasp action systematically

* Corresponding author.
E-mail address: awebb@cs.otago.ac.nz (A. Webb).

0001-6918/$ - see front matter
doi:10.1016/j.actpsy.2009.09.001

Please cite this article in press as: Webb, A., et al. Eye movements during transitive action observation have sequential structure. *Acta Psychologica* (2009), doi:10.1016/j.actpsy.2009.09.001

saccade to the target of the action well before the agent's hand reaches it. Observers appear able to identify the agent's intended target, and make an anticipatory saccade to it. The information needed to infer the agent's intended target must come from an analysis of the agent's attitude or movements very early during execution of the reach. Therefore, we might expect that observers will attend to some part of the agent before saccading to the target.

If there is a strong sequential structure in observers' saccades to the agent and target object, then it may be that observers make use of this structure to help distinguish the agent and patient of an observed reach-to-grasp action. They may even make use of the structure to help distinguish agent and patient in event representations. If an event representation holds a record of the sequence of attentional events, then agent and patient may be identifiable by their serial position in this sequence.

The experiment we report in this paper investigates whether there is indeed a canonical sequence of eye movements during the observation of a simple reach-to-grasp action. Naturally, this is a question which must be answered empirically. There are reasons for expecting the agent to be attended to before the target, but it is not a foregone conclusion that this will happen. In fact, the only study to date which has looked for sequential structure in the pattern of saccades to agents and patients in transitive actions (Griffin & Bock, 2000) did not find any. Our study has two aims: first to examine whether observers execute a canonical sequence of saccades to the agent and the target when watching a reach-to-grasp action, and second, if there is such a sequence, to investigate what mechanisms allow the observer to transition from the agent to the target.

Our experiment addresses some outstanding questions raised by the experiments of Flanagan and Johansson and Rotman et al. In both those studies, observers watched a single agent perform a series of reach-to-grasp actions. The identity of the agent was therefore not at issue; observers' attention was focussed on the agent's workspace (the tabletop on which the targets rested), and it seems likely that any stimulus-driven anticipation of the agent's intended target was due to attention to the agent's early hand trajectory.[1] These studies investigate action recognition in a narrow sense; observers are computing representations of 'what the agent did', rather than of whole events. In our experiment, observers watch a scene in which there are two possible actors, so the identity of the agent in each observed action needs to be perceptually established. In this situation, there are two reasons to saccade to the agent: firstly to identify the agent, and secondly to infer the intended target. In either case, as already discussed, there are reasons to predict that observers make an early saccade to the agent before saccading to the target.

We investigate this, by examining which sources of information about the agent observers use to anticipate the target. The agent's gaze probably provides information about identity and intended target, while the agent's initial hand movement probably only provides direct information about the intended target. We predict that observers can use gaze alone to identify the intended target, and frequently do so when recognising whole events.

Our experiment also seeks to address some unresolved questions in the study of Griffin and Bock (2000). In their experiment, observers were presented with a series of action stimuli, each depicting a different scene featuring an agent and a patient, and were asked to report what they saw using a complete sentence, so observers were certainly obliged to compute full event representations in this paradigm. However, Griffin and Bock's action stimuli took the form of static line drawings. This mode of presentation makes it hard to interpret an observer's eye movements. When the observer first looks at a scene depicting an action, she may need to compute a representation *of the scene* (see e.g. Oliva, 2005) which is distinct from her representation of the action itself. There is no easy way of distinguishing which eye movements relate to 'scene perception' and which relate to perception of the action. To avoid this confound, we used video action stimuli in our experiment. Observers were first shown a still video frame presenting the scene, including the two potential agents and a range of potential targets. After a delay, the video stimulus was played. It was assumed that any eye movements occurring after this point related to the action recognition process, rather than to scene recognition.

In each video stimulus there were two actors seated at a table, facing the observer, and three target objects placed within reach on the table (see Fig. 1 for an example of the scene). One of the actors (the agent) performed a normal reach-to-grasp action upon one of the objects, while the other actor (the non-agent) remained still, fixating a neutral position. The order of agents and targets selected was randomised across trials, so the observer could not predict it. Also, during production of the video stimuli, the actors themselves were not made aware of the target for a particular reach action until the start of that action so as to minimise any unconscious orientation on their part. After each video stimulus, the observer had to indicate the agent and patient (i.e., selected target) of the observed action. Observers were therefore required to recognise complete events, rather than just actions. Despite this, our subjects still performed anticipatory saccades to reach target in the majority of trials (see Section 3.1 for details).

2. Methods

2.1. Participants, equipment and materials

There were 8 participants (4 female), with a mean age of 29.8 (range 21–53), all with normal or corrected-to-normal visual acuity, and all naive to the hypotheses of this experiment. The stimulus videos were recorded at 25 Hz, and were presented on a 405 mm (720 pixel) wide by 303 mm (576 pixel) high CRT monitor refreshing at 100 Hz, placed 610 mm in front of the observer's eyes. Eye movements were recorded monocularly at 240 Hz via video-oculography using an iView X Hi Speed (SMI, Berlin).

Each stimulus video contained the same two actors, and same three objects, in the same configuration in the scene. In each video, one actor (the agent) reached for one of the objects (the target). Upon reaching the target, the agent grasped and lifted it slightly before replacing it. The agent then returned his/her hand to its starting position, and his/her gaze to the neutral fixation position. The other actor remained still throughout the action. The first discernible movement from the agent, in any video, was the onset of a saccade to the target of the reach, which occurred 40 ms (one frame) after the start of the video, but because the agents were told to perform the reaches normally, there was variation in the lengths of trials, and in the onset of the agents' other movements. The videos had a mean length of 5044 ms. There were three examples of each agent/target reach combination, giving us a total of 18 unique stimulus videos.

2.2. Procedure

At the beginning of each session, gaze position was calibrated using the iView X software. The observer fixated each of 13 targets

[1] In fact, in Flanagan and Johansson's experiment, the agent performed a predetermined sequence of reach-to-grasp actions. Observers' anticipations of the target could have been due to knowledge of this sequence, rather than to information in the stimuli. However, in Rotman et al.'s study, the agent's targets were not predictable in advance, and observers still reliably anticipated the target. In this case, anticipation must have been due to information in the stimuli.

Please cite this article in press as: Webb, A., et al. Eye movements during transitive action observation have sequential structure. *Acta Psychologica* (2009), doi:10.1016/j.actpsy.2009.09.001

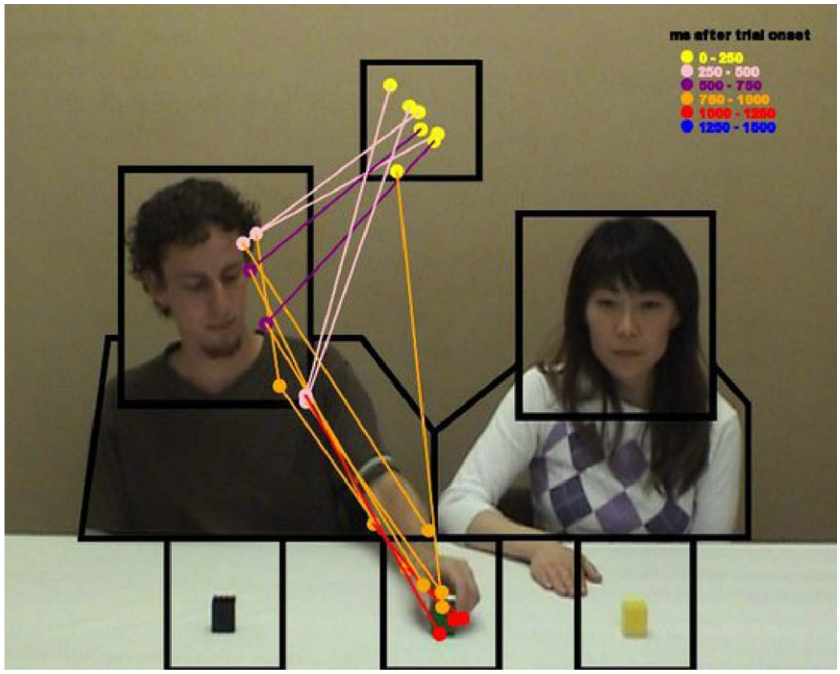

Fig. 1. A frame from one of the stimulus videos, showing the agent about to grasp the target. The black boxes show the Regions of Interest used in the sequence analysis. Overlaid is illustrative gaze data from a single subject, from all trials in which the left agent reached for the green target. Circles indicate the position of fixations, connected to the next fixation in the trial by a straight line. Colours indicate the time period (in 250 ms bins from the start of the trial) in which the fixation occurred.

distributed uniformly over the entire screen. Prior to each trial, the ongoing maintenance of accurate calibration was monitored by having the subject fixate five target points in a more restricted region encompassing the regions of interest (ROIs). If gaze position did not match one or more of the targets then the full calibration process was repeated prior to continuing with the trial. In practice, most subjects did not require re-calibration. At the beginning of each trial, the subject again fixated each of the five targets, presented for 1000 ms each in a random order every trial. These gaze data were recorded as a permanent record of objective calibration quality for each trial. Calibration targets were much smaller (16×16 pixel) than the regions of interest used in the analysis. Hence any undetected calibration errors could contribute to erroneous ROI assignment only in the case where fixations were very close to the borders of adjacent ROIs.

A single trial ran as follows:

1. The first frame of the video stimulus, which was the frame prior to the first observable movement made by the agent, was shown as a still for 2500 ms. The observer was instructed to freely explore the scene with their eyes during this period, allowing them to become accustomed to the scene. We assume that most eye movements occurring while the action was under way should therefore, be related to the process of action perception rather than scene perception.
2. A small (16×16 pixel) red fixation target was overlaid on the still frame, above and between the heads of the two agents, for a variable delay of 1300–1900 ms. The observer was asked to fixate this target when it appeared so that there was a constant gaze position at the beginning of each trial. Any trial which began with the observer not fixating the target location continued but was discarded from analysis.
3. The fixation target disappeared and the video began to play. The observer was free to move his/her gaze so as they saw fit. That is, no instructions were given as to what they should do during this period. They were only told that after the video ended, they would be asked to answer two questions about it (see below).
4. Each video ended 2000–3000 ms after the actor grasped the target.
5. The video was replaced with stylised black male and female silhouettes, corresponding spatially to the location of the previously displayed actors, above which was the text "Which person reached for the block?". The observer was asked to fixate the silhouette indicating which actor had performed the grasp. When satisfied that the correct answer was being fixated, the observer pushed a key which moved to the next display. Gaze position at the time of the keypress indicated their choice.
6. Three large squares were displayed, corresponding in colour and spatial order (left, middle, right) to the blocks in the video. Above them was the text "Which block was reached for?" Once more, gaze position at the time of the observer's keypress was used as the event report.
7. Calibration quality was checked and the next trial commenced.

2.3. Analysis

Each subject participated in three blocks of trials, where a block contained a single viewing of each unique video, presented in a random order. Thus there were a total of 432 trials recorded. Of these, 18 trials (one block's worth) were discarded because the data were corrupt; 56 trials were discarded because the subject was not fixating the fixation point at the start of the trial; and 12 trials were discarded because the subject failed to correctly indicate either the agent of the trial or the target of the trial. This resulted in 346 valid trials.

2.3.1. Regions of interest

For the purpose of analysis, we divided the video stimulus into eight regions of interest (ROIs, outlined in black in Fig. 1). These regions corresponded to the three possible targets ('left target', 'middle target', 'right target'), a head region and a body region for each of the two actors ('left head', 'left body', 'right head', 'right body'), and a region surrounding the central fixation square ('fixation square'). Using the data collected from a given trial (i.e. a single action-observation episode), we represented that trial as a sequence of ROIs. Any fixation that fell within one of our defined ROIs was labelled with the region's name; any fixation outside of these regions was labelled as having no region. As we were interested in the observers' eye movements relative to the action being observed, the label was defined relative to that action. For example, if the observed action was performed by the left actor, fixations in the left head region would be labelled as 'agent head' and fixations in the right-head region would be labelled as 'non-agent head'. For the sake of clarity, the agent head and agent body regions will be reported as a single region, called 'agent', and likewise with the non-agent regions. Similarly, if the target being acted upon were the middle target, then fixations to the middle target region would be labelled as 'target', whilst fixations in either of the other target regions (left target, right target) would be labelled as 'non-target'. Thus, each observation episode, independently of which actor and target were involved in the observed action, can be represented as an ordered list of region labels.

2.3.2. Temporal structure of saccades

We used this representation to search for any temporal regularities in the data, in other words, whether observers were looking at similar places at similar times. The fixations from each trial were placed into 250 ms bins, the first of which began at the start of the trial, such that if a fixation *started* in a particular ROI (e.g. the agent head region) during a particular time-period (e.g. at 700 ms after stimulus onset) then the corresponding bin (in this case, the 500–750 ms bin) was said to contain that fixation. The number of fixations in each ROI in each bin was then calculated. The bin data were then collapsed over all trials from all observers. The results are presented in Section 3.2.

2.3.3. Sequential structure of saccades

We also represented, for each trial, the subject's saccades as a sequence of ROIs, with within-ROI saccades collapsed to a single instance of that region. The motivation for this was that such within-ROI saccades do not actually change the general sequence of eye movements, but only change the timing of these movements, which is not a factor in this analysis.

We were then able to calculate whether particular sequences occurred more often than would be expected by chance. These results are presented in Section 3.3.

3. Results and discussion

3.1. Anticipatory saccades

As in the studies of Flanagan and Johansson (2003) and Rotman et al. (2006), our observers' saccades anticipated the agents' movement. In 292 (84.4%) of our 346 valid trials, the subject fixated the target at some point during the recording process. Of these, 262 (89.7%, or 75.7% of total valid trials) were anticipatory fixations, where the subject fixated the target prior to the actor's hand reaching it. This happened even though, before the trial started, the observer knew neither the actor who was to perform the action, nor the target involved in the action. As can be seen in Fig. 2D, the observers' fixations anticipated the grasp by a mean of 369 ms (95% confidence interval: 256–482 ms). Bearing in mind that the times given are for the start of the target fixation, and thus do not include saccade preparation time (on the order of 200 ms (Montagnini & Chelazzi, 2005) or saccade duration (which, in the current experiment, was 51.8 ms on average), the results strongly support Flanagan and Johansson's finding that the observers are anticipating, rather than reacting to, the agents' movements. And,

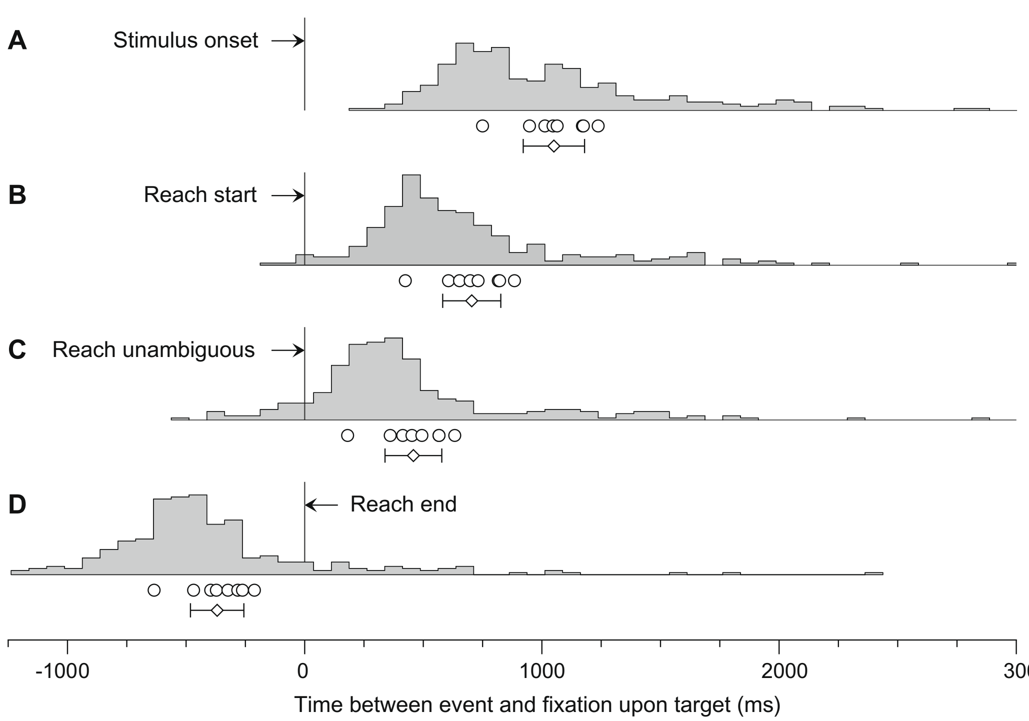

Fig. 2. The time at which the observer began to fixate the target relative to different events in a trial. (A) shows target fixations relative to stimulus onset; (B) target fixations relative to the start of the agent's reach; (C) target fixations relative to the time at which the movement of the agent's hand unambiguously indicated the target; and (D) target fixations relative to the time at which the agent's hand reached the target. Negative times indicate that the fixation reached the target prior to that event occurring. The histograms indicate the distribution of all trials, circles represent individual subject means, and diamonds indicate the group mean relative to each event (with 95% confidence intervals).

Please cite this article in press as: Webb, A., et al. Eye movements during transitive action observation have sequential structure. *Acta Psychologica* (2009), doi:10.1016/j.actpsy.2009.09.001

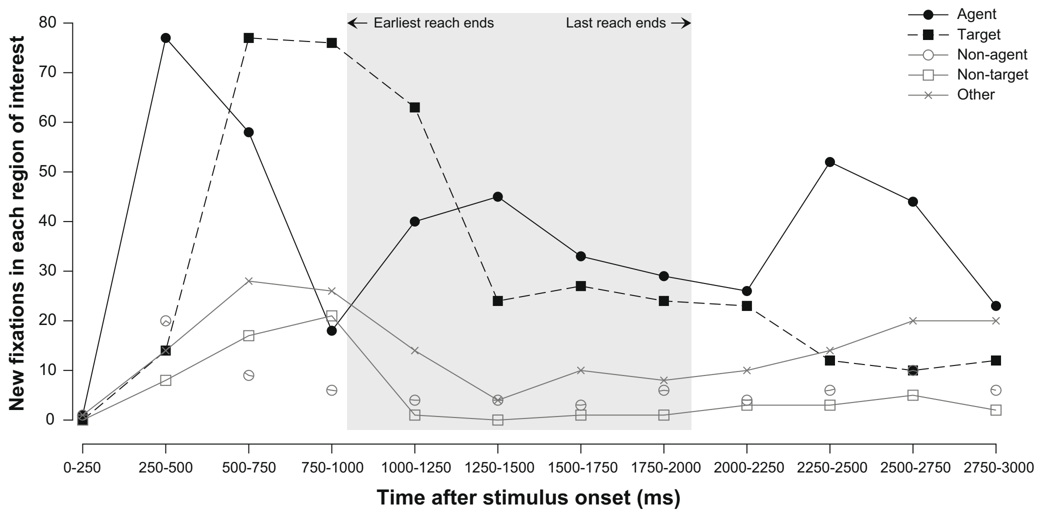

Fig. 3. The number of fixations in each Region of Interest over time. Each point represents the number of new fixations, occurring in a particular bin (bin size = 250 ms), that fell within each region. The shaded area shows the zone during which all reaches were completed.

as shown in Fig. 2D, the anticipatory pattern of saccades is not just an aggregate effect across subjects; it is shown for each individual subject.

We also want to investigate what sources of information subjects are using to anticipate the target. The two main candidates discussed above are the agent's gaze and the early movement of the agent's hand. In the first case, the subject would be using the agent's own preparatory saccade to glean information about the target of the reach by establishing joint attention (Driver et al., 1999), and in the second case, the subject would be extrapolating from the early trajectory of the agent's hand to anticipate its eventual destination. Both of these are certainly plausible. In the studies of Flanagan & Johansson (2003) and Rotman et al. (2006), for instance, the subjects had no access to the agent's gaze, and therefore could not be using it as a source of information; it is likely that hand trajectory was the main cue used to anticipate the target in both cases.

The most obvious way to examine which source of information subjects are using is to analyse where subjects look immediately before an anticipatory saccade to the target. In our case, the region most commonly fixated immediately prior to saccading to the target, was the agent ROI. This occurred in 42% of the anticipatory trials. However, it is also likely that the subjects are getting information by attending *covertly* to the action, as is evidenced by the fact that in 24% of the anticipatory trials, the fixation point ROI was the *only* ROI fixated before saccading to the target. Thus it is possible that, similarly to the studies mentioned above, subjects are attending to the agent's hand trajectory to extract information, albeit covertly. Fig. 2 provides information relevant to this hypothesis.

To investigate to what extent hand trajectory was a cue in our experiment, we analysed each video stimulus to determine the earliest time at which the agent's hand trajectory unambiguously indicated the intended target. All 18 videos were assessed for the time at which the reach became unambiguous (the validity of this subjective rating process was checked by having an independent rater make the same assessment for each video, with good agreement (Pearson's $r = 0.91$) between the raters), and the time at which subjects saccaded to the target was plotted relative to this time. The results are shown in Fig. 2C.

As the figure shows, there are a number of target fixations that occur too near (either before, or soon after) the time where the agent's hand trajectory allows the target to be assessed unambiguously, for this to be the only information used. It is worth reiterating that the times reported do not include saccade preparation time, which means that the subject must have identified the target earlier still. In these cases it seems likely that the subject is relying on other sources of information, such as the agent's gaze, rather than the agent's hand trajectory. However, it is possible that, in the trials where the target is fixated before the agent's reach became unambiguous, the subject was employing an opportunistic strategy of selecting and fixating *any one* the possible targets, rather than anticipating based on cues relating to the *actual* target. If this were the case, then, on those trials where one of the possible targets was fixated while the reach was still ambiguous, the likelihood of the correct target being fixated would be 33%, or chance. In fact, the subjects do rather better than chance, with 24 of the 37 instances of early fixations to possible targets involving the actual target of the action. The binomial probability of 24 or more correct fixations happening (given $n = 37$ and $p = 0.3333$) is small ($p = 0.00008$), which makes it unlikely that the subjects have no information about the actual target prior to the agent's reach becoming unambiguous.

In general, however, we cannot rule out the possibility that subjects supplement overt attention with covert attention, either to the agent's gaze or hand trajectory or both. For the majority of trials, by the time the saccade to the target had started, the agent's gaze had already alighted on the target, and the agent's hand was already in motion, and either of these (or one of a myriad other postural cues) would have given sufficient information for the subject to correctly anticipate the target.

3.2. Temporal structure of saccades

As Fig. 3 shows, there is a characteristic pattern to the binned data (described in Section 2.3.2). From shortly after stimulus onset to about 500 ms, as subjects move their fixation from the fixation square, the majority of fixations are in the agent ROI. After the 500 ms mark, however, observers have been able to discern what is to be the target of the reach action, and, from then on, the majority of fixations fall on the target ROI. Some typical scan paths are shown in Fig. 1. These results suggest that the observation of a reach-to-grasp action frequently involves a characteristic sequence

[2] The majority of the fixations on an agent fall on the agent's head (49 of 77 agent fixations in the first bin; 48 of 58 agent fixations in the second bin). This could suggest that the observer is following the agent's gaze in order to anticipate the target of the action. However, as previously discussed, anticipation is also possible without saccades to the agent's eyes.

Please cite this article in press as: Webb, A., et al. Eye movements during transitive action observation have sequential structure. *Acta Psychologica* (2009), doi:10.1016/j.actpsy.2009.09.001

of saccades: first a saccade to the agent of the action,[2] and then a saccade to its target. However, Fig. 3 is only intended to give a general picture of the order of saccades; it does not give explicit information about saccade sequences, as it groups data over subjects and trials. Patterns in individual saccade sequences are considered in the next section.

3.3. Sequential structure of saccades

Using the representation described in Section 2.3.3 above, there were 118 sequences of exactly length two in our data, of which 69 were ⟨agent, target⟩ sequences (note that this disregards cases which have the ⟨agent, target⟩ sequence as part of a longer sequence, such as when the subject looked at the agent, then at the target, and then looked elsewhere). There are 30 possible sequences of length two: six ROIs (target, non-target, agent, non-agent, fixation-point, none) in the first position and five ROIs in the second position (since we allow no repetition). If the probabilities of any of these sequences occurring are equal, as in the null hypothesis whereby the observers randomly move their eyes between ROIs, then the chance of an ⟨agent, target⟩ sequence occurring would be 0.033 or approximately 4 out of our 118 sequences of length two. The binomial probability of this event occurring 69 or more times (given $n = 118$ and $p = 0.033$) < 0.0001. This sequence of eye movements would be highly unlikely if the null hypothesis, that fixations were randomly directed, were the case.

4. Conclusion and future work

There are two important findings in this study. Firstly, it reinforces the earlier finding (Flanagan & Johansson, 2003) that observers of a reach action anticipate the target by saccading to it prior to the actor grasping it. Secondly, it presents evidence that observers' eye movements frequently follow a characteristic sequence, specifically, moving from the fixation point to the agent (generally to the agent's head) and then on to the target. This sequence is overwhelmingly more likely than chance.

What is responsible for this characteristic sequence? One possibility is that it is a contingent result of the structure of the stimulus, plus the observer's goal to establish the agent and patient (see e.g. Gesierich, Bruzzo, Ottoboni, & Finos, 2008). In our stimulus videos, the first detectable movement is of the agent, so it makes sense for the observer to identify the agent first. (In addition, initiation of movement captures attention, so there is a simple bottom–up reason why the observer first saccades to the agent.) The agent's gaze and/or hand trajectory provides the first evidence about which target is selected, so it makes sense for the observer to use these sources of information to anticipate the target.

Another possibility is that the sequence is a natural consequence of specialised neural mechanisms involved in action perception. If action perception uses the same sensorimotor machinery as is involved in executing actions, as Flanagan & Johansson (2003) suggest, then perhaps observers first attend to the agent to engage this machinery (for instance by adopting the agent's perspective), or as a necessary prerequisite of anticipating the target.

In fact, there may not be a conflict between these hypotheses. It would be natural for our action perception machinery to exploit the sequential structure in action stimuli to bootstrap the automation of a successful tactic for anticipating the target of reaches. That is, the first possibility, assuming it is successful, could naturally lead to its adoption as an automatic sequence of actions for the perception of this type of action. In any case, there is nothing in this study to favour one of these possibilities over the other, as both are entirely consistent with our results.

References

Abrams, R., & Christ, S. (2003). Motion Onset Captures Attention. *Psychological Science, 14*, 427–432.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*(3), 547–619.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509–540.

Flanagan, J., & Johansson, R. (2003). Action plans used in action observation. *Nature, 424*, 769–771.

Gesierich, B., Bruzzo, A., Ottoboni, G., & Finos, L. (2008). Human gaze behaviour during action execution and observation. *Acta Psychologica, 128*, 324–330.

Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.

Montagnini, A., & Chelazzi, L. (2005). The urgency to look: Prompt saccades to the benefit of perception. *Vision Research, 45*, 3391–3401.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *The encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Rotman, G., Troje, N., Johansson, R., & Flanagan, R. (2006). Eye movements when observing predictable and unpredictable actions. *Journal of Neurophysiology, 96*, 1358–1369.

Please cite this article in press as: Webb, A., et al. Eye movements during transitive action observation have sequential structure. *Acta Psychologica* (2009), doi:10.1016/j.actpsy.2009.09.001