

RUNNING TITLE: Sequential action representation in infancy

## **Sequential-action representation in infants: Evidence for concurrent-activation models**

Verschoor<sup>1</sup>, S. A., Paulus<sup>4</sup>, M., Spapé<sup>2</sup>, M., Biro<sup>3</sup>, S. & Hommel<sup>1</sup>, B.

<sup>1</sup>Leiden University Institute for Psychological Research & Leiden Institute for Brain and Cognition

<sup>2</sup>Helsinki Institute for Information Technology

<sup>3</sup>Center for Child and Family Studies, Leiden University

<sup>4</sup>Entwicklungs- und Erziehungspsychologie, Universität Erfurt

Keywords: action-effects; eye tracking; pupillometry; goal-directed action; infancy; cognitive development, sequential action.

Wordcount: 8149

**ABSTRACT**

*Infants interpret third-person sequential actions as goal directed by 6 months of age, around 9 months of age they start to perform sequential actions to accomplish higher order goals. The present study employed an innovative pupillometric and oculomotor paradigm to study how infants represent first-person sequential actions. We aimed to contrast chaining-, concurrent- and integrated models of sequential-action representation. 9- and 12- month olds were taught action sequences consisting of two elementary actions. Thereafter the secondary action was selectively activated to assess any interactions with the primary action. Results suggest that concurrent models best capture the representations formed.*

(98 words)



## INTRODUCTION

Infants are active, goal-directed agents. As demonstrated in numerous studies, they are able to perform a multitude of goal-directed actions, ranging from shaking rattles to enjoy themselves to grasping moving objects (e.g., McCarty, Clifton, Ashmead, Lee, & Goubet, 2001). Interestingly, some of these actions are sequential in nature, such as reaching for a rattle in order to shake it — a rather simple sequence, that nevertheless comprises of two dissociable components that differ in function and motor demands. Piaget (1936) and others (Claxton, Keen, & McCarty, 2003; Hauf, 2007; Willatts, 1999; Woodward and Sommerville, 2000; Woodward, Sommerville, Gerson, Henderson & Buresh, 2009) have stated that true goal-directed (sequential) action emerges around 9 months of age when infants begin to be able to organize means-end action sequences in the service of overarching goals. Although the goals in such sequences are commonly considered “higher-order” in one sense or another, the ontology of the implied (probably hierarchical) representational format of sequential action remains open. Until recently, most research has focused on infants’ intention reading of others’ sequential actions (e.g., Baldwin, Baird, Saylor, & Clark, 2001; Olofson & Baldwin, 2011; Woodward & Sommerville, 2000; Biro, Verschoor and Coenen, 2011; Verschoor and Biro, 2012), thus complicating the matter by entering interpretation into the equation. Only recently research has started to focus on the representational format of simple first-person action in infants and even less is known about the representational format of sequential action in infants. The purpose of the current study is to explore such representational formats.

There are three obvious prerequisites for infants to represent sequential action: that they can represent their actions, that they can represent sequential information, and that they can combine those abilities to represent sequential action. Let us turn to the first prerequisite, can

infants represent their own action and what is the ontology of such representations? There is ample evidence that simple actions are represented in terms of their effects. James (1890) in his ideomotor theory states that elementary actions are learned on the fly through sensorimotor exploration; an automatic and unintentional mechanism creates bidirectional associations between perceived effects and the action producing them (Hommel, 2009; Hommel, Müsseler, Aschersleben, & Prinz, 2001). These associations bring the actions under voluntary control, the agent can now activate the action by “thinking of” (i.e., endogenously activating the representation of) the corresponding effect. The theory can thus account for learning new actions and new goals and proposes that actions are represented in terms of their effects. This idea is typically tested in a two-stage paradigm. Experimenters first let participants perform voluntary actions that lead to specific effects. After acquisition, they test if exogenously cueing an effect also cues the action that previously caused it (Elsner & Hommel, 2001; Greenwald, 1970). This approach resulted in many demonstrations of bidirectional action-effect acquisition for a wide range of actions and effects in children (Eenshuistra, Weidema & Hommel, 2004; Kray, Eenshuistra, Kerstner, Weidema & Hommel, 2006) and adults, suggesting the mechanism responsible to be fast-acting (Dutzi & Hommel, 2009), automatic (Elsner & Hommel, 2001; Band, Steenbergen, Ridderinkhof, Falkenstein & Hommel, 2009), implicit (Elsner & Hommel, 2001; Verschoor, Spapé, Biro & Hommel, in press), and modulated by the same factors that influence instrumental learning (Elsner & Hommel, 2004) (for a review on action-effect learning see: Hommel & Elsner, 2009). Furthermore, action-effects have also been found to be important for action evaluation (Band, Steenbergen, Ridderinkhof, Falkenstein & Hommel, 2009; Verschoor et al., in press).

Infant research, on the other hand, has until recently mainly been restricted to the importance of action effects for third-person action interpretation instead of first-person action (e.g., Hauf, 2007; Kiraly, Jovanovic, Prinz, Aschersleben, & Gergely, 2003; Paulus, 2012; Woodward, 1998; Biro, Verschoor, Coalter and Leslie, under review). Evidence from this field cannot be considered as direct evidence for bidirectional action-effect acquisition. However, in the light of the upsurge of theories that stress similar representational formats for first person experience and observed action (e.g. Fabbri-Destro & Rizzolatti, 2008; Hommel et al. , 2001; Meltzoff, 2007; Tomasello, 1999), and that conceptualize action understanding as inverse planning (Meltzoff, 2006; Baker, Saxe & Tenenbaum, 2009), these findings can be seen as corroborative. Indeed action-effects have been implicated as instrumental for action understanding in 7 months-olds (e.g.; Biro & Leslie, 2007; Paulus, Hunnius, & Bekkering, 2012; for a review, see: Hauf, 2007; Kiraly, Jovanovic, Prinz, Aschersleben, & Gergely, 2003) and imitation in 9-months-olds (Hauf & Aschersleben, 2008; Klein, Hauf & Aschersleben, 2006; for a review, see: Elsner 2007; Meltzoff, 2006). Lately, more direct evidence regarding infant action-effect learning was obtained from first-person paradigms more similar to that of Elsner & Hommel (2001). A recent study from our lab showed that although 7- and 12-month olds used action effects for action evaluation, only the 12-month olds showed evidence of reversing bidirectional associations for action control (Verschoor et al., in press). By eight months, infants show stronger motor resonance when listening to previously self-produced action-related sounds than when hearing other sounds, suggesting that they represent actions in terms of their effects (Paulus, Hunnius, Elk & Beckering, 2011). The youngest age group thus far that showed evidence for using action-effects for action control were nine-month olds (Verschoor et al., 2010). Comparable results were found in 12-month olds (Verschoor et al., in press) and 18-

month olds (Verschoor, Weidema, Biro & Hommel, 2010). It thus seems safe to conclude that, by the time infants reach 7 months of age, they can represent and monitor own- and others action, and do so in terms of action effects. By the time they reach 9-12 months of age they can actively use these action effects for action control.

The second prerequisite for representing sequential action is the ability to encode sequential information. In other research domains, infants have indeed been shown to be responsive to sequential information in stimuli streams. For instance, it has been shown that neonatal infants are susceptible to sequential grammar information in speech stimuli (Gervain, Berent, & Werker, 2012; Teinonen, Fellmann, Näätänen, Alku & Huotilainen, 2009). Infants of 8 months are susceptible to similar information in artificial sound (Marcus, Fernandes & Johnson, 2007) and from 3 months of age they are susceptible to spatiotemporal sequences (Wentworth Hait & Hood, 2002). Studies like these suggest an early-appearing, domain-general information-acquisition mechanism for sequential information (Marcovitch & Lewkowicz, 2009). However, these studies only show a capacity for passively representing sequential information and do not specify the format of the representations.

Since the first two prerequisites seem to be met, the question remains whether infants can combine these abilities to merge simple actions into action sequences. Tentative evidence suggests they can: Looking-time studies provide evidence that by 7 months of age infants can detect whether a third-person action sequence is efficient given the goal and situational constraints (Biro, Verschoor & Coenen, 2011; Csibra, 2008; Gergely & Csibra, 2003; Verschoor & Biro, 2012). Similarly, Woodward and Sommerville (2000) found that 12-months-olds can relate elementary third-person actions to overarching goals of action sequences. Ballargeon, Graber, DeVos & Black (1990) reported that infants as young as 5.5 months of age can judge

whether an action sequence is sufficient to acquire an occluded toy or not. Other, more direct evidence was found in a study by Claxton et al. (2003), who reported that 10-months-olds adjust the kinematics of reaching for an object depending on what they intend to do with it. Taken together these studies suggest a rudimentary ability to represent first- and third person sequential action. Although Woodward and Sommerville (2000) propose a “causal” relational framework for evaluation of sequential action, and others (e.g. Gergely & Csibra, 2003) propose evaluation in terms of the efficiency, the studies mentioned above shed little light on the format of the sequential representations itself on which such evaluation might work. Understanding the ontology of sequential representations is essential in interpreting what it is that infants actually evaluate and is thus crucial in interpreting the results of studies in which evaluation of sequential action is the subject matter.

### **REPRESENTING SEQUENTIAL ACTION**

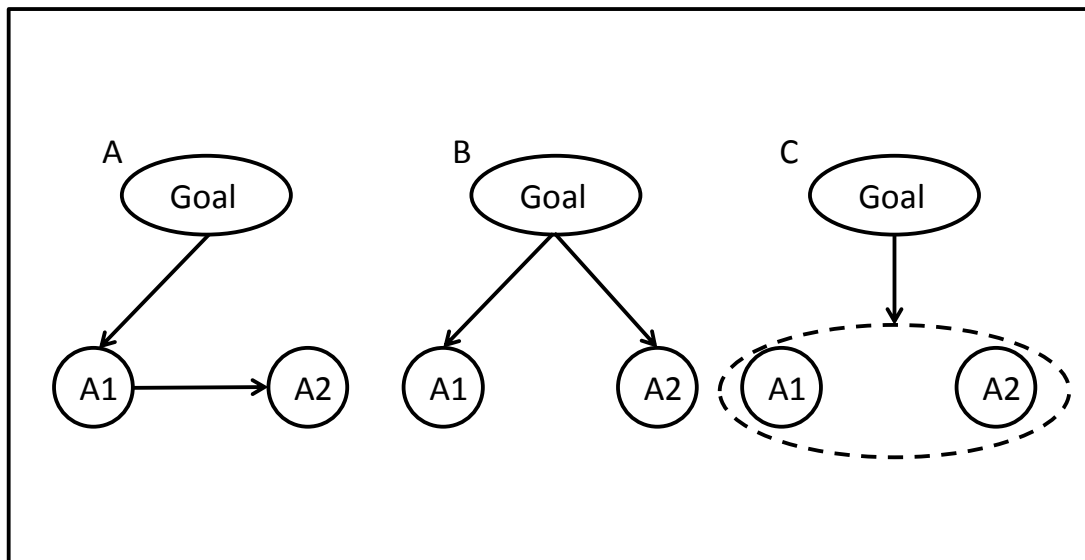
As there is little specific developmental literature on the subject, we will turn to general psychological theories of sequential action control. Through the years many influential incarnations of sequential action representation have been conceived (de Kleijn, Kachergis, & Hommel, 2013). They diverge on several aspects of the representational format stressing different mechanisms involved. Although these theories are by no means all inclusive or complete, they do posit a useful approximation for understanding sequential actions. Depending on exact circumstances relating to practice, content, time pressure and strategy, some models are more adept than others at explaining the empirical phenomenon involved. However, all of these theories hold that sequential actions consist of elementary actions that are somehow combined into a sequence. Furthermore, they agree that sequential actions are (to some extent) planned



ahead of execution, as suggested by the observation that the speed to initiate a sequence increases with the number of elements it comprises of (e.g., Henry & Rogers, 1960; Rosenbaum, 1987) or that people spontaneously start reaching movements with awkward hand postures if that optimizes the comfort achieved at the end of the action sequence (Rosenbaum et al., 1990).

Theories regarding sequential action representation can broadly be distinguished into three ontological types, which we will refer to as chaining, concurrent, and integrative theories of action-sequence representation (see Figure 1). Chaining theories stress that elementary actions are selected and combined into sequential ordered representations through association processes. During performance, activation cascades down the ordered elementary action representations resulting in an orderly performance of the sequence. Concurrent (Hebbian) theories focus on competitive processes that account for the orderly sequential production of an action sequence. Integrated approaches highlight crosstalk between elementary actions resulting in newly chunked elementary actions.

=== FIGURE 1 ===



*Figure 1. Models of sequential action.* :Schematic representation of activation in **A**: Chaining models of sequential action, activation cascades forward thru the different elementary actions, **B**: Concurrent models of sequential action, all elementary actions are activated simultaneously where after competition thru inhibition ensures the correct order of execution, **C**: Integrated models of sequential action, the sequence of actions has been integrated into a new elementary action.

The prototype of chaining theories of sequential action is the chaining theory proposed by James (1890). It holds that elementary actions can be chained into sequential actions by sequentially activating the anticipated effect of each element, thereby performing the subsequent actions in sequence via the ideomotor principle (Greenwald, 1970). With practice, the sensory effect of each elementary action becomes associated with the next elementary action through stimulus-response learning, thereby eliminating the need for sequential activation. The model thus effectively reduces sequential actions representation to a combination of ideomotor and stimulus-response learning. In terms of our example, this would mean that thru initially performing the sequence of reaching and shaking by chance, the infant would learn to unidirectionally associate reaching with shaking, thus enabling the chained action sequence. Ziessler (1994, 1998) indeed found that action-effects associations play an important role in implicit sequence learning. In a serial reaction-time task he systematically varied a sequence of locations of the stimuli to respond to. Indeed the action-effect (action-location) relations were more important and dissociable from stimulus-response relation learning for performance.

Although the James' version of chaining theory is temptingly simple, it has a number of important drawbacks resulting in several elaborations on the model to accommodate these.

Firstly, Hull (1931) pointed out that in order to be goal-directed and flexible in performance, the representation of the last action in the sequence and thus the end state of the action sequence should remain active during execution to compare the actual outcome to the expected outcome. He thus introduces a mild form of hierarchy into the representation by proposing additional activation of an overarching goal representation. This is indeed necessary for the action sequence to be driven by a conception of its function. Secondly, in the conception of James (1890) the second action of the sequence is cued by the sensory effect of the first action, meaning that sequential action would have to rely on sensory feedback. However, empirical evidence suggests that feedback mechanisms are too slow to accommodate such a reliance on feedback (e.g. Sternberg, Monsell, Knoll & Wright, 1978). Greenwald (1970) consequently suggested that instead of the sensory effects, the anticipated action effects of the preceding action should be associated to those of the next action, thereby circumventing the criticism. This addition also enables the initiation of the sequence by anticipating its end effect. In terms of our example, this would mean that through initially performing the sequence of reaching for and shaking the rattle by chance, the infant would learn to associate the effect of reaching with the effect of shaking. The infant thus only has to activate the representation of the rattling sound which in turn activates the effect of reaching, which would then activate the reaching action. This would start of the chained action sequence. However the model does not specify how the end effect activates the sequence instead of just the elementary action.

An important criticism of chaining models of sequential action is that they seem to imply that the effects of the elementary actions are equally associated with the preceding and subsequent action effects, thus making orderly performance of sequences impossible. In other words, chaining models of sequential action just assume, but fail to describe how activation

moves forward through the sequence. In terms of our example, these models fail to describe why the infant reaches first and shakes later, not vice versa. This lack of temporal dynamics in chaining models resulted in the emergence of the second ontological class of theories, the concurrent activation theories. To tackle the temporal shortcomings, Estes (1972) suggested an initial concurrent activation of all elementary actions in the sequence by a superseding goal representation. Thereafter, temporal inhibitory processes ensure that activation moves forward through the sequence in an orderly manner. To guarantee this forward flow he introduced inhibitory links flowing from each element to the next and secondary self-inhibition of completed elements to prevent repetition (e.g., Henson, 1998) equivalent to inhibition of return in visual attention (Posner & Cohen, 1984; Houghton & Tipper, 1996). Again, in terms of our rattling example, this would mean that by activating the overarching goal of rattling the rattle, the infant would activate both the shaking action and the reaching action which would then compete thru inhibitory processes until the reaching action is performed first and then the shaking action. The concurrent model can thus additionally account for the empirical finding for more prospective- than retrospective-intrusion errors. For instance, in a serial reaction-time tasks using over-learned sequences by Li, Lindenberger, R nger & Frensch (2000), more intrusion errors on the current goal were detected when the future goal rather than the past goal was cued, suggesting greater availability of intended than completed goals. Similar phenomena are known from linguistic research. For example, Dell, Burger, and Svec (1997) found more anticipatory than preserveratory errors in practiced speech. Inhibitory mechanisms to ensure serial order were successfully modeled by Rumelhart and Norman (1982) in their seminal study on mechanisms responsible for motor control in typewriting (for a review see: Houghton & Tipper, 1996).

Yet another class of theories on sequential action representation, integrated sequential action representation, does not presuppose that action elements remain independent from each other when combined into an action sequence. Such theories state that through practice elementary actions can be integrated into one common action plan or “chunk”, implying considerable crosstalk between the elementary actions (Miller, 1956; Sakai, Hikosaka, & Nakamura, 2004). Their main support comes from studies that find reductions in the sequence-length effect by extensive practice (e.g. Klapp 1995). Chunking of the elementary actions is possible by relating the sequence to internal or external context thus creating a unique identifying criterion for the associations (Hull, 1931). In terms of our example, the infant incorporates the reaching and shaking action into a new elementary action by relating it to internal context (e.g. the goal of the action) and/or external context (the situational constraints of the action to be performed). This is crucial since it would indeed be very inefficient for all reaching actions to be exclusively associated with shaking rattles. Logically the association needs to be context dependent and as such, to some extent unique from other associations with reaching. Stöcker & Hoffmann (2004) found evidence indicating that action effects can facilitate chunking in a serial reaction time task.

The chaining-, concurrent- and integrated models imply partly different predictions with regard to the spreading of activation from one (representation of a) sequence element to another. Consider a sequence of two elements, with element A preceding element B. All models imply that priming or otherwise activating A might spread activation to B, but they differ regarding their predictions in case that B is primed/activated. James' (1890) chaining model would not predict that priming B might lead to the activation of A, since the sequence is assumed to be represented by unidirectional stimulus-response bindings ( $A \rightarrow B$ ). Greenwald's (1970) version,

instead, would predict the spreading of activation from B to A, as sequences are represented through associations formed between the effects of the elementary actions. Conversely, concurrent activation models would predict that activating B leads to the inhibition of A, as activation is allowed to spread in a forward direction only, while “backward” connections are inhibitory. Finally, integrated models would predict that activating one element would activate the representation of the entire sequence, including A. The aim of the present study was to pit these different predictions against each other.

### **EXPERIMENTAL APPROACH**

To tackle our question regarding sequential-action representation, we modified a recently developed gaze-contingent paradigm to assess ideomotor action-effect learning in infants (Verschoor et al., in press). This eye-tracking paradigm overcomes problems that arise due to underdeveloped motor abilities in infants (Verschoor et al., in press; Wang et al., 2012) and enabled us to administer a paradigm to infants conceptually identical to the original Elsner and Hommel (2001) paradigm. The paradigm uses eye movements as actions, which is appropriate since infants can accurately control their eye movements from at least 4 months of age (Scerif et al., 2005) and these can be considered voluntary goal-directed actions (Gredebäck & Melinder, 2010; Falck-Ytter, Gredebäck, & von Hofsten, 2006; Perra & Gattis, 2010; Senju & Csibra, 2008;). An additional advantage of our paradigm is the concurrent recording of Task-Evoked Pupillary Responses (TEPRs) which is relatively new in developmental research (Falck-Ytter, 2008; Jackson & Sirois, 2009; Laeng, Sirois & Gredebäck, 2012; Verschoor et al., in press). Larger dilations indicate motivational phenomena such as increased arousal (Bradley, Miccoli, Escrig & Lang, 2008; Laeng & Falkenberg, 2007), attention allocation (e.g. Hess & Polt, 1960), cognitive load (Kahneman & Beatty, 1966), and mental effort (Kahneman, 1973; Hess & Polt,

1964). Whatever the exact interpretation of the measure, using it enables us to contrast acquisition contingent vs. non-contingent responses since all interpretations suggested that dilations should be larger for actions that require more processing (Verschoor et al., in press). In the Verschoor et al., (in press) paradigm 7-, 12-month olds and adults were to make eye-movements towards a visual stimulus appearing at the left or right of a fixation point, where the two directions produced different auditory effects. In the following test phase, participants were presented with the effect tones and could freely choose to make a saccade to the left or right of two simultaneously presented visual stimuli. We found evidence for the acquisition of bidirectional associations in all age groups. Latencies for responses incongruent with the previously experienced action-effect mapping were increased in 12-month olds and adults only, while pupillary responses were enhanced in incongruent conditions for all age groups—suggesting that even the 7-month olds picked up action-effect knowledge.

Since this paradigm works well with infants, and other research groups have also successfully demonstrated saccade-effect learning in adults (Huestegge & Kreutzfeldt, 2012; Herwig & Horstmann, 2011), we decided to adjust the paradigm and present participants with action sequences during acquisition. We adapted our gaze-contingent paradigm (Verschoor et al., in press) to teach infants two contrasting, sequential actions, each consisting of two perpendicular eye movements—each having a specific auditory effect. The sequences were taught in a free choice acquisition phase. Thereafter we cued one of the end effects of the sequences (i.e., we presented the auditory effect of one of the final sequence elements) and presented participants with the choice to perform one of the two first action components—a left- or right-ward saccade. If infants represent sequential action as sequences at all, cueing the effect of the second element should affect the activation of the first element. Hence, finding such an

effect (whether facilitatory or inhibitory) would provide evidence for the cognitive representation of sequential action as a true action sequence in infants. Moreover, the direction of the effect would speak to the internal structure of that representation: While chaining and integrative models would lead one to expect facilitation of the first element, concurrent activation models would rather predict the opposite, that is, a negative priming effect.

## **METHODS**

### **Subjects**

Two groups of infants were tested: 14 9-month-olds (mean: 8.94 months,  $SD = .37$ ,  $SE = .9$ , 5 female) and 16 12-month-olds (mean: 11.99 months,  $SD = .42$ ,  $SE = .10$ , 9 female). They were recruited through the municipality and they received small gifts as reward. An informed consent and a questionnaire regarding their general health and development were obtained from all caretakers. The infants were all healthy full-term and without pre- or perinatal complications. Another four 9- and seven 12-month-olds were excluded for not meeting the criterion for the minimal amount of test trials.

### **Test environment and apparatus**

During the experiment the infant sat in a specially designed, stimulus-poor curtained booth on the lap of his/her caretaker, who was seated in a chair in front of the eye-tracker apparatus. The distance between eyes and apparatus was approximately 70 centimeters (the screen's viewing angle was  $34.1^\circ$  by  $21.8^\circ$ ). The behavior of the infants was monitored and controlled online by the experimenter from a separate control room by means of a camera located above the apparatus. A 17 inch TFT-screen (1280 x 1024 pixels), equipped with an integrated Tobii T120 eye-tracker operating at 60 Hz, was used for visual and auditory data presentation, and for data collection. The Tobii T120 has an average accuracy of .5 visual



degrees and allows for a reasonable amount of free head movement by the subject (30x22x30cm). It recorded gaze direction and pupil-size. Stimulus presentation was controlled by a PC running E-prime® software (Schneider, Eschman & Zuccolotto, 2002).

### **Procedure**

Infants were tested in the laboratory at a time of day when they were likely to be alert and in good mood. Instructions were given to the caretakers prior to the experiment. The caretakers were instructed not to move the chair after calibration and gently hold the infant in order to maintain eye-tracker alignment and to entertain the infant during the 1-min interruption between calibration and the experiment. The eye-tracker was calibrated using a 9-point calibration consisting of a small animated dancing infant accompanied by music. The calibration was accepted with a minimum of eight points acquired successfully. The experimenter could play an attention-grabbing sound during the experiment to regain attention. If the attention grabbing sound did not work caretakers were encouraged to direct the infant's attention to the middle of the screen by pointing to it. Lighting conditions were kept constant during testing and across subjects. Furthermore the luminance levels were controlled for by presenting the stimuli in a random fashion. After completion of the test-phase, further information on the research and an explanation of the experiment was provided.

### ***Acquisition phase***

The experiment began with an acquisition-phase of 36 trials (see Figure 2). If during the acquisition phase the subject showed declining attention to the screen or was otherwise distracted the acquisition phase could be shortened (minimum amount of acquisition trials was set at 24). In each trial participants could freely choose to perform one of two saccade sequences. Each saccade sequence consisted of two distinct actions, first one to the left or right whereafter an up-

or downward action followed (depending on the mapping assigned). Each saccade was followed by an effect-sound which was consistently designated to left-, right-, up- and downward response areas (RA's) during the entire acquisition phase.

The background color of the screen was grey. A trial started with the presentation of a brightly colored dot with a superimposed line drawing ( $4.3^\circ$  by  $4.3^\circ$ ) being displayed at the center of the screen that served as start signal and fixation point. To keep infants interested, the color of the dot changed randomly from trial to trial (selected from eight bright colors) and the superimposed line drawing was randomly selected (without replacement) from a selection of 50 drawings (Snodgrass & Vanderwart, 1980). The dot disappeared after the subject fixated for an interval that varied from trial to trial (to remove any bias or habituation that might be caused by fixed intervals between trials) between 150 and 350 ms. Immediately after the dot disappeared, photographs of two different faces (randomly selected without replacement from 100 grayscale pictures from the "Nottingham scans" emotional faces database, <http://pics.psych.stir.ac.uk>, displaying emotionally neutral faces of 50 men and 50 women from a frontal perspective) appearing left and right from the dot. The faces served as response locations. Faces were chosen to elicit spontaneous saccades as they are known to attract attention in infants (Goren, Sarty & Wu, 1975; Johnson, Dziurawiec, Ellis & Morton, 1991). The  $5.3^\circ$  by  $5.3^\circ$  pictures appeared at  $9.7^\circ$ , center to center, to the left and right of the center of the screen. To avoid perseverance to either left or right across acquisition trials the images immediately started to pulsate. One of the faces started shrinking to  $4.1^\circ$  while the other started growing to  $6.5^\circ$  (which picture started shrinking was randomized); one cycle from intermediate size to small, to intermediate, to large and back to intermediate again, took 2 s.

When a saccade towards one of the two face locations was detected, the face at the other location disappeared and the targeted face stopped pulsating. Depending on the targeted side, one of two distinct 200 ms effect-sounds (“tring” or “piew”) was presented (with the mapping being balanced across participants). RA’s were defined as the maximum size of the pulsating images:  $6.5^\circ$  by  $6.5^\circ$ . A saccadic response was defined as eye movement into the left or right response area, which corresponded to a minimal  $4.3^\circ$  eye-movement. Immediately after the first effect had sounded the current face disappeared where after it reappeared  $7.8^\circ$  above or below (depending on the mapping) that location in the same dimension and continuing to pulsate. This face served as a RA again (defined again as the maximum size of the pulsating image). Upon detection of a gaze at that location one of two distinct 200 ms effect-sounds (“high note” or “low note”) was presented (with the mapping being balanced across participants). The minimal saccade was  $1.3^\circ$ . Reaction Times (RTs) were defined as the time it took from the disappearance of the central dot until a saccade into the secondary RA was detected. The maximum allowable RT was defined as 2000 ms; if the subject did not respond within this time the same trial was repeated. After each trial, an inter-trial-interval of 500 ms was presented.

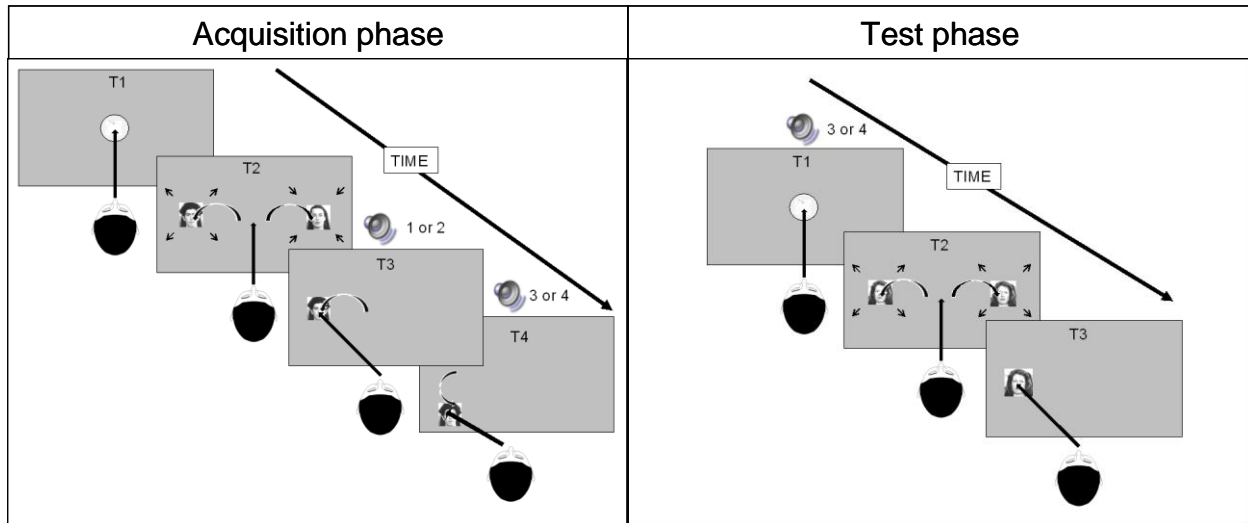
### ***Test phase***

The acquisition phase was followed directly by a test phase of 32 trials (see Figure 2). The minimum amount of test trials to enter analysis was 22. A test trial started with a similar dot as in the acquisition phase, again serving as a start- and fixation-stimulus. However, after the subjects fixated (fixation time identical to acquisition), the dot stayed on the screen for another 200 ms during which of the effect-sounds that was previously triggered by one of the two secondary eye-movements, was played. Upon the end of the sound the dot immediately disappeared. Then two identical  $5.3^\circ$  by  $5.3^\circ$  images of the same face (again randomly selected

without replacement from Nottingham scans' emotional faces database) appeared  $9.7^\circ$  to the left and right of the screen centre serving as RA. The two images were identical to minimize influence on the subject's gaze preference. To further reduce influence on the choice the faces now pulsed in synchrony, meaning that they either both grew or shrank (randomized and with the same motion parameters as in the acquisition trials). Again, the images were expected to evoke a saccade and the question of interest was whether the direction of this saccade would be biased by the tone.

In previous studies (that all used single-component actions), the definition of congruency was straightforward: If participants were exposed to two action-effect mappings in the acquisition phase, which responses A and B were followed by action effects 1 and 2 ( $A \rightarrow 1$ ;  $B \rightarrow 2$ ), performing action A in response to (or as a result of being primed by) action effect 1 in the test phase ( $1 \rightarrow A$ ) would be considered congruent, while performing the same action in response to (or as a result of being primed by) action effect 2 ( $2 \rightarrow A$ ) would be considered incongruent. Introducing actions that each consist of two components renders the definition somewhat more complicated. Our participants were exposed to two pairs of actions and action effects in the acquisition:  $A \rightarrow 1 + C \rightarrow 3$  and  $B \rightarrow 2 + D \rightarrow 4$ . In the test phase, we presented the action effect of one of the second components (3 or 4) and we tested whether this would affect processes related to one of the two first components. Hereby, the pairings of effect 3 and component A ( $3 \rightarrow A$ ) or of effect 4 and component B ( $4 \rightarrow B$ ) were considered congruent and the pairings of effect 3 and component B ( $3 \rightarrow B$ ) or of effect 4 and component A ( $4 \rightarrow A$ ) incongruent. Except for absence of the effect after the saccade, the remaining procedure was as in the acquisition phase.

==== FIGURE 2 ====



*Figure 2. Acquisition trial.*: Each trial starts with an intertrial interval of 500 ms. T1: A fixation dot is displayed at screen center. T2: After successful fixation, faces appear at either side of the screen where they started to pulsate. T3: Depending on the saccade target, the face at the other side disappears and an effect sound is played for 200 ms. T4: Depending on which side was chosen the face moves up or down. **Test trial.** Each trial starts with an intertrial interval of 500 ms. T1: A fixation dot is displayed at screen center. After successful fixation one of the previous action effects is played. T2: The dot disappears whereafter the same face appears on both sides. T3: The participant freely chooses where to saccade.

### Data acquisition

E-prime® was used to collect RTs, the number of left and right responses, and the number of congruent and incongruent responses during test. Furthermore, the E-gaze data files produced by E-prime® were imported into BrainVision Analyzer software (Version 1.05, BrainProducts, Germany) to analyze gaze position and pupillary data. Pupil sizes of both eyes

were averaged as to create more stable data. Artifacts and blinks as detected by the eye-tracker were corrected by using a linear interpolation algorithm where after a 10 Hz low-pass filter, commonly used for pupil dilation data (e.g., Hupe, Lamirel & Lorenceau, 2009) was applied. To ensure there were no erroneous pupil data we then rejected artifacts using the parameters of a minimal pupil size of 1mm and a maximum of 5mm, furthermore the maximum allowed change of pupil-size was defined as .03mm in 17ms. Eye gaze data were recorded in pixel coordinates and filtered using a 10 Hz low-pass filter.

Given that the acquisition of action-effect associations is sensitive to the same factors as stimulus-response learning (Elsner & Hommel, 2004), the Number of completed Acquisition Trials (NOCAT) was taken as an individual measure for how well action-effect associations were learned. The Mean Acquisition RT (MART) was taken as an individual measure for general speed and activity. Both NOCAT and MART variable were used as covariates in the analyses when appropriate.

## **RESULTS AND DISCUSSION**

### **Acquisition phase**

First we tested for age group differences in a number of dependent variables collected during the acquisition phase to ensure that the learning experiences of the two age groups were comparable (see also Table 1). All ANOVA's were performed with age group as a between-subjects factor. There were no effects for the percentage of completed acquisition trials ( $p > .5$ ), mean RT ( $p > .5$ ), or with respect to the percentage of right-vs.-left responses ( $p > .2$ ) or upward-vs. downward responses ( $p > .2$ ). The only two reliable effects were obtained for RTs. For one, horizontal response location interacted with age group,  $F(1,25) = 6.25$ ,  $p = .019$ ,  $\eta^2_p = .20$ . Separate analyses showed no main effect in the 9-months-olds (RT-left = 999ms, RT-right =

1046ms) and a tendency toward faster right-ward responses in the 12-months-olds,  $F(1,12) = 3.84$ ,  $p = .074$ ,  $\eta^2p = .24$  (RT-right = 982ms, RT-left = 1089ms). For another, vertical response location interacted with age group,  $F(1,25) = 4.63$ ,  $p = .04$ ,  $\eta^2p = .16$ . Separate analyses showed no main effect in the 9-months-olds (RT-up = 1008ms, RT-down = 1037ms) and a tendency toward faster downward responses in the 12-months-olds,  $F(1,12) = 3.82$ ,  $p = .07$ ,  $\eta^2p = .24$  (RT-up = 1089, RT-left = 982ms).

=== TABLE 1 ===

AGE GROUP SCORES	Percentage of completed acquisition trials	Percentage of left responses	RT in ms	RT left	RT right	RT up	RT down
9-month-olds	92.9 (11)	42.8 (38)	1007 (81)	999 (107)	1046 (90)	1008 (116)	1037 (83)
12-month-olds	92.0 (14)	38.5 (40)	1023 (83)	1089 (167)	982 (73)	982 (78)	1089 (164)

*Table 1.* Mean scores of acquisition phase (standard deviation in brackets).

Taken altogether, we can conclude that the learning experiences of the two age groups were comparable. The interaction of horizontal response location and age might reflect the fluctuating emergence of a general right-side preference during the first year (Corbetta & Thelen 1999; Michel, 1998), which also affects infants' eye movements (Cohen, 1972). A similar, but orthogonal effect may be reflected in our analysis of upward vs. downward RTs. However little is known to us about such preferences. Importantly, however, these observations are in no way detrimental to our main question since both age groups received approximately the same amount training for all combinations of response locations.

### **Test phase**

All ANOVA's were performed with age group as a between-subjects factor. There was no effect on the percentage of completed test trials ( $p > .4$ ). We considered four dependent variables: Response frequency, reaction time, pupil dilation, and gaze position. Especially the first three of these measures have been used to assess action-priming effects before (e.g. Verschoor et al., in press), commonly showing shorter reaction times, higher frequencies and lesser pupil dilations for responses that are congruent (as compared to incongruent) with the previously acquired action-effect association. However, there is strong evidence that these measures differ substantially in their sensitivity to congruency effects depending on the age of participants and the type of task involved (forced choice vs. free choice and manual vs. saccadic). Whereas forced- and free-choice reaction time, and manual-response frequency (e.g., Elsner & Hommel, 2001) are reliable indicators in adults, choice errors in forced-choice versions of the task are more sensitive to pick up congruency effects in children (Eenshuistra et al., 2004; Verschoor, Eenshuistra, Kray, Biro & Hommel, 2012). In infants, reaction time is a sensitive measure in free-choice versions of the task (the only version that infants can handle; Verschoor et al., 2010; Verschoor et al., in press), while response frequency reliably diagnoses congruency effects in older infants (from 18 months of age, only in manual versions of the task; see Verschoor et al., 2010). In contrast, pupil dilations (in free choice versions of the task) turned out to be highly sensitive to congruency in all age groups, showing lesser dilation (i.e., lesser effort) during congruent action. To provide a complete picture we will report findings on all three measures (together with findings for gaze direction) in the following. However, given that our task and basic setup adopted that of Verschoor et al. (in press), we expected reliable congruency effects mainly for reaction times and the pupil dilation measure.



**Response frequency**

Overall, participants looked more often (64%) to the right than the left side,  $F(1,28) = 9.00$ ,  $p = .02$ ,  $\eta^2p = .19$ , but this time the effect did not interact with age. More importantly for our purposes, an ANOVA with congruency as within-subjects factor and MART as covariate did not reveal any effect,  $ps > .1$ . We conclude that, if infants do represent sequential actions at all, this does not seem to affect the probability to choose a particular sequence.

**Reaction times**

There were no reliable effects with regard to overall RT,  $p > .6$ , left vs. right response location,  $p > .3$  (see Table 3), or inter-trial interval,  $p > .5$  (which we analyzed because the test-phase was self-paced). More importantly for our purpose an ANOVA with congruency as within-subjects factor revealed a significant effect indicating 29ms-slower responses for trials with congruent cues,  $F(1,28) = 4.15$ ,  $p = .05$ ,  $\eta^2p = .13$ ; the interaction with age was not significant,  $p > .3$ .

=== TABLE 2 ===

AGEGROUP SCORES	Percentage completed test trials	Percentage left responses	Percentage congruent responses	ITI ms	RT ms	RT Congruent ms	RT Incongruent ms
9-month-olds	93.5 (11)	43.2 (38)	47.5 (8)	1637 (323)	431 (90)	440 (104)	424 (91)
12-month-olds	96.5 (10)	29.0 (21)	49.3 (7)	1563 (393)	447 (74)	468 (83)	425 (83)

Table 2. Mean frequency and RT scores of test phase (standard deviation in brackets).

The main finding is that facing a cue of the final element of an action sequence interfered with executing the first. The fact that we find an effect at all can be considered as evidence that infants represented the entire action sequence in some coherent way. Thus it seems that repeatedly performing two consecutive actions is sufficient to integrate them into a coherent sequence representation even in infants. More specifically, infants apparently represent these action sequences in a format that allows for interactions between the codes of their individual elements (which excludes abstract or fully symbolic representational formats). Moreover, the direction of the effect does not support chaining or integrative theories of sequential action, as these would predict facilitation of earlier actions in a sequence by priming later actions. In contrast, our finding provides specific support for concurrent activation theories, as only these would predict interference.

=== FIGURE 3 ===

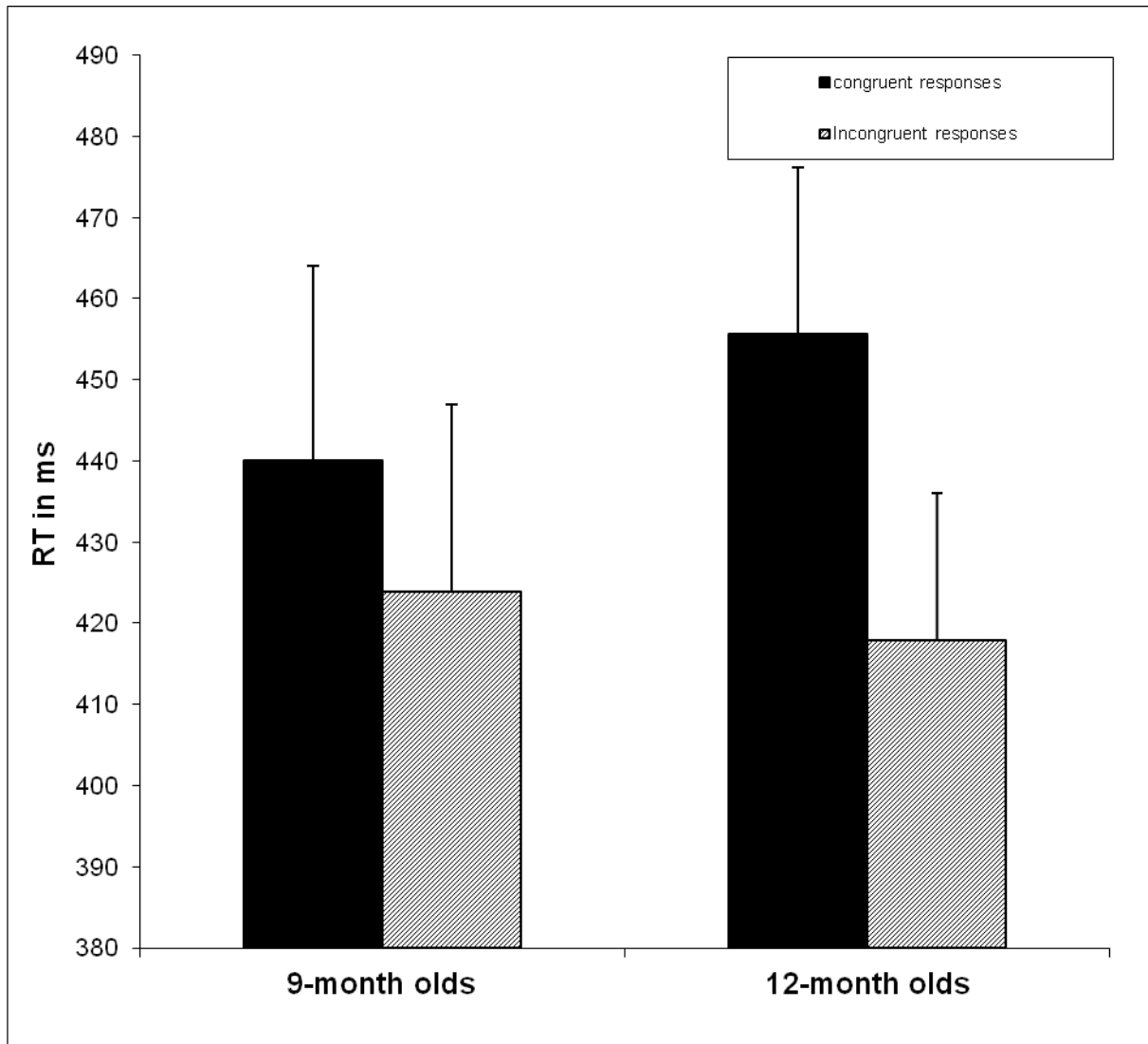


Figure 3. Mean reaction times (+SE) for 9-month-olds (N=14) and 12-month-olds (N = 16) in congruent and incongruent test trials.

### ***Pupil dilation***

Pupillary responses were sorted according to congruency of the response and the stimulus- and response-locked time functions were averaged (Verschoor et al., in press). Segments were created, depending on the analysis, from 2000ms before the presentation of the sound or RT to 8000ms after, while allowing for overlapping segments. Following the method

used by Bradley et al. (2008), pupil-diameter measurement began after the initial pupil reflex caused by the fixation stimulus. Visual inspection showed the light reflex to end around 500ms after effect presentation (see Figure 4) (see also Verschoor et al., in press). To accommodate for the variable RTs across age groups and conditions, we considered both stimulus-locked and response-locked Task-Evoked Pupillary Responses (TEPR's). Furthermore dilations were calculated as the percentage of dilation relative to the baseline to make the data more comparable across age groups. The percentage of trials rejected due to erroneous data points (leaving 29 valid trials on average) did not differ across age groups,  $p > .8$ .

The stimulus-locked analysis of pupil dilations for congruent and incongruent responses used a 500 ms pre-effect baseline (Beatty & Lucero-Wagoner, 2000). A repeated measures ANOVA on pupil dilations with congruency as within subjects factor revealed no a priori effects of congruency on baselines (-500 to 0 ms),  $p$ 's  $> .7$ . TEPRs in adults start from 200 to 300 ms after stimulus onset and peak around 1200 ms post-stimulus (Beatty & Lucero-Wagoner, 2000) in the range of 500ms to 2000ms (Beatty, 1982). We therefore calculated the mean pupil sizes for congruent and incongruent responses as the mean percentage of change from baseline to 500-2000 ms post effect onset. An ANOVA with MART as covariate revealed that, overall, participants exhibited larger relative pupil dilations during congruent responses?  $F(1,27) = 4.12$ ,  $p = .05$ ,  $\eta^2_p = .13$ , and this effect was not modulated by age group ( $p > .7$ ). Since the time window was based on adult findings and, thus, underestimates the pupillary reactions of the slower infants (Verschoor et al., in press), we reran the analysis with an extended time window of 1000-2500 ms post effect onset with the same baseline. Again, pupil dilations were significantly larger in congruent trials,  $F(1,25) = 5.03$ ,  $p = .03$ ,  $\eta^2_p = .16$ , and this effect was again independent of age,  $p > .09$  (see Figure 4).

=== FIGURE 4 ===

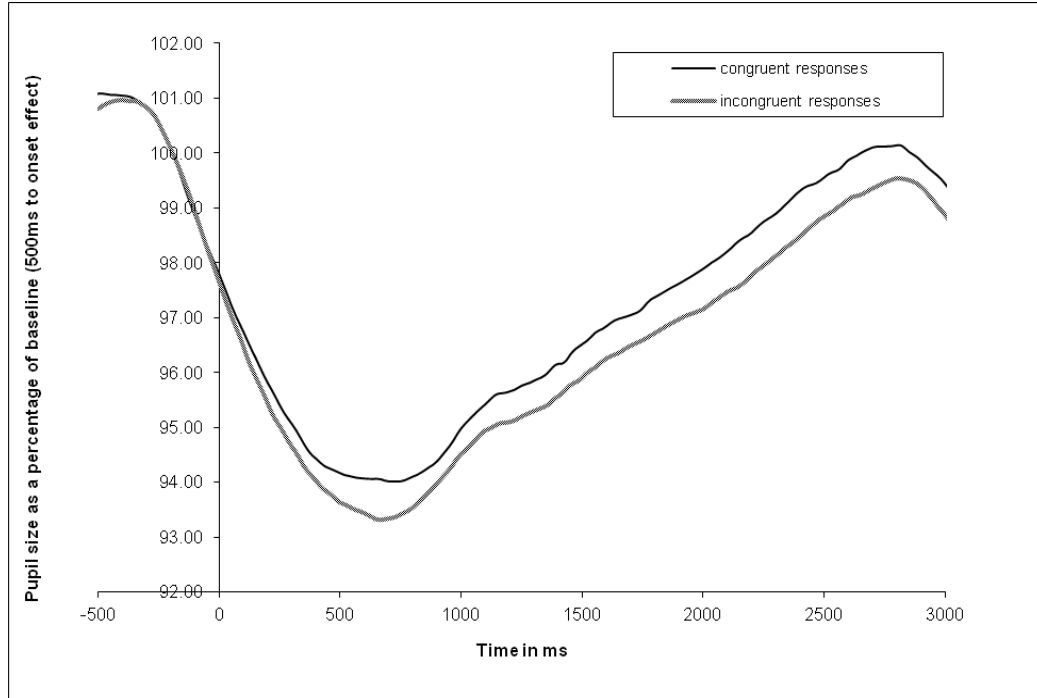
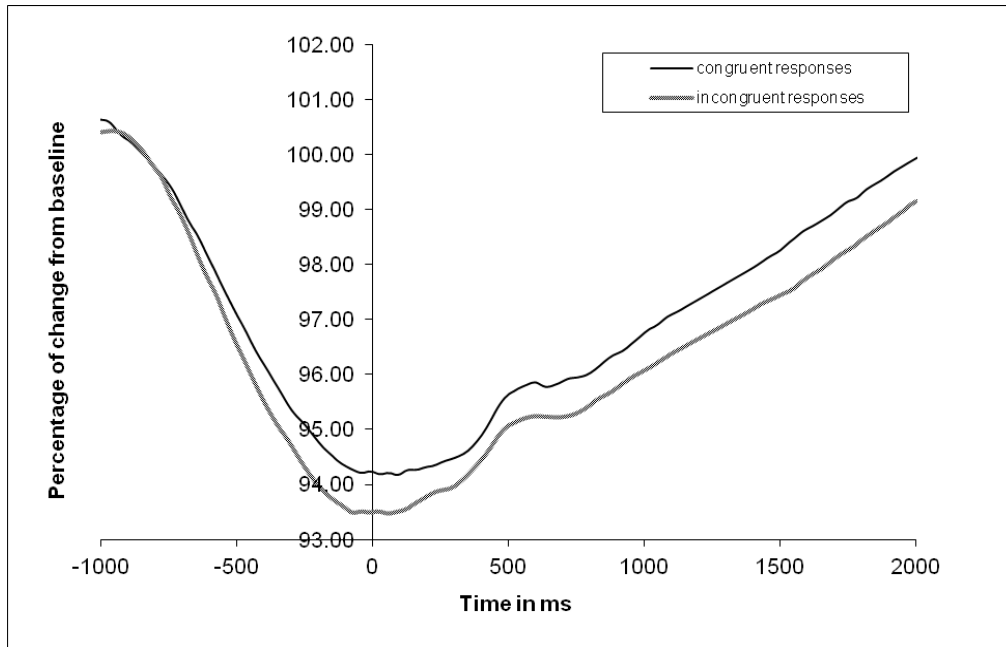


Figure 4. Relative pupil sizes for congruent and incongruent responses to baseline, stimulus-locked.

For the response-locked analysis, we calculated the percentage of dilation from a 700-ms time window starting at saccade onset, to the same 500-ms pre-stimulus baseline. An ANOVA with MART as covariate yielded a tendency for larger relative dilation in congruent trials,  $F(1,27) = 3.51$ ,  $p = .07$ ,  $\eta^2p = .12$ , while the interaction with age group was again not significant,  $p > .8$  (see Figure 5). Adding NOCAT as additional covariate resulted in a significant effect ( $F(1,26) = 5.48$ ,  $p = .03$ ,  $\eta^2p = .17$ ), again without an interaction with age group,  $p > .9$ .

=== FIGURE 5 ===



*Figure 5.* Relative pupil sizes for congruent and incongruent responses to baseline, response locked.

The finding that relative pupil dilations were larger for congruent responses in both stimulus-locked and response-locked analyses corresponds nicely to the outcome of the RT analysis: Cuing the second component of an action sequence makes the execution of the first slower and more effortful. Both findings provide evidence for action-sequence representation in infants and fit with the predictions from concurrent-activation theories.

### *Gaze position*

Our aim in the test phase was to prime the second element of the two-element sequences carried out in the acquisition phase by presenting the corresponding action effect. The three previous dependent measures (response frequency, reaction time, and pupil dilation) were taken to reflect the impact of this priming on the performance of left- and right-going saccades—the two first elements of the action sequences. All three measures thus tap into indirect priming

effects, effects induced by the spreading of activation from the primed second element to the associated first element. However, priming and activating the second elements of the sequences might also affect action planning directly and the effect on action planning could be modulated by the choice the participant makes during the trial (congruent vs. incongruent reaction). One of the second elements was an upward movement while the other was a downward movement. Priming these elements might thus induce a vertical bias: If presenting an action effect of the upward movement activated this element, this activation might induce an upward-going deviation from the horizontal midline; likewise, priming the downward movement might produce a downward-going deviation. But another scenario is possible: Activating the second of two elements might spread activation to the first, which then (because only horizontal movements were valid in the test phase) inhibits the vertical second element. This might induce the opposite bias—downward deviation when priming upward movements, and vice versa.

To investigate these effects, we analyzed the mean vertical deviation from the horizontal midline toward the primed action element as a function of congruency. To do this we collapsed all vertical deviations from horizontal midline toward the direction cued to one side and divided the data segments into congruent and incongruent from stimulus onset to 650ms thereafter (corresponding roughly to the mean RT plus mean random ITI) and compared these segments to a 150ms pre-effect baseline (the minimum fixation time before effect onset).

There were no a priori effects of congruency on baselines,  $p > .5$ . An ANOVA with MART as covariate showed that during congruent responses gaze position deviated vertically significantly less toward the direction cued by the effect sound,  $F(1,27) = 4.83$ ,  $p = .04$ ,  $\eta^2p = .15$  (effect size = 22 pixels; see Figure 6) than in incongruent responses, and this effect did not vary with age,  $p > .5$ . However, although we found a difference depending on congruency one cannot

distinguish between inhibitory processes or less activation unless the congruent responses (for which one may wrongly assume solely inhibition) were to deviate significantly away from the cued direction. To test this we conducted a t-test on the congruent responses to see if they significantly deviated away from the direction cued. This was not the case,  $t(29) = -1.64$ ,  $p > .1$  (we also performed a similar analysis on the segment from 100ms to 200ms since the graph shows a mean deviation away from midline in the incongruent reactions but this was far from significant ( $p > .4$ )). In addition we performed an analysis irrespective of congruency on vertical deviation toward the cued direction and found no significant deviation ( $p > .4$ ).

=== FIGURE 6 ===

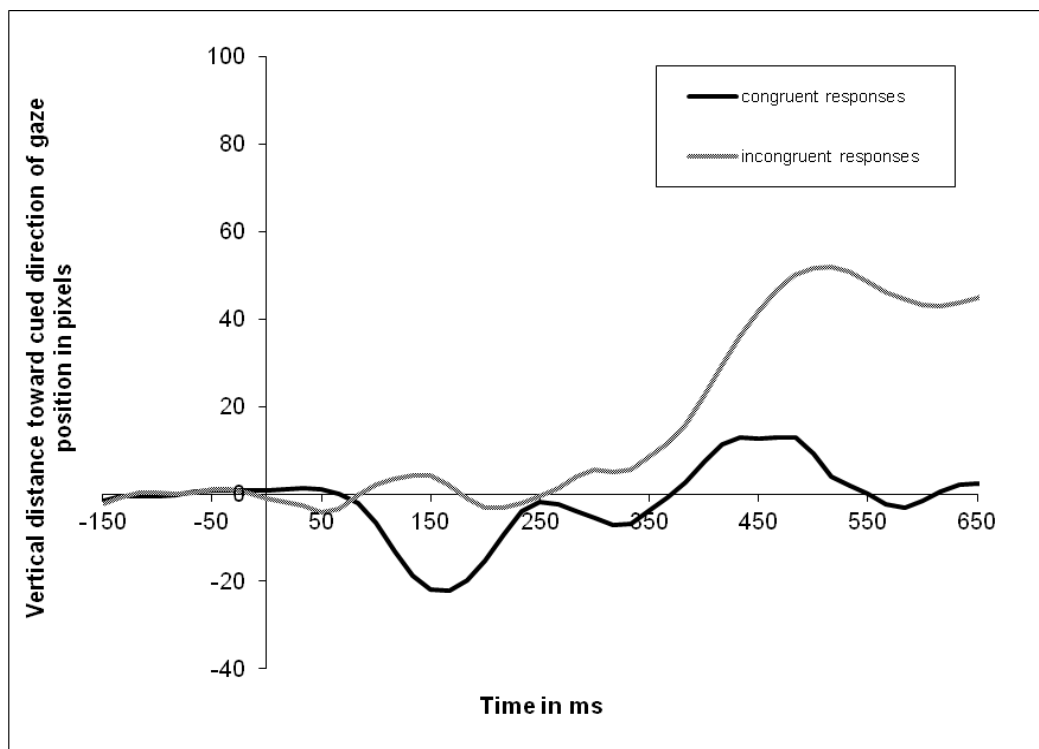


Figure 6. Vertical distance from midline toward cued direction of gaze position in pixels for congruent and incongruent responses. Time is 0, is the moment the effect starts.



Assuming that the vertical drifts reflect the amount of activation of the primed second action component, our finding suggests that priming results in inhibition of the primed component in congruent trials (as compared to incongruent trials)—inducing a bias away from the direction that this component represents. This might be due to competition between the activated components in congruent trials, as concurrent models would hold: activating the second component of a sequence might prime the associated first component up to a degree that induces the forward inhibition of the second. This finding provides further evidence for action-sequence representation in infants and fits with the predictions from concurrent-activation theories.

### **GENERAL DISCUSSION**

The aim of the current study was to examine whether and how 9- and 12-month old infants represent first person sequential actions. We hypothesized that priming an action component later in the sequence by cueing its effect should affect the availability of action components earlier in the sequence. We indeed found converging effects on two different measures indicating that activating the second component of a sequence does have an effect on the activation of the first component of that sequence. Additionally we found an effect on gaze position indicating that action control inhibits the second component of an action sequence whilst preparing the first part of the sequence. These findings can be taken as evidence that infants are indeed able to represent first-person sequential actions. Our second major finding is that the effect of (priming) the second on the first component is inhibitory in nature, as indicated in RT's and our pupillary measure. This finding is in agreement with the concurrent activation approach to sequential action (Estes, 1972) but not with integrative and chaining theories, which would have predicted a facilitatory effect instead.

Taken together, our observations suggest that cuing the second of two action components primes the corresponding action code, which leads to the inhibition of the first action component of the sequence. This slows down the initiation of this sequence, as indicated by the RT findings, and makes its execution more effortful, as suggested by the pupil data. It also biases action control away from the location implied in the second component, as implied by the gaze data, but this is insufficient to entirely prevent the corresponding action (as the frequency was unaffected). This means that our action prime had an impact on action control without dominating it altogether. Very interestingly, we had no evidence for any age effect in the relevant measures, suggesting that the underlying mechanisms are in place from rather early on.

Even though we consider the present findings as a first step towards the understanding of sequential-action representation in infants, the details of the suggested scenario are not entirely clear yet. Concurrent activation models would suggest that the inhibition we observed is a necessary outcome of the priming effect. Concurrent activation induces competition between alternative codes which, given the capacity limitations of the cognitive system and/or the functional need of action control to focus on one process at a time, induces inhibition in all losers of the competition—the winner takes all. This would explain our present findings, but alternative versions of this scenario are possible. For instance, it could be that the inhibition of the first sequence components was due to temporal differences in the process of activating the individual action components on the one hand and of the overarching goal representation on the other. It is well documented that initiating more complex sequential actions takes longer than initiating simpler actions (Henry & Rogers, 1960; Rosenbaum, 1987). One could thus speculate that cuing the second action component activated the underlying representation quickly but it took much more time to also activate the overarching goal representation. The eventual activation of this

goal representation could have facilitated both components of the sequence (as proponents of integration theories might suggest), but that may have taken too long to be picked up by our measures. As a consequence, the inhibition that our findings point to may reflect an only initial state of an actually more dynamic action-planning process. Another possibility would be that cueing an action component that is not yet appropriate (as none of the secondary components was a valid action in the test phase) resulted in the inhibition of not only the first component but of the entire sequence, perhaps including the goal representation. We cannot exclude that the second component of each sequence was also inhibited—although the lack of bias “away” from the direction cued by the second component suggests that it was not.

The current experiment was not set up to distinguish between these more detailed scenarios and more research will be necessary to fully understand the dynamics of action planning and infants. Nevertheless, our present findings demonstrate that young infants are able to construct action plans comprising of more than one element and to do so in a manner that puts the available elements into the right order. They also suggest that code competition and mutual inhibition are involved in the planning process. It seems essential to develop a more comprehensive theory of (the development of) sequential action representation, which would need to address how novel components are integrated into a sequential plan, how the sequencing is generated, and whether this requires hierarchical representations. We are confident that the present paradigm can be helpful in answering some of these questions, especially by introducing further modifications of the task (e.g., by cueing the first component and examine what effect this would have on the availability of the second). The paradigm might also help to increase our insight into general cognitive mechanisms underlying action planning, since saccade-effect learning is now well established (Huestegge & Kreutzfeldt, 2012; Herwig & Horstmann, 2011)

and as the paradigm provides various measures (frequency-, RT-, pupillary- and gaze position measures) that pick up different aspects of the planning process.

### **DISCLOSURE**

The authors declare no competing interests.

### **ACKNOWLEDGEMENT**

This research was supported by the Netherlands Organization for Scientific Research. We especially thank Thijs Schrama for technical support and Henk van Steenbergen for analytical support.

### **REFERENCES**

- Baker, C.L., Saxe, R., & Tenenbaum, J.B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 3, 329–349.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708-717.
- Band, G.P.H., van Steenbergen, H., Ridderinkhof, K.R., Falkenstein, M., & Hommel, B. (2009). Action-effect negativity: Irrelevant action effects are monitored like relevant feedback. *Biological Psychology*, *82*, 211-218.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292.
- Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. In J. Caccioppo, L.G. Tassinari, & G. Berntson (Eds.), *The handbook of psychophysiology* (2nd ed.) (pp. 142-162). Cambridge University Press.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*, 602-607.

Biro, S., & Leslie, A. (2007). Infants' perception of goal-directed actions. Development through cues-based bootstrapping. *Developmental Science*, *10*, 379-398.

Biro, S. Verschoor, S., Coalter, E. & Leslie, A. (under review). Outcome potential influences twelve-month-olds' interpretation of a novel action as goal-directed. *Cognitive Development*.

Biro, S., Verschoor, S. & Coenen, L. (2011). Evidence for a unitary goal-concept in 12 months old infants. *Developmental Science*, *14*(6), 1255–1260. Claxton, L.J., Keen, R., and McCarty, M.E. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science*, *14*, 354-356.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, *43*, 869-879.

Corbetta, D., & Thelen, E (1999). Lateral biases and fluctuations in infants' spontaneous arm movements and reaching. *Developmental Psychobiology*, *34* (4), 237-255.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*, 705-717.

Clearfield, M. W., & Thelen, E. (2001). Stability and flexibility in the acquisition of skilled movement. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 253–266). Cambridge, MA: MIT Press.

de Kleijn, R., Kachergis, G., & Hommel, B. (2013). Action representation, sequencing, and execution: an overview. Manuscript submitted for publication.

Dell, G.S., Burger, L.K., & Svec, W.R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123-147.

Dutzi, I. B., & Hommel, B. (2009). The microgenesis of action-effect binding. *Psychological Research*, *73*, 425-435.

Eenshuistra, R. M., Weidema, M. A., & Hommel, B. (2004). Development of the acquisition and control of action-effect associations. *Acta Psychologica*, *115*, 185-209.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 229-240.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research*, *68*, 138-154.

Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and action effects. *Acta Psychologica*, *124*, 44-59.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: Winston.

Fabbri-Destro, M. & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology*, *23*, 171-9.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*, 878-879.

Falck-Ytter, T. (2008). Face inversion effects in autism: A combined looking time and pupillometric study. *Autism Research*, *1*, 297-306.

Gervain, J., Berent, I., & Werker, J.F. (2012). Binding at birth: the newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, *24*(3), 564-574.

Gergely, G. & Csibra, G. (2003). Teleological reasoning in infancy: The one-year-olds' naive theory of rational action. *Trends in Cognitive Science*, 7(7), 287-292.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56, 544-549.

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

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. *Psychological Review*, 77, 73-99.

Hauf, P. (2007). Infant's perception and production of intentional actions. In C. von Hofsten & K. Rosander (Eds.), *Progress in Brain Research: From Action to Cognition*, 164, 285-301.

Hauf, P., & Aschersleben, G. (2008). Action-effect anticipation in infant action control. *Psychological Research*, 72, 203-210.

Henry, F. M., & Rogers, D. E. (1960). Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. *Research Quarterly*, 31, 448-458.

Henson, R. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, 36, 73-137.

Herwig, A., & Horstmann, G. (2011). Action-effect associations revealed by eye movements. *Psychonomic Bulletin & Review*, 18, 531-537.

Hess, E.H., & Polt, J. (1960), Pupil size as related to interest value of visual stimuli, *Science*, 132, 149-150.

Houghton, G. & Tipper, S.P. (1996). Inhibitory mechanisms of neural and cognitive control: applications to selective attention and sequential action. *Brain and Cognition*, 30, 20-43.

Hommel, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research*, 59, 176-186.

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849 – 937.

Hommel, B., & Elsner, B. (2009). Acquisition, representation, and control of action. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (eds.), *Oxford handbook of human action* (pp. 371-398). New York: Oxford University Press.

Huestegge, L., & Kreutzfeldt, M. (2012). Action effects in saccade control. *Psychonomic Bulletin and Review*, 19, 198-203.

Hull, C. L. (1931). Goal attraction and directing ideas conceived as habit phenomena. *Psychological Review*, 38, 487-506.

Hupe, J. M., Lamirel, C., & Lorenceau, J. (2009). Pupil dynamics during bistable motion perception. *Journal of Vision* 9, 1–19.

James, W. (1890). *The principles of psychology*. New York: Macmillan/Harvard University Press.

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science*, 12, 670-679.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40, 1–19.



Kahneman, D. & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583-1585.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, New York: Prentice-Hall.

Kiraly, I., Jovanovic, B., Prinz, W., Aschersleben, G. & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition*, *12*, 752-769.

Klein, A., Hauf, P., & Aschersleben, G. (2006). The role of action effects in 12-month olds' action control: A comparison of televised model and live model. *Infant Behavior and Development*, *29*, 535-544.

Kray, J., Eenshuistra, R., Kerstner, H., Weidema, M., & Hommel, B. (2006). Language and action control: The acquisition of action goals in early childhood. *Psychological Science*, *17*, 737-741.

Laengle, B., Sirois, S. & Gredebäck, G. (2012) Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*, 18-27.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.

Li, K. Z. H., Lindenberger, U., Rüniger, D., & Frensch, P. A. (2000). The role of inhibition in the regulation of sequential action. *Psychological Science*, *11*(4), 343-347.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, *18*, 387-391.

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science*, *12*, 6, 1020-1025.

McCarty, M. E., Clifton, R. K., Ashmead, D. H., Lee, P., & Goubet, N. (2001) How infants use vision for grasping objects. *Child Development, 72*, 973-987.

Meltzoff, A. N., & Prinz, W. (Eds.) (2002). *The imitative mind: Development, evolution and brain bases*. Cambridge, UK: Cambridge.

Meltzoff A. N. (2006). The ‘like me’ framework for recognizing and becoming an intentional agent. *Acta Psychologica, 124*, 26-43.

Meltzoff, A. N. (2007). The “like me” framework for recognizing and becoming an intentional agent. *Acta Psychologica, 124*, 26–43.

Michel, G. F. (1998). A lateral bias in the neuropsychological functioning of human infants. *Developmental Neuropsychology, 14* (4), 445-469.

Müsseler, J. & Prinz, W. (2002) (Hrsg.). *Lehrbuch Allgemeine Psychologie*. Heidelberg: Spektrum Akademischer Verlag.

Olofson, E. L., & Baldwin D. (2011). Infants recognize similar goals across dissimilar actions involving object manipulation. *Cognition, 118*, 258-264.

Paulus, M. (2012). Action mirroring and action understanding: An ideomotor and attentional account. *Psychological Research, 76*, 760-767.

Paulus, M., Hunnius, S., & Bekkering, H. (2012). Neurocognitive mechanisms subserving social learning in infancy: Infants’ neural processing of the effects of others’ actions. *Social Cognitive and Affective Neuroscience*.

Paulus, M., Hunnius, S., van Elk, M., & Bekkering, H. (2012). How learning to shake a rattle affects 8-month-old infants’ perception of the rattle’s sound: Electrophysiological evidence for action-effect binding in infancy. *Developmental Cognitive Neuroscience, 2*, 90-96.

Perra, O., & Gattis, M. (2010). The control of social attention from 1 to 4 months. *British Journal of Developmental Psychology*, 28, 891–908.

Piaget, J. (1936, 1963). *The origins of intelligence in children*. New York: W.W. Norton & Company, Inc.

Posner, M. I., & Cohen, Y. A. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance* (pp. 531–554). Hillsdale, NJ: Erlbaum.

Prinz, W. (1990). A common coding approach to perception and action. In Neumann O., Prinz W. (Eds.), *Relationships between perception and action* (pp. 167-201). Berlin: Springer.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129-154.

Rosenbaum, D. A. (1987). Successive approximations to a model of human motor programming. *Psychology of Learning and Motivation*, 21, 153–182.

Rosenbaum, D.A., Marchak, F., Barnes, H.J., Vaughan, J., Slotta, J., & Jorgensen, M. (1990). Constraints for action selection: overhand versus underhand grips. In: Jeannerod, M. (ed.), *Attention and performance XIII* (pp. 321–342). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & Norman, D. 1982. Simulating a skilled typist: A study of skilled cognitivemotor performance. *Cognitive Science*, 6, 1–36.

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research*, 152, 229–242.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Scerif, G., Karmiloff-Smith, A., Campos, R., Elsabbagh, M., Driver, J., & Cornish, K. (2005). To look or not to look? Typical and atypical development of oculomotor control. *Journal of Cognitive Neuroscience*, 4, 591-604.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18, 668-671.

Snodgrass, G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for naming agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174-215.

Sternberg, S., Monsell, S, Knoll, R.L., & Wright, C.E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In G. E. Stelmach (Ed.) *Information Processing in Motor Control and Learning*. New York: Academic Press. Pp. 117-152.

Stocker C, Hoffmann J (2004) The ideomotor principle and motor sequence acquisition: tone effects facilitate movement chunking. *Psychological Research*, 68, 126–137.

Teinonen T., Fellmann V., Näätänen R., Alku P., Huotilainen M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10, 21.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Verschoor, S. A., Weidema, M., Biro, S., & Hommel, B. (2010). Where do action goals come from? Evidence for spontaneous action-effect binding in infants. *Frontiers in Cognition*, 1, 201.

Verschoor, S. A., Eenshuistra, R. M., Kray, J., Biro, S. & Hommel, B. (2012). Explicit learning of arbitrary and non-arbitrary action–effect relations in adults and 4-year-olds. *Frontiers in Psychology*, 2, 354.

Verschoor, S. & Biro, S. (2012). Primacy of Information About Means Selection Over Outcome Selection in Goal Attribution by Infants. *Cognitive Science*, 36, 714–725.

Verschoor, S., Spapé, M., Biro, S., Hommel, B. (in press). From outcome prediction to action selection: Developmental change in the role of action-effect bindings *Developmental Science*.

Wang, Q., Bolhuis, J., Rothkopf, C.A, Kolling, T., Knopf, M., & Triesch, J. (2012) Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm. *PLoS ONE*, 7(2).

Willatts, P. (1999). Development of means-end behavior in young infants: Pulling a cloth to retrieve a distant object. *Developmental Psychology*, 35, 651–667.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1-34.

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, 11, 73-76.

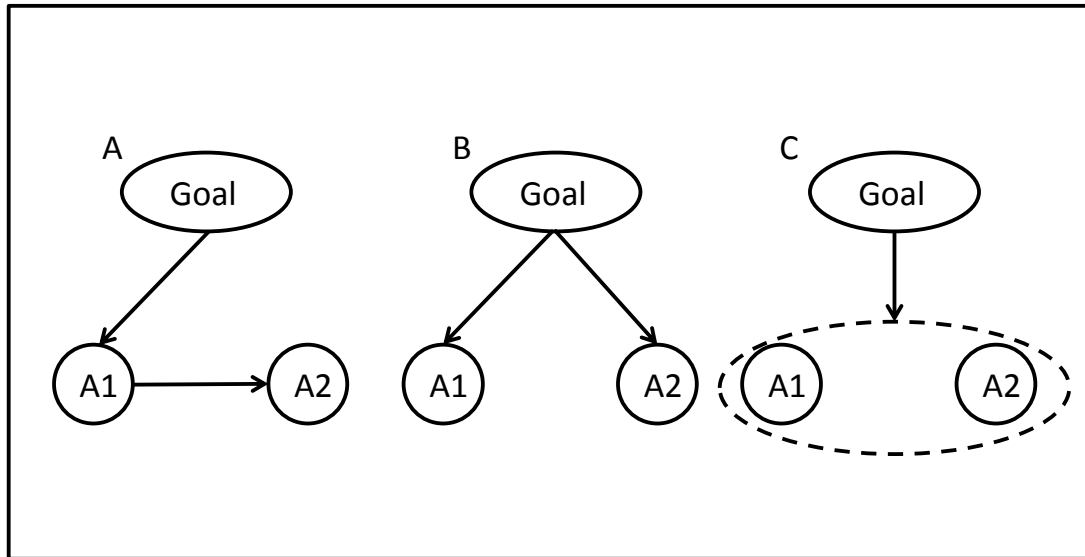
Woodward, A. L., Sommerville, J.A., Gerson, S., Henderson, A. M. E., & Buresh, J. S. (2009). The emergence of intention attribution in infancy. In B. Ross (Ed.) *The Psychology of Learning and Motivation, Volume 51*. Academic Press.

Wentworth, N., Haith, M. M., & Hood, R. (2002). Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy*, 3, 303 – 321.

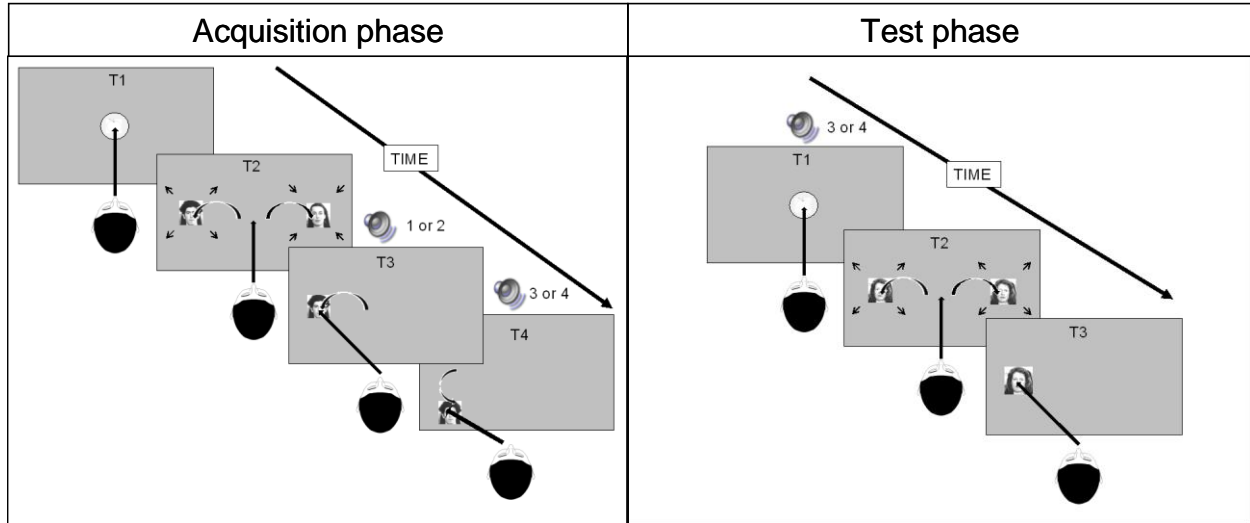
Ziessler, M. (1994). The impact of motor responses on serial pattern learning. *Psychological Research*, 57, 30-41.

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 962–978.

**FIGURES**



*Figure 1. Models of sequential action.* :Schematic representation of activation in **A**: Chaining models of sequential action, activation cascades forward thru the different elementary actions, **B**: Concurrent models of sequential action, all elementary actions are activated simultaneously where after competition thru inhibition ensures the correct order of execution, **C**: Integrated models of sequential action, the sequence of actions has been integrated into a new elementary action.



*Figure 2. Acquisition trial.* Each trial starts with an intertrial interval of 500 ms. T1: A fixation dot is displayed at screen center. T2: After successful fixation, faces appear at either side of the screen where they started to pulsate. T3: Depending on the saccade target, the face at the other side disappears and an effect sound is played for 200 ms. T4: Depending on which side was chosen the face moves up or down. **Test trial.** Each trial starts with an intertrial interval of 500 ms. T1: A fixation dot is displayed at screen center. After successful fixation one of the previous action effects is played. T2: The dot disappears whereafter the same face appears on both sides. T3: The participant freely chooses where to saccade.



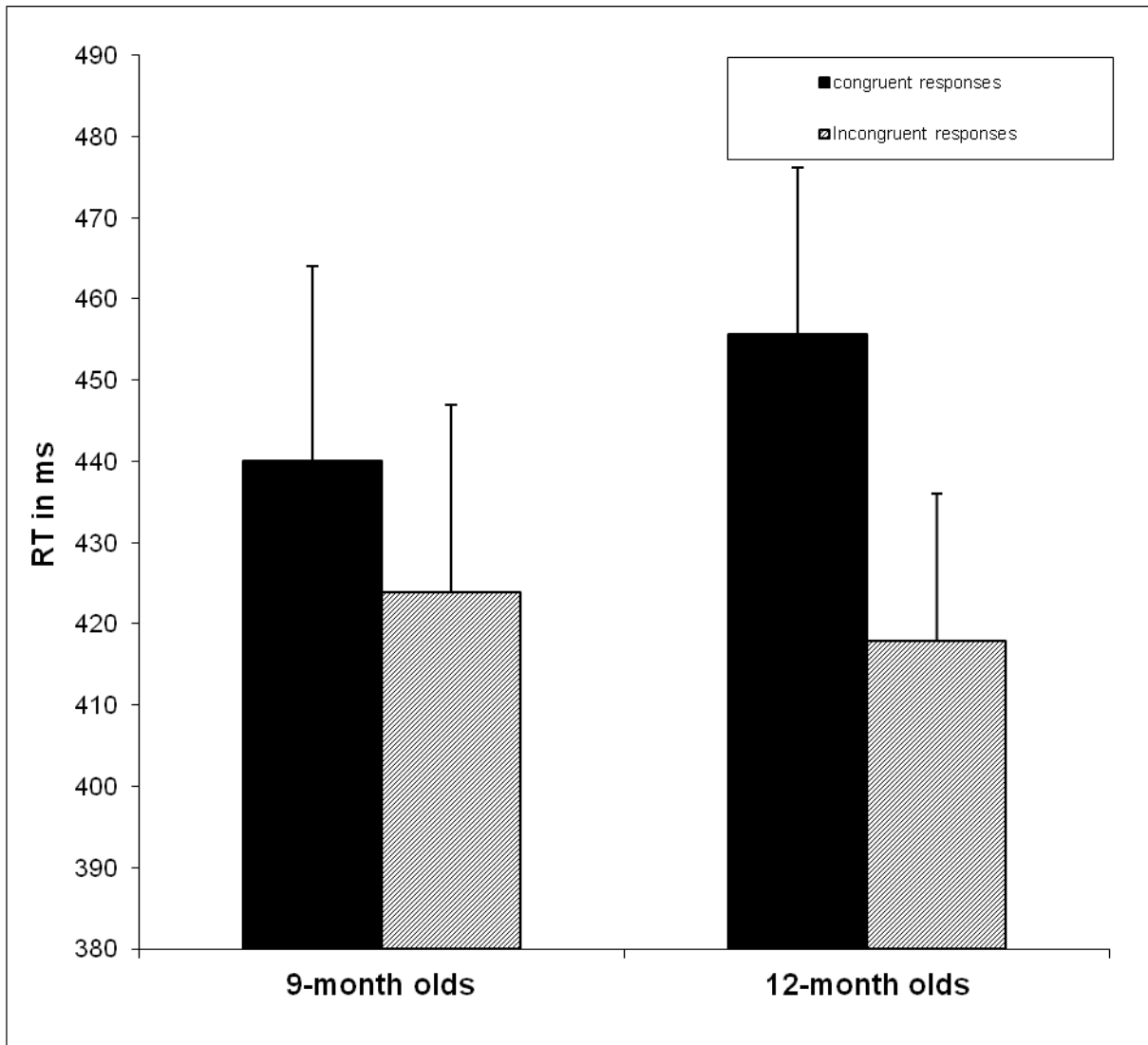


Figure 3. Mean reaction times (+SE) for 9-month-olds (N=14) and 12-month-olds (N = 16) in congruent and incongruent test trials.

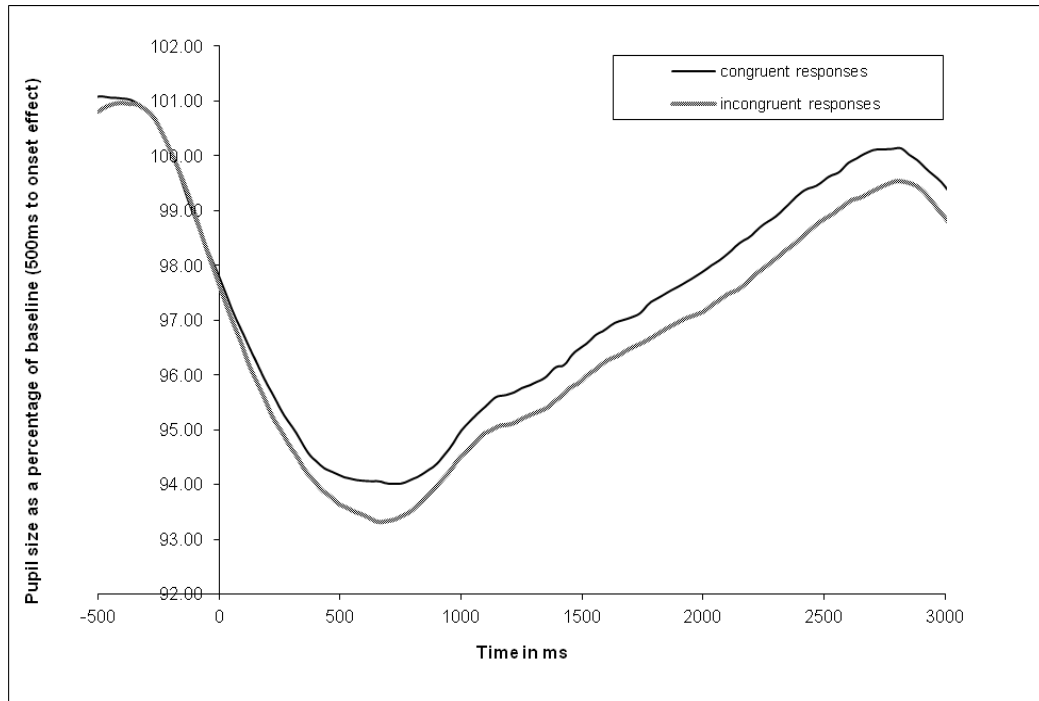


Figure 4. Relative pupil sizes for congruent and incongruent responses to baseline, stimulus-locked.

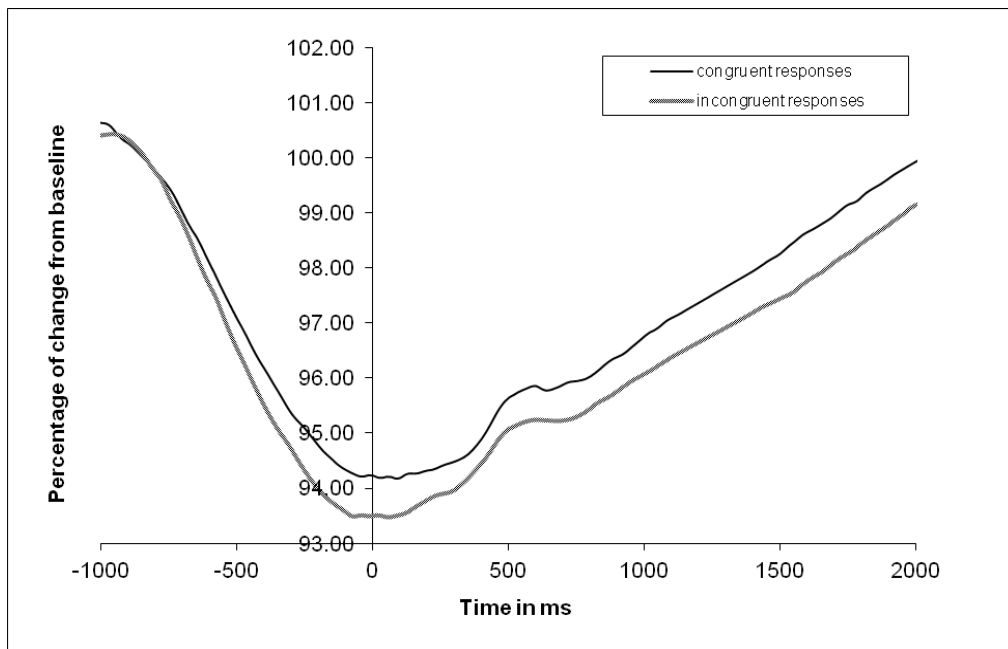


Figure 5. Relative pupil sizes for congruent and incongruent responses to baseline, response locked.

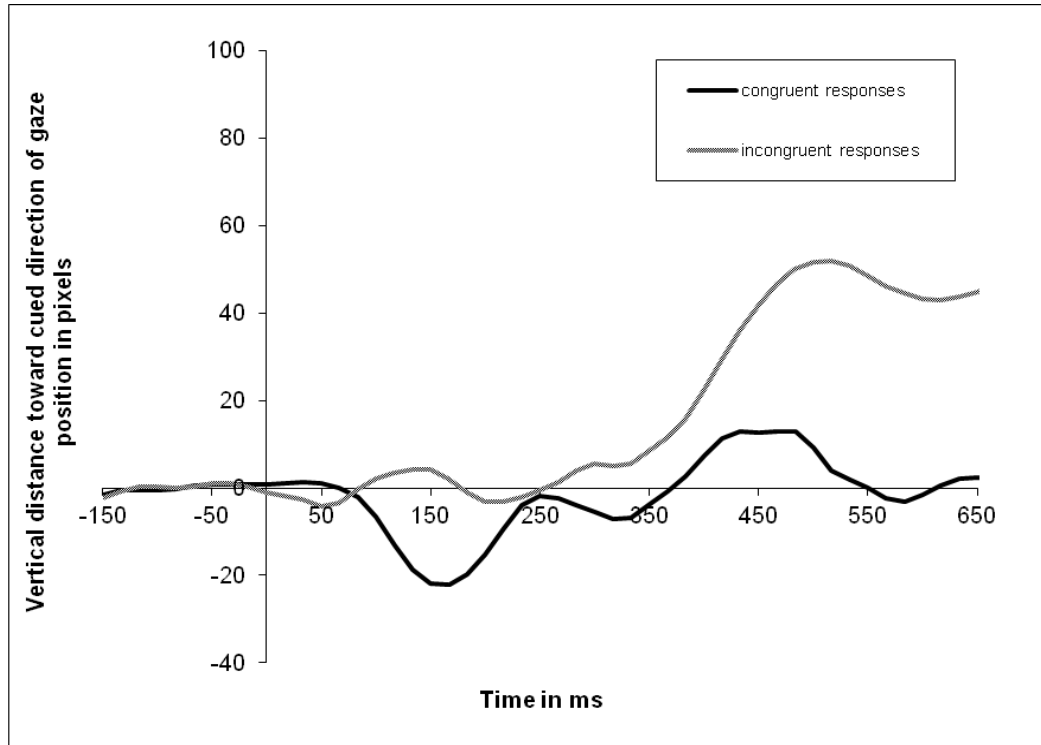


Figure 6. Vertical distance from midline toward cued direction of gaze position in pixels for congruent and incongruent responses. Time is 0, is the moment the effect starts.

AGE GROUP SCORES	Percentage of completed acquisition trials	Percentage of left responses	RT in ms	RT left	RT right	RT up	RT down
9-month-olds	92.9 (11)	42.8 (38)	1007 (81)	999 (107)	1046 (90)	1008 (116)	1037 (83)
12-month-olds	92.0 (14)	38.5 (40)	1023 (83)	1089 (167)	982 (73)	982 (78)	1089 (164)

Table 1. Mean scores of acquisition phase (standard deviation in brackets).

AGEGROUP SCORES	Percentage completed test trials	Percentage left responses	Percentage congruent responses	ITI ms	RT ms	RT Congruent ms	RT Incongruent ms
9-month-olds	93.5 (11)	43.2 (38)	47.5 (8)	1637 (323)	431 (90)	440 (104)	424 (91)
12-month- olds	96.5 (10)	29.0 (21)	49.3 (7)	1563 (393)	447 (74)	468 (83)	425 (83)

*Table 2.* Mean frequency and RT scores of test phase (standard deviation in brackets).