Measuring the noticing of an unexpected event in Magical Garden with a Teachable Agent using Eye-Tracking

Mäta upptäckten av något oväntat i Magical Garden tillsammans med en Teachable Agent via Eye-Tracking

Ludvig Londos

Handledare / Supervisors

Agneta Gulz
Richard Andersson

KOGM20

2015-06-21

Masteruppsats i kognitionsvetenskap
Avdelningen för kognitionsvetenskap
Filosofiska institutionen
Lunds universitet

Master's thesis (2 years) in Cognitive Science
Lund University Cognitive Science
Department of Philosophy
Lund University

Measuring the noticing of an unexpected event in Magical Garden with a Teachable Agent using Eye-Tracking

Ludvig Londos
ludvig.londos@gmail.com

Scientific views on what children are capable of have been revised through history again and again, usually when new methods of studying children's capabilities are presented. What has often been concluded is that children are capable of more than what was previously thought.

New technology has introduced a genre of educational games which utilize the captivating power of computer games which have shown a positive effect on learning and motivation. In this study, the educational game Magical Garden was used as a platform to train, teach, and test number sense. The pedagogical instrument Teachable Agent (TA) is a part of Magical Garden's design which utilizes the protégé effect. A new method of measuring number sense, detecting an "unexpected event" by attending to it, is proposed and tested. The unexpected event was a tree elevator malfunction. The purpose of the unexpected event was to create a task where only the children who were attentive and knew which branch the elevator would go to would react to and detect the unexpected event. A model of detection of the unexpected event, looking back at the correct branch after the elevator passed the correct one, was proposed. Eye-tracking was used as the method of capturing detections of the unexpected event, as well as measuring the interaction between the children and the TA during the unexpected event.

In this study, 42 preschoolers participated. The results show that children attend the TA significantly more when the TA was in charge of the decisions in the game. This indicates that preschoolers understand that the TA was in charge. The model of detection used in this study was not comprehensive. However, detecting an unexpected event could still be a promising method of measuring number sense. Therefore, future research could utilize this method to unveil more exciting capabilities of children with a more inclusive model of detection.

1 Introduction

How would the world look like if you do not understand numbers, relations, or quantities? This is the reality for many young children. These difficulties are a very hard thing to even imagine because numbers, relations, and quantities are a fundamental part of human cognition and thinking. Growing up, children train and develop a feeling for numbers, relations, and quantities. This feeling or knowledge is often referred to as *number sense*. Unfortunately, for some children developing a number sense takes longer time. Without intervention, this could lead to failing early math (Griffin, Case & Siegler, 1994) and developing learning disabilities (Gersten & Chard 1999).

With new technology and the popularity of computer games, a genre of educational games for mathematics has emerged. Utilizing the motivational and captivating power of computer games, educational games for mathematics have shown a positive effect on learning and motivation (Schwartz & Martin, 2004; Moreno & Mayer, 2005).

In this thesis, an *unexpected event* is introduced to the educational game Magical Garden, and thought to act as a method of testing number sense capabilities in preschoolers. The unexpected event was constructed in a way that only children with sufficient number sense would notice it. This method of testing number sense has not been applied before this study. Eye-tracking technology was used as the method of choice to confirm if a child had noticed the unexpected event. In this study, a proposed model of noticing will be presented and tested.

Magical Garden utilizes a *Teachable Agent* (TA) as a pedagogical instrument to motivate children and elicitate the *protégé effect* (Chase, Chin, Oppezzo & Schwartz, 2009). Question has been raised whether preschoolers can benefit from a TA due to the fact that most preschoolers do not have a fully developed *theory of mind* (ToM) (Graham & Perner, 2001). Axelsson, Anderberg and Haake (2013) have researched preschooler's relation to a TA. Their results indicated that preschoolers can interact socially with a TA, and they requested further research in this relationship. In this thesis, I will investigate the interaction between the TA and the children during the unexpected event, and hopefully add to the growing body of knowledge regarding the interaction between TAs and preschoolers.

The structure of the thesis

Firstly, a report on previous research from the fields of number sense, learning, and *TAs* is presented. This will set the stage for the thesis. Secondly, noticing something unexpected is proposed as a way of examining children's level of number sense. Then, the method of eye-tracking technology to capture the act of noticing the unexpected is proposed. Then, an "unexpected event" is introduced as an experimental manipulation, and a model for noticing is proposed. Next, there will be a short overview of the experiment. Following this, there will be a presentation of the research questions for this thesis. In the method section, material, participant information, ethics, the experiment procedure, measurements, and data analysis is covered. After that, the results are presented and discussed. Finally, there is the conclusion and ideas for future research.

Number sense and learning

In the following section, an attempt will be made to explain number sense. What is number sense? Possessing an understanding of amounts, proportions, numbers, and the language of numbers and arithmetic is referred to as having a "number sense" (Dehaene, 1997). An important skill when learning number sense is to be able focus attention towards what is important. Gelman (1990) emphasizes the importance of attending the right things in order to gather more knowledge and build up the cognitive skill set. Number sense is the ability to understand numbers in a broader sense, not only as numbers but also as a quantity, something you can manipulate, something flexible, and a way of looking at the world (Griffin, 2004). Number sense has been described as something easy to recognize, but hard to fully grasp (Case, 1998; Griffin, 2004). If a child has a good number sense, they are able to fully comprehend that quantities in the world correspond to mathematical expressions and numbers is a way of representing these quantities. Also, number sense can be understood as an understanding of that e.g. five could be represented in a lot of different ways. For instance, that "five" is the same as 5 and | | | | |, 2 + 3 etc. A crucial step in developing one's number sense is to gain an understanding of numerical magnitude, e.g. the understanding that 9 is greater than 3 (Dehaene, 1997; Griffin, 2004). A mental number line from left to right is a way of representing how big a specific number is in relation to other numbers. Smaller numbers are placed to the left and bigger to the right (Dehaene, 1997). This is crucial when learning basic arithmetic because an understanding that numbers are ordered in systems gives the child an ability to recognize patterns, which is a step towards fully understanding that specific quantities can also represent general amounts.

Number sense starts to develop during childhood. Gersten and Chard (1999) argue that number sense should be viewed as a skill that could be trained and taught as opposed to something innate and static. Additionally, Lipton and Spelke (2003) describes in their research that infants already have a basic understanding of quantity and the ability to discriminate between different quantities. Moreover, the precision of discriminating between quantities increases over the child's development. It is common that parents begin to teach their children about numbers early on through verbal communication. The parents often falsely conclude that the child knows numbers when she can recite them (Carey, 2004). According to Carey (2004), often children do not understand what parents tell them due to the fact that they cannot attach the numerical words to concepts of numbers. The first things children learn about numbers is their order, and to recite them correctly as "1, 2, 3, 4" not "1, 3, 4, 2". This is called a *count list*. Besides the numerical order, the count lists do not carry much other semantical content (Carey, 2004). Carey (2004) hypothesized that children learn "one" and its semantical content in the same way they learn the semantical content of a single determiner such as "a" in spoken language.

Another important principle of number sense is *cardinality*. Cardinality is the number of elements in a set. In this case, the set is the word "one" which refers to one and only one element, and the set "two" refers to two and only two elements (Wagner & Johnson, 2011). Giving semantical content to the unknown number "three" is done by taking its order in consideration. For instance, understanding that "three" is after "two", and that "three is one more than two". Each step and each new number takes time to learn. If the child has not learned the cardinality of a number, it is represented as "many" because it is more than any known number to them (Carey, 2004). Therefore, children could be at different levels of number sense depending on their development. Gersten and Chard (1999) argue that the process of learning the concepts of numbers is an automatic gateway into mathematics, and the key to solving basic arithmetics. All of the concepts presented above, order, cardinality, numerical magnitude, and semantic abstraction, constitute the foundation of number sense. In the section above I described the foundation and some theoretical concepts of number sense. Based on these concepts, numerous studies have presented different exercises as a way of operationalizing and measuring number sense e.g. counting, number identification, quantity discrimination, and missing number (Geary, 2004; Clark & Shinn, 2004; Case & Okamoto, 1996; Chard, Clarke, Baker, Otterstedt, Braun & Katz, 2004). All of the exercises have roots in the foundation of number sense. Geary (2004) have measured number sense with counting exercises, such as; count from 1 to 20, count from a given number 3 or 6 steps forward, and count by 2, 5 or 10. Geary (2004) argued that these counting exercises were a good way of measuring number sense. Clarke and Shinn (2004) conducted an experiment testing number sense based on exercises consisting of number identification with exercises, where children identified a given number between 0-20, quantity discrimination, mental number line with exercises where children would select which one of two numbers is the bigger one, and missing number by filling in the blank number missing in a sequence. Case and Okamoto (1996) created the number knowledge test which is regarded to be an exhaustive test for assessing children's conceptual knowledge of numbers. It consisted of a number of exercises, similar to the ones presented above, on three levels with increasing difficulty. Chard et al. (2005) used the number knowledge test as the baseline measure of number sense when they evaluated other operationalized measures of number sense in order to create a screening measure to detect children with lacking number sense. It seems possible to construct an operationalized measure for number sense based on the basic concepts of number sense. There is no one way of doing it.

What are the consequences if a child does not develop their number sense? Children who do not add enough 'meat', concepts of number sense, to their conceptual skeleton are very likely to fail early math in school (Griffin, Case, & Siegler, 1994) and develop learning disabilities when they grow older (Gersten & Chard, 1999). Chard et al. (2005) successfully used the combination of the operationalized exercises, presented above, to screen and sieve out preschoolers and first graders with early difficulties in mathematics.

How do you catch children's attention and promote learning? One intervention that has been tried was to systematically drill children with math facts, in order to automatize their thinking procedure and make them more interested in math (Pellegrino & Goldman, 1987). This type of intervention worked poorly and had the opposite effect because it was both unpleasant for the children which caused them to lose interest in math (Gersten & Chard, 1999). The importance of play as a pedagogical tool for learning and to get children motivated in their acquisition for new abilities has been common knowledge for many years. Play and playing is by Geary (1995) considered a primary way of adding meat to the conceptual skeleton in form of new abilities and knowledge. Also, different types of games and play involving numbers have been found all over the world. Griffin, Case, and Siegler (1994) suggested board games as a good way of raising mathematical awareness and early number sense for children with low socioeconomic status who do not get the informal preparation of mathematics through parents, siblings and other social interaction. Gee (2003) describes that when the play and fun in learning disappears, motivation and educational opportunities plummets.

In this introduction to number sense, basic concepts of number sense has been unveiled, as well as different ways to operationalize exercises to measure number sense. The tight connection between number sense and mathematics has been displayed, and the problems which will follow with a lack of number sense have been emphasized. In the following section Magical Garden will be introduced as a way of incorporating play and games in the educational process of learning number sense.

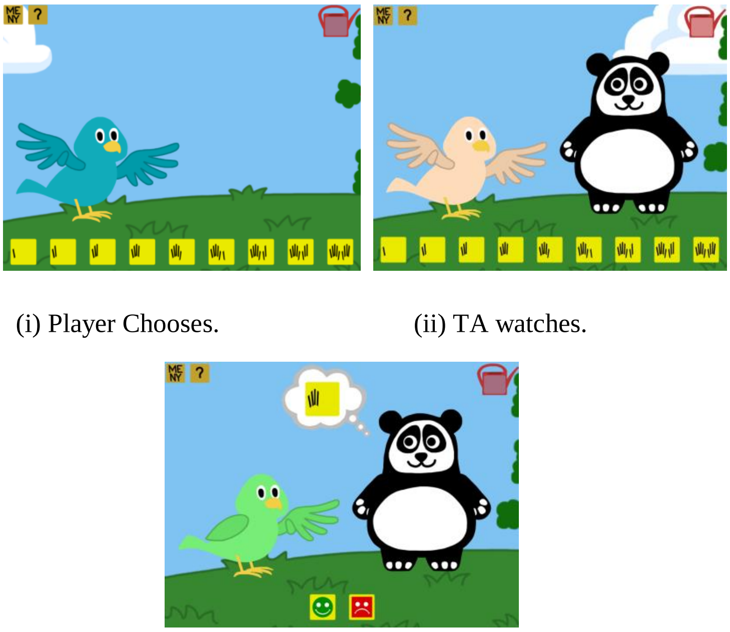
Magical Garden

In this and the two coming sections I will present the educational game Magical Garden, and argue that it could act as an operationalized exercise and be used to measure number sense capabilities. After this, I delve into the concept of a TA and discuss its addition to the Magical Garden thoroughly.

In this study, preschoolers ages 4-6 will play the educational game Magical Garden which is designed for that age group to learn and practice number sense and basic math skills. One of the advantages of Magical Garden and other educational games is that it is not obvious for the children that they are learning mathematics by playing the game. Magical Garden is designed by Anderberg, Gulz, Haake, and Husain, Lund University. In the version of Magical Garden (1:2014-11-26) used in this study, there are four different mini-games. The first mini-game is “Bird Hero”, where the mission is to help the baby birds to take a tree-elevator to their parents (see figure 1). The second mini-game is “Bee Flight”, where the mission is to help a bumblebee find nectar from the right flower. The third mini-game is “Lizard”, where the mission is to help a chameleon shoot its tongue and hit ants on a tree. The fourth mini-game is “Balloon”, where the mission is to help a woodlouse go on a treasure hunt inside different caves on a cliff. In all games, the child has to take the reins and press the correct button in order to help the baby bird, bumblebee, lizard, or woodlouse. The range of numbers which are trained in the mini-games are either 1-4 or 1-9. Griffin (2004) emphasizes the importance of using representations of numbers in the same way they are represented in our culture e.g. an amount of objects, a specific pattern of dots, a mark on a line, and a value of a scale. The buttons in Magical Garden are represented either as dots, lines, fingers, or numbers. To train familiarity with number representations commonly used in our culture is important because a more developed number sense makes it easier to understand the meaning of numbers both in and outside of school (Griffin, 2004).

In Magical Garden, the goal is to collect water droplets. The player receives water droplets when a mini-game is completed. The water is used to water the garden. Each time the garden is watered, magical plants and candy grow. Gulz, Haake, and Silverberg (2011) discuss the value of an *off-task*, such as social conversation with the TA or something not related to the main task in between playing the main task of an educational game. The purpose of an off-task is to encourage learning, motivate, and give the students a positive experience of the game. The watering in Magical Garden fills a similar role as an off-task, having nothing to do with the main task. This is meant to be a fun activity for the children, a goal to keep them engaged, and make them willing to play more.

Each of the mini-games follow the same structure. There are three different modes of gameplay (i) the player chooses which button to press, (ii) the TA watches the player choose which button to press, or (iii) the TA thinks of a button (displaying a proposal for which button might be correct) in a thought-bubble, and asks the player if the button it thinks about is correct. If the player thinks the TAs proposal is correct, the player presses the green button with a “smiling face”. Otherwise, the player presses the button with a red “frowning face”. If the player presses the red button, they have to show the TA which the correct answer is by pressing the correct button (see figure 1). Each mode is played three



(i) Player Chooses. (ii) TA watches.

(iii) TA thinks and the player agrees or not agrees.

Figure 1. The three pictures displays the three different types of gameplay (i, ii, and iii) in Magical Garden.

times in each mini-game. Depending on the performance in previous mini-games, the difficultly level goes up or down, as well as the TAs competence. In the later stages of the game, early mathematical concepts, such as addition and subtraction, are added to the gameplay.

Below follows a detailed description of the characteristics of the mini-game Bird Hero, which will be the focus of this study. In this mini-game, the task is to help a baby bird to its parent at the right branch of a tree by sending it up the tree-elevator made of a rope and bucket. The baby bird shows the player which branch it wants to go to with its feathers. The task is to press the yellow button with the same number of fingers as the feathers the baby bird shows (see figure 1:i). The baby bird jumps into the tree-elevator and the elevator displays, with a symbolic number, the current branch at the front of the bucket. When the elevator stops at the correct branch, the bird jumps off, and is welcomed by its parent as they cheer and tweet. If an incorrect button was pressed, feedback is given to the player that this is not where the baby bird lives by saying “this is not my parent, I live further up (or down)”.The baby bird returns to ground floor and the player may try again. After helping three baby birds, the TA (Panda, Mouse, or Hedgehog depending on which character the player has chosen to play with), enters the game and says s/he wants to help the baby birds as well. At first, the TA watches the player continue helping baby birds, game mode (ii). When three more birds have been helped, the TA says s/he wants to try her- or himself and suggests an answer through a thought-bubble, game mode (iii) (see figure 1:ii and 1:iii). Three more birds are helped this way, and when all the birds have been saved, the TA and the birds cheer saying “thanks for all the help”. Afterwards, the player receives water droplets which now can be used to water the magical garden.

Is it then possible to evaluate children’s number sense from their performance in Magical Garden? There are three different types of number representations in Bird Hero. As mentioned above, the elevator buttons represent the amount of objects. The baby bird’s feathers also represent an amount. Finally, the tree is represented as a scale where the current branch is the value of the scale (see figure 2). The combination of these three representations creates an operationalized exercise. I would argue that the exercise trains and tests number sense because it incorporates key elements such as number identification, the abstraction that quantities can be represented in different ways, and the order of numbers both on the elevator buttons and the order of branches of the tree. Because the exercise is build up by elements of training number sense capabilities, failure to complete an exercise would be an indication of not fully understanding the concepts or an insufficient level of number sense. Therefore, I would argue that Magical Garden is a well suited platform to train number sense because it utilizes the key concepts and have operationalized them in exercises which can be measured.

Teachable-Agents

Magical Garden is a TA based game. As previously mentioned the TA is a part of the (ii) and (iii) game mode (see figure 1). A TA is a computer application represented as a digital student. TAs were introduced to the paradigm of intelligent learning environments by Biswas, Katzlberger, Bransford and Schwartz (2001) in an effort to enhance student’s motivation and learning. A TA is trained and taught by the student and cannot learn in any other way. Therefore, the TA is a reflection of what it has been taught. Also, the TA is able to display what it has been taught. The idea behind this is that it would to motivate and give the student an opportunity to reflect upon it (Biswas et al., 2001). In this process of tutoring the TA, the student acts as a teacher and learns for herself while teaching the TA. Educational literature shows that during the act of tutoring, the tutor as well as the tutee gain understanding and benefits. The tutoring in a form of social learning (Chi, Siler, Jeong, Yamauchi & Hausmann, 2001). This idea of social learning is captured by the ”learning by teaching” paradigm and it is supported by educational literature (Biswas, Leelawong, Schwartz & Vye, 2005; Schwartz, Blair, Biswas, Leelawong & Davis 2007; Biswas et al., 2001; Chase et al., 2009; Gulz, Haake & Silvervarg, 2011; Pareto, Arvemo, Dahl, Haake & Gulz 2011; Pareto, Haake, Lindström, Sjödén & Gulz, 2012; Lindström, Gulz, Haake & Sjödén, 2011). The addition of TA to learning games encourages students to care for someone else’s learning and by doing so increases their own learning. This is known as the *protégé effect* (Chase et al., 2009). The idea behind the design of Magical Garden is for the player to have a TA as a friend and to elicit the *protégé effect.*

Previous research on TAs has mainly been carried out on undergraduate students (Schwartz et al., 2007; Biswas et al., 2001) and 8-14 year old children (Chase et al., 2009; Gulz et al., 2011; Pareto et al., 2011; Pareto et al., 2012; Lindström et al., 2011). Other than Axelsson et al. (2013) preschoolers, age 4-6, interaction with a TA have not been extensively tested. One question that arises is whether preschoolers can even benefit from the pedagogical help a TA is offering. One argument against the idea, that a TA would be a strong pedagogical aid for preschoolers, is that preschoolers do not yet have fully developed ToM. ToM is the ability recognize to that other people can have a different understanding, feelings, and knowledge of things than your understanding, feelings, and knowledge of the same things (Premack & Woodruff, 1978). Premack and Woodruff (1978) described ToM as the most important factor to understand and to act socially. To teach someone, one needs to be able to recognize that the one who is being taught does not know everything you know yourself. Without a developed ToM, the concept of teaching seems thus diluted and meaningless. ToM has traditionally been measured by a false belief test (Graham & Perner, 2001). To pass a false belief test the child must be able to identify that someone else can have a false belief, i.e. believe that something is true which is not true. The development of ToM comes stepwise and is usually fully developed by the age of six (Graham & Perner, 2001). A premise for interacting with a TA is to be able to act in a social learning environment. Does this mean that preschoolers cannot benefit from a TA? Clements and Perner (1994) reported that even though some children did not pass the false belief test they seemed to have an implicit knowledge of the false belief. Their implicit knowledge of the false belief was measured by anticipating eye movements towards the correct answer. Axelsson et al. (2013) used Magical Garden, as the educational game of choice, in their study which indicated that the game was engaging both with and without a TA. Another result was that preschoolers were able to attend to the TA and not be disturbed by its presences. This led them to conclude that it is possible for preschoolers to interact socially with a TA. Axelsson et al. (2013) emphasized the importance of further research on preschoolers. If the children acknowledge the TA as another player in the game, they would reveal ToM capabilities and be able to interact socially with the TA and thus benefit from the pedagogical help the TA is offering. In this study, in order to investigate if children acknowledge the TA as another player in the game a fourth game mode had to be created. In game mode (iiii), the TA in charge of the decision making. When the TA is in charge, he thinks of a button and then acts by himself pushing the elevator button. By observing how the children act, when the TA is in charge of the decision making, if they attended to, interacts and acknowledge the TA as another player in the game, I would argue that the child has ToM capabilities and can benefit from the TA as a pedagogical aid.

These introducing sections have started to set the stage for the study, by introducing key concepts and questions. What is important when learning and developing number sense? As previously mentioned, key components of learning are attending to the right things, motivation, play, and learning the concepts and that they can be represented in various forms. Also, number sense can be operationalized and measured by using exercises. With operationalized exercises it is possible to sieve out children with a lacking number sense. As previously mentioned educational games seem to be able to incorporate many of the important aspects for developing number sense as well as having a motivational and captivating nature. Another advantage of computer games and virtual environments are, that they can be scripted and programmed to include experimental manipulations. These aspects give grounds to use Magical Garden as a platform to train, evaluate and to test preschooler’s number sense.

In the next section, I will discuss and argue for why noticing an unexpected event could be indicative of developed number sense. Thereafter, I will discuss and describe how the noticing could be measured and captured.

Noticing

How does one notice anything at all? Schmidt (1990) describes that noticing is the basic indication of being aware of something. Noticing is also a private experience. One way of describing it is if a verbal report could have been given. One must actively attend to something in order to be able to notice it. Thus, there is a distinction between perceiving information and noticing information. To be aware and act upon the awareness, information must be actively processed and manipulated by the brain (Schmidt, 1990). Corbetta and Shulman (2002) describe two attentional systems, with functional differences, that work together to guide the attention in a visual environment in order to gather information.

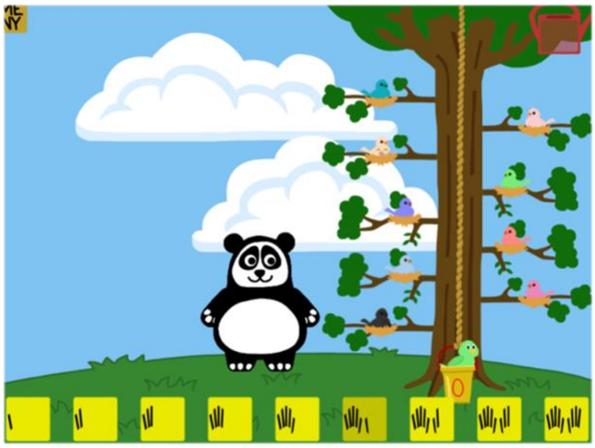


Figure 2. Displays the mini-game Bird Hero with all of its components: the tree with nine different branches, the baby bird, the elevator, the buttons, and the TA Panders.

Attention is either driven by stimulus (bottom-up) or goal-directed (top-down) attention. Bottom-up driven attention depends on exogenous stimuli that are salient, or unexpected stimuli, such as the change of color, shape, movement, contrast, or luminance. The top-down system depends on endogenous or cognitive stimuli such as knowledge, expectation, and goals. In most cases, attention is driven by a combination of bottom-up and top-down stimulus (Corbetta & Shulman, 2002).

Could a detection of something unexpected, in a math context, then be indicative of developed number sense? If an unexpected event, in a math context, is attended via bottom-up stimuli, it is not sound to give credit to number sense. In order to conclude that number sense is a contributing factor in noticing, it must be evoked by top-down attention; be driven by previous knowledge, expectations, and mismatch with the intended goal. It seems possible to design an unexpected event in an exercise that allows a detection, noticing, to be made by top-down control. Furthermore, the exercise could be operationalized in a way to test developing number sense. A detection of the unexpected event would then be noticing a mismatch in the visual environment by top-down control. Could *working memory* (WM) be a contributing factor in noticing or not noticing an unexpected event? WM could be defined as when information and knowledge are “kept in mind” (readily having access to it), and use it without relying on sensory information (Corbetta & Shulman, 2002). In Magical Garden, to keep in mind which branch the baby bird indicated with its feathers is a WM task, but using the information kept in mind with WM I would argue to be credited to number sense. The child has to have the sufficient level of number sense, know that the feathers represent the branches, and use the information kept in mind by the WM in order to notice an unexpected event.

How could a detection be manifested and how could it be measured? A possible way to manifest noticing would be via a concurrent verbal report or think aloud protocol. According to Schmidt (1990), verbal reports are used to both confirm and dismiss if something was noticed or not. Although this type of verbal report is a very strong indication of detection, it might be too narrow in the sense that someone might not spontaneously comment out loud. Another reason may be that the participant is shy or uncomfortable in talking to an experimenter. Another possible way to manifest a detection would be to let the participants press a button, ring a bell, or move something physically in order to indicate that they have noticed the unexpected event. This would also be a clear indication that could easily be measured. An issue with this as an indicator for noticing is that the unexpected event has to be disclosed in order for the participant to know when and why they are supposed to "press the button". Another issue, when working with children, is that they have to remember to follow the instructions of "pressing the button" when noticing the manipulation. Both of these presented ways to manifest noticing are problematic because they do not catch all of the noticing of the unexpected event. Smith (2012) presented noticing manifested as increased visual attention. The method used to measure increased visual attention was eye-tracking. The coupling of eye movements and visual attention is based on the fact that eye movements are not only directed by the visual input (bottom-up) but also controlled by top-down constraints based on previous knowledge (Henderson, 2003). Deuble and Schinerder (1996) presented evidence for the hypothesis of a coupling between eye movements and visual attention when directed towards an object. They proposed a single attentional mechanism model for both perceptual processing and recognition. Corbetta et al. (1998) showed that there is an anatomical overlap in active regional networks when processing attention and eye movements. This is consistent with their hypothesis that attention and eye movements are tightly connected at neural level.

The close connection between top-down control and the control of eye movements provides grounds for investigating detection of an unexpected event with eye-tracking technology as an online-measurement for the detection. Before this study's proposed model of detection is introduced, a short introduction to eye-tracking is presented.

Eye-tracking

Visual information from the world is collected by our eyes. The retina in the eye is built up by rods and cones, which record the incoming light and convert it to electric signals. These signals are transported by the optic nerve to the visual cortex in the brain for processing. To be able to gather this input, the eyes move and orient themselves in order to face a small part of the retina, the fovea, towards objects in world. The eyes move by making saccades. A saccade is a very fast movement of the eye, normally lastning 30–40 ms. Depending on its amplitude it can be faster or slower. The eye makes on average three saccades per second (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka & Van de Weijer, 2011). During a saccade, the visual input is obtained but limited and impaired because the visual information gets smeared over the retina (due to the rapid movement of the saccade), and the brain actively stops processing the visual information (Holmqvist et al., 2011). A fixation is when the eye is "still"[1] and the center of gaze is directed towards an object. During a fixation, visual input is collected and actively processed. A fixation differs in length and could be up to several seconds long. Holmqvist et al. (2011) stated that when a fixation is measured, what is actually measured is the attention to that location which the eye was fixated on.

An eye-tracker tracks the movements of the eye with a camera by recording the pupil and corneal reflection. From these recordings, the raw data (the movements of the eye) are categorized as fixations, saccades, and blinks by the eye-tracker (Holmqvist et al., 2011). Gaze estimation is where on a *stimulus* (the item that is being observed) the fixations and saccades are located. The gaze estimation is calculated by taking into account the distance between the eye-tracker and the stimulus, as well as the spatial relations between the pupil and the corneal reflections center (Holmqvist et al., 2011). Using eye-tracking to capture underlying cognitive processes is supported by the close coupling between visual attention and eye movements (Deuble & Schinerder 1996; Corbetta, Akbudak, Conturo, Snyder, Ollinger, Drury & Shulman, 1998; Holmqvist et al. 2011). A lot of measures can be calculated from the recorded eye movements, such as the number of fixations, dwell time, attentional shifts, and fixations on an *Area of Interest* (AOI). Different eye-tracking systems differ in functionality; therefore, different eye-trackers are used depending on the task and stimulus. The three most common systems are tower-mounted, head-mounted, and remote eye-tracker (Holmqvist et al., 2011). The tower-mounted systems often fixate the head with a chinrest, as well as presenting the stimulus on a screen in front of the participants. Hence, the tower-mounted system can have a higher sampling rate than other systems, because it can assume that the eye is captured by the camera instead of utilizing eye-detection algorithms. Tower-mounted systems are used when the objective is to capture small and precise eye movements. If the task is to measure eye movements while walking around in a supermarket, the head-mounted systems are preferred. The distance to the stimulus is not the same when one moves around an environment such as a supermarket. Thus, the head-mounted system struggles with defining the stimulus. Therefore, a head-mounted system cannot calculate gaze estimation the same way as tower-mounted systems. Often, the recorded eye movements have to be manually coded on a gaze map frame by frame which is a tedious process. The third system is the remote eye-tracker. Its sampling rates are on average higher than the mobile head-mounted systems and lower than the immobile tower-mounted systems. It is attached to a monitor; therefore, it could be used out in the field as well as in a laboratory setting. In order to record the eye movements, the participant has to be within a distance of 70 cm of the eye-tracker. The participant has a possibility to move their head a little bit without losing connection to the eye-tracker (Holmqvist et al., 2011). An advantage of the remote eye-tracker is that it is not as intrusive and apparent to the participant that the eyes are being recorded, as if a helmet or a chinrest is used.

Eye-tracking also has a diagnostic role because it provides an objective and quantitative measure of the eye movements that reveal the underlying cognitive processes (Duchowski, 2002). As a tool, eye-tracking can be used in a

[1] During a fixation, the eye is not really still. A fixation consists of three distinct micro-movements: tremor, microsaccades and drifts (Holmqvist et al., 2011).

number of different domains as it captures the attentional window the participant has on a given stimuli. A lot of eye-tracking research has been devoted to investigating reading exercises. In contrast, little progress has been made in eye movement research involving mathematical tasks. One of the few previous studies on eye-tracking with mathematical tasks is Green, Lemaire and Dufau (2007). They conducted a study on adults' eye movements when adding together three-digit numbers. They showed that the adults used "strategic" eye movements when adding the numbers. Another study is Schneider et al (2008) who presented a validation of eye movements as a measurement for developing number sense in children. They examined the exercise "mental number line" and to what degree eye movements, in this case fixational accuracy, were related to competence. They also looked at the spread of fixations, when the participants were trying to come up with a solution to the task. They found, inter alia, that fixations increased with competence. Furthermore, they concluded that using eye-tracking as a measurement for developing number sense was both valid and useful (Schneider et al., 2008). Hitherto, no research has investigated eye movements during dynamic mathematical exercises such as during an educational game. Using eye-tracking to capture the noticing of an unexpected event in Magical Garden would then be pioneering work.

In the next section, I will introduce the experimental manipulation, an unexpected event in Magical Garden, of this study.

Experimental manipulation "the unexpected event"

What is "the unexpected event" in Magical Garden, and could noticing it measure children's level of number sense? I would argue that noticing an unexpected event has to contain three crucial elements in order to be indicative of a developed number sense. Firstly, it has to only evoke a response from children who fully understand the exercise (have a developed number sense). Secondly, it has to be measurable. Thirdly, the act of noticing the unexpected event has to have a meaningful and logical explanation tied to the measured behavior. In this study, the unexpected event is incorporated in the mini-game Bird Hero. The unexpected event is defined as:

The elevator continues past the intended branch and gets stuck in the tree top[2]

The player is given the opportunity to detect the unexpected event before it reaches the tree top. During regular gameplay the elevator always moves, as expected, to the level chosen by the player. An expectation is created by the top-down system that the elevator will move to the chosen level. The expectation depends on previous knowledge, WM, and number sense. In this situation, previous knowledge include the number the baby bird showed with its feathers, the lift button's number, and corresponding branch of the tree. Therefore, I would argue that the unexpected event is only unexpected for the players who have an expectation (the elevator should move to the branch corresponding to the pressed elevator button). There are no bottom-up salient or unexpected stimuli exposing the unexpected event before the elevator gets stuck in the tree top. The only thing that changes in the scene is that the elevator keeps ascending. If a child does not know that the elevator was supposed to stop at a specific branch, you could assume that the child thinks that "The elevator is still moving towards the chosen branch". The reality is that it has already passed the chosen branch. This distinction separates the children that have a sufficiently developed number sense from the ones that do not. All of the children will likely understand that something went wrong when the elevator gets stuck in the tree top. When the elevator gets stuck in the tree top, marking the end of the unexpected event, a service bird enters the screen, coming to the rescue. The service bird says "Oh dear, oh dear, the elevator is broken! Don't worry, I will fix it [the service bird repairs the elevator]. You're welcome!". When the elevator is repaired, the service bird leaves the screen, and the elevator starts to descend to the correct branch. The reasons for adding the service bird was primarily as a precaution so that the children would not get upset or wonder if the game was broken. Another reason was to explicitly say that something went wrong.

What type of eye movements would then be indicative of noticing? As previously mentioned, the method of eye-tracking provides a number of different possible measurements. In the next section, a novel model of detection will be presented. Eye-tracking will be the method of capturing the eye movements. In addition, I will further argue why noticing the unexpected event would be a good method of measuring number sense.

Model of detection

Where will the children look when they notice the manipulation? As previously proposed, noticing could be manifested as increased visual attention towards a specific area. This increased visual attention towards an area could be measured with *AOI hits*. Holmqvist et al. (p. 189, 2011) defined an AOI hit as: "AOI hit, which states for a raw sample or a fixation that its coordinate value is inside the AOI". A detection of an unexpected event could then be measure as fixation on an AOI. This AOI has to have a logical and meaningful explanation why looking at it should be a detection of the unexpected event. In this study, the model that is proposed for a detection of the unexpected event is a *look back*. A look back is defined as:

If the child looked back at the correct branch after the elevator passed the correct branch.

The reasoning behind this proposal was that attention would be given towards the correct branch during an unexpected event as a way of indicating that they had noticed the mismatch. Only top-down attention would direct the children's eye movements toward the correct branch. Children who do not have understanding of which branch was correct, due to

[2] Only if the correct level was chosen by the player. Otherwise, the elevator will go to the incorrectly chosen level and the baby bird will say it was the wrong level prompting the player to try again.

either a lack of working memory or number sense, will not notice that something is wrong with the elevator, and have no reason to perform look back. The most likely eye movement behavior of the children, if no detection is made, would be to follow the elevator with their eyes and wait for it to stop. This is because the movement of the elevator would be the most bottom-up salient object in the scene and draw the most attention (Corbetta, Shulman, 2002). According to Henderson (2003), even though the movement of the elevator is a strong salient feature, visually salient objects are less attended to during an active task on a meaningful scene. In this case, the unexpected event is a meaningful scene to the child who knows which branch was the correct one. Therefore, noticing the unexpected event would be considered an active task which would then negate the salient feature of the moving elevator (Henderson, 2003). Having an AOI hit as the measure for attending the AOI seems sound because the children who noticed the unexpected event only use their top-down attention.

The proposed model of detection is novel and although the reasoning behind the model seems sound, it has yet to be tested. By testing the model, it will be determined whether it is a good way of measuring number sense. It certainly could be the case that the proposed model is not the best for testing noticing of an unexpected event. Therefore, in addition to the proposed model, other eye movements, such as looking in anticipation and looking at the buttons, will be collected and analyzed. In this study, the proposed model will be tested and below follows an overview of the experiment.

Overview of the experiment

In order to evaluate the proposed model, as well as, study if preschoolers attend to the TA an experiment was conducted. It consisted of two parts; 1) familiarization with Magical Garden and collecting longitudinal data and 2) an eye-tracking experiment. First, the children played Magical Garden individually on tablets (iPads) at their preschools during dedicated sessions, a number of times each week for three weeks. Every mini-game played on the tablets was logged and the logged data was stored on an online server. After the familiarization phase, an eye-tracking experiment was conducted at the preschools. A new version of Magical Garden was designed and created for the eye-tracking experiment, which included the unexpected event. In the experiment, the child and the TA alternated between being in charge of the decision to push the button to answer. The exercises also alternated between containing an unexpected event and not containing an unexpected event (control condition). The children played at least five trials and a maximum of ten trials. A remote eye-tracker was attached to a monitor and the game was played on the monitor with a mouse. After the experiment was done, each child was given a diploma as a reward for completing the experiment. Besides the proposed model, other eye movement measures was collected and analyzed such as looking at the button/thought-bubble, looking in anticipation, and looking at the TA. A full review concerning procedure, experimental design, as well as the exact measurements follows in the method section.

Hypotheses

(H1) Will the model of looking back at the correct branch during an unexpected event be able to predict the performance in the game?

(H2) Do children attend to the TA during an unexpected event, and is there a difference in attending the TA depending on who is in charge – the child or the TA?

In addition to the two hypotheses an explorative analysis will examine the measured eye movements; to see how they correlate with each other and with performance. Examining eye movements from an explorative perspective could unearth potential findings for future research.

2 Method

Participants

The target population for this study were children, ages 4.5 – 6.5 years. The sample for the study was taken from three different preschools, all of them located in the south of Sweden. The three preschools were selected by a convenience sample. In the study 42 children (21 girls, $M=4.6$, $SD=0.72$) participated.

Ethics

This study was a part of project which was ethically approved by the Regional Ethical Review Board of Lund. All of the children willingly participated and were given an anonymous ID instead of their name when the data was analyzed.

Materials

The experiment was performed on a laptop using Experiment Center (Version 3.5; SensoMotoric Instruments). The stimulus, Magical Garden, was presented on a separate 22" widescreen monitor (resolution = 1,680 × 1,050 pixels). Eye movements were measured using an iViewX SMI RED250 remote eye tracker (SensoMotoric Instruments) that recorded binocularly at 250 Hz. The eye-tracking data was analyzed in BeGaze (Version 3.5; SensoMotoric Instruments). An additional laptop was used to take notes on. Figure 3 displays the set-up for the eye-tracking experiment.

Procedure and the design of the experiment

During the first part of the study the children played Magical Garden on tablets. An experimenter introduced the game to the children, one at the time. The children played the game individually on tablets at the preschool 1-3 times per week. Each session lasted 15-20 minutes. The playing sessions were directed by the teachers at the preschools, during 3 weeks. The teachers were instructed in how to help the children log in and to distribute the tablets to children during the dedicated lessons. Each child logged in on an individual account to his/her own Magical Garden and digital friend (the TA). All the data from the played games were logged. The logged data from each game were; the number of correct answers, the number of tries, proportion of correct tries, the reaction time, how many games a child played, which level in the game their pace/progress through levels of difficulties and if they got "stuck" for a longer time on a certain level of difficulty. Each preschool had several tablets available to let more than one child play during each session. Therefore children generally focused on their own playing rather than that of other children.

The eye-tracking experiment was conducted at the different preschools in a group activity room or studio separate from the classroom. The set up (see figure 3) was a one-computer set up, one laptop connected to an external stimulus screen with a remote eye-tracker, a mouse and a second laptop was used for taking notes. The children performed the eye-tracking experiment individually. The eye-tracking experiment was split up over five sessions, to be able to collect data from every child, three times on the first preschool (18 children) and once on each of the two other preschools (10 and 12 children respectively). The average duration of the eye-tracking experiment including calibration was about 30 minutes. A teacher was present during the experiment for some children, sitting on a chair a few meters away, most of the experiments were conducted alone with the experimenter. To begin the experiment a calibration and validation of both eyes was performed. The children were told keep their head still and to look at a dot when it moved across the screen. To convey the importance of calibrating the eyes, without telling the children that their eye movements were being recorded, a cover story was told. The cover story was that looking at the dot and following it was the password in order to unlock the game and to start playing. To ensure an acceptable calibration, the experimenter gently held the head of the children, who were unable to keep their head still during the calibration. The calibration method was an 8-point with a black circle on silver background. The calibrations points were accepted by the experimenter hitting the space bar. The validation was 4-point with a white dot on a silver background. The validation points were automatically accepted when the fixation was stable on the dots. In this study a re-calibration was performed if a child deviated more than 1.7° on either eye during calibration. If the calibration dragged on too long the children would lose interest in participating. Therefore a maximum of four re-calibrations were performed and children who still had a deviation of more than 1.7° where allowed to continue the experiment any way and data was collected. This was a tradeoff between an accurate calibration and the children compliance and willingness to participate.

The children were asked which TA (Panda, Mouse, or Hedgehog) they wanted to play with during the experiment. The Bird Hero scenario was modified and the unexpected event was added as the experimental condition (for detail description, see Construction of the experiment program).

Each child played a minimum of five trials and maximum ten trials. The amount of trials each child undertook was decided by the willingness to participate. Some gladly played more than five trials and some needed encouragement just to complete five trials. Five trials were set as the minimal limit to have enough data to construct the measurements. A trial consisted of four tasks; (1a) control condition with child in charge and TA watching, (1b) experimental condition with child in charge and TA watching, (2a) control condition with TA in charge and child watching, and (2b) experimental condition with TA in charge and child watching (see table 1). The tasks were in the same order for each trial (1a, 1b, 2a, 2b).

In each task only one bird is saved. Before each new trial the bird parents were randomly assign to the different branches in the tree and the bird's colors are randomized. The color of the baby birds was also randomized because the correct branch for each task was randomized. The control condition had the correct branch be in the number range of 1-9 and in the experimental condition had the correct branch be in the number range of 1-6. In 1a and 1b the child was in charge and the one making the decisions which button should be pressed. The TA was standing in the background watching the child's actions. In the control condition (1a) the elevator would move to the chosen branch. If the child answered correctly the baby bird reached the correct branch and the baby bird and its parent cheered. If the child answered incorrectly the bird went to the chosen branch and was given the standard feedback "This is not my parent! I live higher up/lower down" and returned to ground floor for child to try again. In the experimental condition (1b) if the child answered correctly the unexpected event would happen (the elevator moving passed the chosen branch and crashes in the tree top). If the child answered incorrectly the standard feedback will be given. In (2a) and (2b) the TA was in charge, the one making the decisions which button should be pressed, and the child had to watch the TA. The way the TA made a decision was by thinking with a thought-bubble and saying "Um, it has to be this one" or "could it be this one?" and then the TA was programed to accept its own thought and send the baby bird to the branch represented by the thought-bubble. The TA was programmed to always choose the correct branch. In the control condition (2a) the elevator would move to the intended branch. In the experimental

Table 1. Displays the four different tasks of the experiment, and names them accordingly 1a, 1b, 2a, and 2b. 1 = the child is in charge, 2 = the TA is in charge. A = the control condition, b = the experimental condition.

Tasks	**Control condition (= a)**	**Experimental condition (= b)**
Child in charge and TA watching (= 1)	1a	1b
TA in charge and Child watching (= 2)	2a	2b

condition (2b) the unexpected event would happen. After completing a trial, four tasks, the TA and the birds cheered and gave water drops as a reward. Due to the children's lacking experience with handling a computer mouse, the experimenter conducted the clicking most of the time. The children instead pointed at the screen to indicate which elevator button should be pressed. The children had pressed the tablets screen to indicate the buttons during the tablet gameplay, so this was a natural solution to the mouse handling problem.

In between each trial the children was given the choice to use the water drops they have collected from the previous trials and watered their garden. This off-task was both necessary to keep up the motivation and acted as a break to preserve energy. Most of the children wanted to completely fill the watering can, complete two trials, before they watered. After completing at least five trials the children were thanked for participating and given a diploma.

Construction of the experiment program

JavaScript and HTML5 were used to program the experiment version of Magical Garden. In the experiment version (1.0:2015-02-16) a number of new elements were created and implemented compared to the original version (1:2014-11-26). The first task (1a) was the same as game mode (ii) except only one bird is saved. To create the second task (1b) the unexpected event was added to the (1a). The third task (2a) was constructed by removing the agree/not agree function of (iii) and adding an execute function for the TA to act and "press" the elevator button it is thinking on. To create the fourth task (2b) the unexpected event was added to the (2a). The service bird from the unexpected event was created in Paint.net.

Measurements and data analysis

In this study two performance measures were constructed, one from the tablet sessions and one from the eye-tracking experiment for each child, and the eye movements were recorded, coded, and analyzed.

As previously mentioned, all of game data from the

Figure 3. Picture of the eye-tracking set-up at the preschool.

tablet sessions were logged. The performance variable was was calculated from these logged data. The performance in Magical Garden, during the tablet session, for each child was calculated from the total number of correct answers divided by number of tries from all of the finished mini-games. A mini-game consisted of the three gameplay modes (i, ii, and iii). Therefore, the total number of tries in each mini-game was nine (the number of correct answers) plus one each wrong answer in the mini-game.

The performance variable from the special version of Magical Garden played during the eye-tracking experiment was based on the performance of the child from exercise 1a and 1b. Total performance was calculated from the total number of correct answers divided by the number of tries. Therefore, the number of tries per trial was two plus the number of incorrect answers. With a minimum of five trials the children performance was based on least ten exercises (1a and 1b). The game data from the eye-tracking experiment was not logged by the game, it had to be retrieved by manually review the recorded eye-tracking videos of the experiment. As previously mentioned the proposed model of detection, look back, would use AOI hits as the eye-tracking measurement of choice. How long must a fixation be for it to count as a hit? There is no consensus for the lower level of fixation duration in order to obtain visual input, fixation durations summarized by Rötting (2001; in Holmqvist et al., 2011) ranged from 60-120 ms and Granka et al. (2008; in Holmqvist et al., 2011) used 200ms as cut off. In this study the lower level of fixation duration for an AOI hit was set to 150 ms. By having a cutoff at 150 ms, fixations shorter than 150ms will not be recorded as AOI hits even though they were located on the AOI. This would reduce the variance of the sample. I would argue, that it is sound premise to interpret that the child has attended the AOI if a fixation is longer than 150 ms.

As mentioned in the introduction, additional eye movement data would be collected for the explorative analysis. Table 2 presents an overview of the four different eye movements measured. In table 2 the exact location of the AOIs on the screen are presented, as well as their specific time frame. The reason the unexpected event only was in the range 1-6, was that the children should have enough time to make a look back or another eye movement before the elevator reached the tree top. It was not counted as a look back if the child was looking at the correct branch when the elevator passed the correct branch. The child needed to have moved her gaze away from the branch and back to it, in order to define it as a look back.

The eye-tracking data was collected during the exercises with an unexpected event (1b, 2b) when the correct branch had been chosen. The data was coded accordingly, if a fixation was inside an AOI it was a hit (coded as a 1) and if there are no fixations inside the AOI during the time frame of the AOI it was a miss (coded as a 0).The missing data was either because gaze cursor was flickering or moving all over the screen or that the children did not look at the screen. When the gaze cursor was flickering or moving all over the screen, thus giving incorrect and not reliable data, it was coded as NA. If it was due the child not looking at the screen it was coded as a miss (0). The reason for coding missing

Figur 4. The three different AOI locations were used in the experiment: the correct branch, the TA and the thought-bubble. When the child is in charge the AOI is on the elevator buttons (see figure 2) instead of the thought-bubble.

Table 2. Displays a review of the different AOIs, their location on the screen and their timeframe.

AOI name	Area of interest	Time frame
1: Anticipate	The correct branch	From the choice of which branch the elevator should move to until the elevator is one branch under the correct level[3,4].
2: Look back	The correct branch	From when the elevator passes the correct branch (changes number on the bucket from the correct level) reaches the tree top.
3: Look at button/thought bubble	The correct button/thought bubble	From when the elevator passes the correct branch (changes number on the bucket from the correct level) until it reaches the tree top.
4: TA	The TA	From when the elevator passes the correct branch (changes number on the bucket from the correct branch) until it arrives at the correct branch.

data differently was that "not looking at the screen" should be the same as not attending the AOIs, if it would have been coded as NA, information would have been lost. The consequence would have been that, the results would have indicated a higher average level of attending to the AOIs than what was actually recorded. The mean probability that an AOI was attended to, was calculated from the number of hits in exercise 1b and 2b for each trial divided by the number of exercises performed in the eye-tracking experiment.

Manual coding of events

Due to eye-tracking of a dynamic stimulus the eye-tracking data was coded on to the AOIs manually in BeGaze. The method the fixations was coded by in this study was, by manually looking frame by frame and reported if the gaze cursor stays "still" fixated on a AOI (see above for the fixation level). Holmqvist et al. (2011) writes that this way of coding, thou time consuming, might be a good option when coding dynamic stimuli or gaze-overlaid videos. It is a somewhat subjective way of coding and opens up for potential of human error and inconsistency, but compared to algorithmic coding one can use "the advantage of being able to utilize the powerful pattern matching ability that humans have" (Holmqvist et al., p.175, 2011). The analysis of the data could have been conducted algorithmically by first manually editing the AOIs for each trial and then code the fixations on the AOIs in Begaze. The data from Begaze could then have been extracted and compiled to a statistical format with a script. The choice to conduct the data analysis manually and thus manually code for the fixations on AOIs was made for a number of reasons. The first reason, there was limited amount of time in this project and it would have taken too long add the AOIs on every trial. The AOIs would have had to be added to each participant because all videos differed in length therefore no static or general AOIs could have been placed on to the videos to save time. The second reason was the author's limited skill of handling and creating scripts to extract the relevant data. Rather than learning to extract all the necessary data by adding AOIs to each video it was concluded that it would be faster and easier to manually code the AOI hits or misses or NAs in an excel file, while going through each video frame by frame. Although manually coding has some disadvantages the measurement used in this study AOI hits allowed for manual coding of fixations. The manual analysis of the recorded eye movements was conducted a couple of weeks after the eye-tracking experiment, thus minimizing the potential bias when analyzing the material.

[3] The definition of "one branch under the correct branch" is when the number on the branch is showing on the bucket.
[4] If the correct level is the first branch, the time frame is from the elevator starts moving from the ground floor.

Table 3. Some overall statistics from the eye-tracking experiment and tablet session with mean and standard deviation.

Overall statistics	*M* (*SD*)
Tracking ratio (%)	50.29 (11.38)
Deviation X (°)	0.83 (0.54)
Deviation Y (°)	0.93 (0.60)
Games played on tablet	10.35 (3.81)
Performance tablet (%)	85.7 (5.5)
Trials played during eye-tracking	5.38 (0.95)
Performance eye-tracking (%)	85.2 (12.2)
- Control exercise (1a)	75.5 (25.5)
- Unexpected event exercise (1b)	91.2 (12.5)

3 Results

The data and analysis presented in this section was calculated from 40 of the 42 children (21 girls, *M=4.6*, *SD=0.72*).Two children were excluded from the study before the eye-tracking experiment due to sickness, and not being able to complete the calibration because of an eye problem. Five children had to do multiple re-calibrations and complete the eye-tracking experiment with X or Y deviation of larger than 1.7°. By doing a manual coding of the fixation, the five children with a calibration deviation larger than 1.7°, could be included instead of having to exclude them because their eye-tracking data during the unexpected event was readable. In table 3 some overall statistics are presented to give an overview of the experiment. The total tracking ratio presented in table 3 is a bit misleading it is the tracking ratio of the whole experiment, including watering the garden. The children attended the monitor during the unexpected event to a larger extent than they did during the rest of the experiment.

Looking at the descriptive statistics, from table 3, the tablet performance and performance during the eye-tracking experiment have similar mean percentages but a higher in standard deviation which is an indication of a larger spread. In an attempt to examine the difficulty levels of the exercises in the eye-tracking experiment, searching for floor effects or celling effects, the performance in the eye-tracking experiment was matched against the performance from the tablet sessions. The performance from the tablet sessions should significantly predict the children performance in the eye-tracking experiment. In order to conclude that the difficulty levels of the exercises matched, a linear regression was used to test if the performance in the tablet sessions could predict the performance in the eye-tracking experiment (see figure 5). The result of the regression showed that the predictors explained 1.75% of the variance (R^2 = 0.0175, F(1,38) = 0.6781, p = 0.41), with a 95% confidence interval (CI) of [-0.088 – 0.209]. The result does not indicate that performanceform the tablet session could predict the performance in the eye-tracking experiment. There was high performance overall 85.7% and 85.2% respectively (see table 3). A possible celling effect could be seen with nine children scoring 100% on the eye-tracking version of Magical Garden (see figure 4).

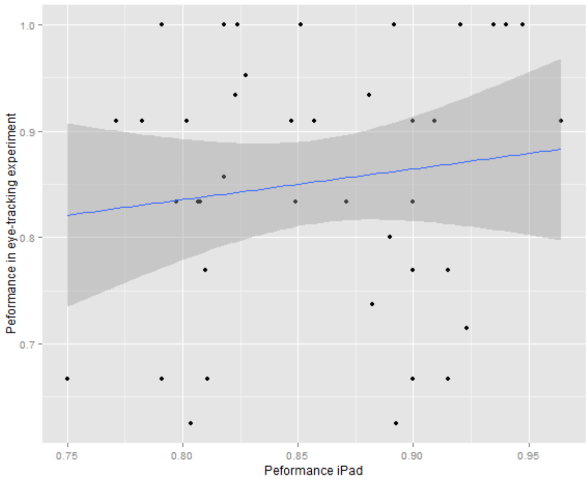


Figure 5. Scatterplot of performance from the tablet sessions as a function of the performance in the eye-tracking experiment (n = 40). A line was added to display the regression, the shadow represent the standard error.

Hypothesis 1

Will the model of looking back at the correct branch during an unexpected event be able to predict the performance in the game?

In order to answer the question a linear regression was used to test if the mean probability of a “look back” significantly could predict the performance in the eye-tracking experiment. The result of the regression showed that the predictor explained 13.8% of the variance (R^2 = 0.138, F(1,38) = 6.099, p = 0.018), with a 95% CI of [0.067 – 0.679]. The result seems to indicate that there is a correlation between a higher probability of looking back at the branch and higher performance in the game during the eye-tracking experiment (see figure 6). In an attempt to reduce the variance in the data, the time frame of a look back decreased to four seconds instead of until it reached the tree top.

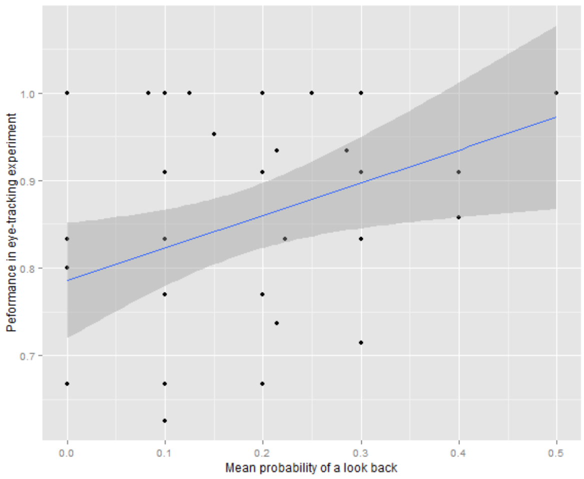


Figure 6. Scatterplot of performance from the eye-tracking experiment as a function of the mean probability of a look back (n = 40). A line was added to display the regression (0.37). The shadow represent the standard error. Some dots overlap.

The reason for this was that every level should have the same length of time. When the time frame for the AOI look back was “until the elevator reaches the tree top”, the time frame for doing a look back was longer for the first level compared to the second level compared to the third level etc. It took the elevator one second to move from one level to the next. From the sixth level until it reaches the tree top it took four seconds, therefore the time frame was set to four seconds. A linear regression was used to test if the mean probability of a “look back” with the time frame of 4 seconds significantly could predict the performance in the eye-tracking experiment (See figure 7). The result of the regression showed that the predictor explained 1.6% of the variance ($R^2 = 0.016$, $F(1,38) = 0.6382$, $p = 0.4293$), with a 95% CI of [-0.248 – 0.571]. When the time frame was adjusted to be of same length for each level, the model of looking back at the branch could not significantly predict the performance of the eye-tracking experiment. The number of look backs with the four second limit was decreased by 35% compared to when there was no limit.

Hypothesis 2

Do children attend the TA during an unexpected event and is there a difference in attending the TA depending on who is in charge of the decision – the child or the TA?

In this section the results from the data of the AOI: looking at the TA is presented. The mean look at TA was 24.4 % (*SD*= 26.2%), which means that the TA was attended to in a fourth of the trials. Figure 8 shows the overall mean look at TA when depending on who was in charge of the decision to press the button. When the children were in charge the mean look at TA was 17.2 % (*SD*= 2%) and when the TA was in charge the children the mean looked at the TA was 31.6 % (*SD*= 29%) (see figure 8). In order to negate the possibility of a few children alone increasing the total mean, the

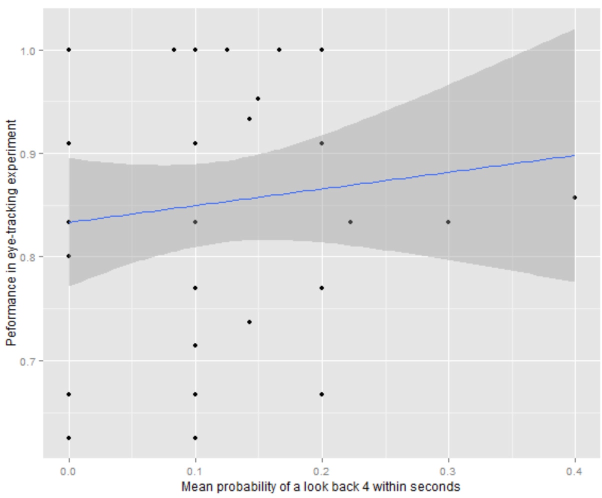


Figure 7. Scatterplot of performance from the eye-tracking experiment as a function of the mean probability of a look back within 4 seconds (n = 40). A line was added to display the regression (0.16), the shadow represent the standard error.

individual difference in looking at the TA when the TA was in charge and looking at the TA when the child was in charge was calculated for each participant. The result was plotted in a histogram (see figure 9). The distribution of the histogram in figure 9 looks like it could be normal distributed, therefore normal distribution is assumed in order to do a t-test. A t-test was performed, the mean difference in looking at the TA when TA was in charge and when the child was in charge for all the children was 14.3% (*SD* = 24.8%), indicating that looking at the TA when the TA was in charge where significantly higher than looking at the TA when the child was in charge, $t\,(39) = 3.66$, $p < .001$. The difference could also be presented with a 95% CI of [0.06 – 0.22].

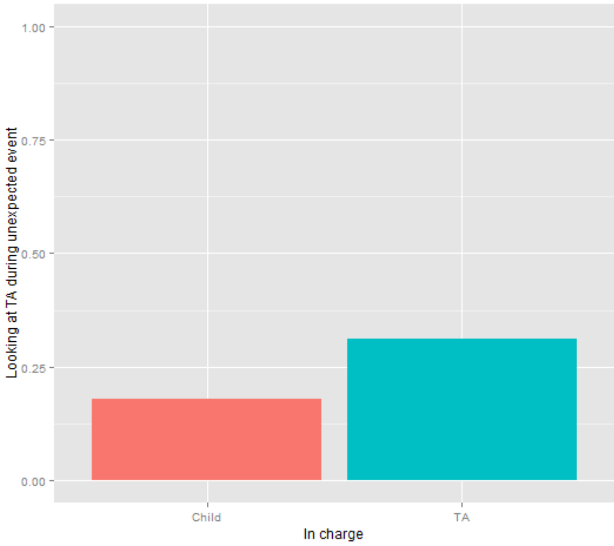


Figure 8. Bar plot of mean number of looking at the TA during an unexpected event combined for all participants (n = 40), divided up by in charge (child or TA).

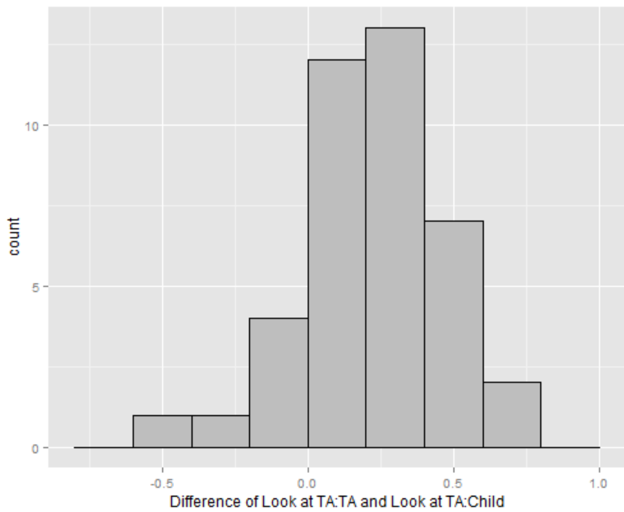


Figure 9. Histogram of all the participants (n = 40) the difference of mean look at TA when TA was in charge and mean look at TA when the child was in charge.

Table 4. Person's correlation of coefficient correlations between performance in the eye-tracking experiment and the different measured eye movements (n = 40).

	1.	2.	3.	4.	5.
1. Performance ET	-				
2. Anticipate	*0.09*	-			
3. Look at button	*0.5***	*0.41**	-		
4. Look back (4 sec)	*0.13*	*0.11*	*-0.01*	-	
5. Look at TA	*0.29*	*0.22*	*0.31**	*0.39 **	-

*p < 0.01, **p < 0.001

In order to investigate how and if the different eye movements measures correlated with each other and with performance, in search for possible positive or negative correlations. A Pearson's correlation of coefficients was conducted and is presented in table 4. The p values of table 4 are not adjusted for multiple tests. A strong correlation between the eye movement look at button and the performance was found $r(38) = .5$, $p < .001$. The 95% CI for predicting performance with look at button was [0.152 – 0.561].

The statistical analysis was conducted in RStudio in the language R (Version 0.98.501).

4 Discussion

In this study, an eye-tracking experiment was conducted examining 40 children between the ages of 4.5–6.5 years. The educational game Magical Garden was argued to be a suitable operationalized exercise for testing number sense. It was argued that noticing an unexpected event, tree elevator malfunction, when playing the game would reveal a child's number sense capabilities. The unexpected event was designed to only call attention from top-down controlled processes. Therefore, the children who had a sufficient level of number sense would only be the ones who understood that an unexpected event had happened before it is revealed. The children had been familiarized with Magical Garden by playing it on tablets for three weeks before the eye-tracking experiment. A model for noticing the unexpected event was proposed. Namely, the children would look back at the correct branch when an unexpected event occurred. Look backs were measured by AOI hits and the children's performance in both the eye-tracking experiment and the familiarization phase was recorded.

The first hypothesis, H1, predicted that higher probability of performing a look back would correlate with higher performance in the eye-tracking experiment. The second hypothesis, H2, predicted that there would be a difference in the probability of attending the TA during an unexpected event depending on who was in charge of the decision making (the child or the TA). The final goal of the thesis was to investigate, from an explorative point of view, how and if the captured eye movements correlated with each other and the performance measure.

The results were the following. Concerning the first hypothesis, H1, the results showed that the proposed model, a look back, significantly predicted the performance in the eye-tracking experiment. This meant that children with a higher probability of performing a look back had higher performance in the experiment. However, higher probability of making a look back did not correlate with higher performance if the time frame was controlled for and set to four seconds. Concerning the second hypothesis, H2, the children attended the TA during an unexpected event, and there was a significant difference in the mean probability of looking at the TA depending on who was in charge. The results showed that children had a higher probability of looking at the TA when the TA was in charge than if the child was in charge. In the third goal, the explorative investigation found a strong correlation ($r = .5$) between the AOI hits on "look at the button" and performance in the eye-tracking experiment. A few additional medium sized correlations were found: between AOI hits on "look at button" and "anticipate" ($r = .41$), between AOI hits on "look at TA" and "look at button" ($r = .31$), and between AOI hits on "look at TA" and "look back 4 sec" ($r = .39$). There was overall high performance in the eye-tracking experiment (*M*=85.2%, *SD*= 12.2%). Some children even scored 100%, indicating a possible ceiling effect. In the control condition (1a) the performance was (*M*= 75.5%, *SD*= 25.5%) lower than in the experimental condition (1b), (M= 91.2%, *SD*= 12.5%). Thus, indicating a difference in difficulty.

In this section I will discuss whether the results were reasonable and how they could be interpreted. Firstly, the performance measure from the tablet session did not correlate with the performance from the eye-tracking experiment. This could be interpreted as an indication of an imbalance between the two performance measures. One reason for this imbalance might be that the performance measure on the tablet sessions and eye-tracking experiment were not actually representative of the children's level of number sense. The potential celling effect and overall high performance in the eye-tracking experiment might have skewed the data. Also, maybe progression is a better or more representative way of calculating a longitudinal measure of number sense instead of average performance from all exercises of the tablet sessions. For future research, a progression measure could be constructed from the logged data from the tablet sessions. Secondly, concerning H1, the model of look back without the four second limit showed that there was a significant correlation with performance. This indicates that the proposed model of detection was not completely off target. However, the look back model did not seem to predict every time a child noticed an unexpected event. When the four second limit was added to the look back model, 35% of the look backs did not fulfill the criteria of a look back. By adding this limit, no significant correlation was found with performance. An explanation could be that there was not enough power left in the data. Additionally, the looming celling effect of performance indicated that the eye-tracking experiment was too easy.

In effect, the four second limit meant that 35% of the look backs happened after four seconds. One question that arose during the analysis was why do children sometime perform a look back only after four seconds? One reason for this could be that the children noticed that something was wrong only when the elevator was at a great enough distance away from the correct branch. Therefore, the look backs were too late to be considered a look back with the four second limit. Another possible explanation might be that looking back at the correct branch is not the only way to indicate detection of the unexpected event. This could be an explanation why some look backs occurred first only after four seconds. An observation made by the experimenter was that during the eye-tracking experiment both verbal and non-verbal detections of the unexpected events were made regularly. Although, these forms of detections were not officially documented because the experimental design was focused on eye movements as the model of detection. The most common non-verbal response to the unexpected event was for the child to turn and look at the experimenter. When the child turned her head and looked at the experimenter, the eye-tracker could not track their eyes. This lead to a decrease in documented detections of the unexpected event within the four second limit. The explorative correlation analysis conducted across all collected eye movements showed that the child looking at the button correlated strongly (*r=.5)* with her performance. It makes sense that looking at the button should be positively correlated with performance, because the chosen button contained information about which branch the elevator was supposed to go to. In this context, it makes sense for the child to look at the button if they start to suspect that something is not right. Also, the correlation between anticipating eye movements and looking at the button could be logically explained. If a child is anticipating the tree elevator to go to a certain branch and it does not, then looking at the button in order to verify one's suspicion that something is wrong makes sense. The fact that preschoolers even make anticipatory eye movements is worth emphasizing for future research to further investigate its relation to noticing and understanding of number sense. It is hard to explain why there should be a positive correlation between look at TA and look back 4sec/look at button. The result of the correlation analysis indicated that the children noticed and detected the unexpected events in different ways. The correlation analysis was not adjusted for multiple tests. With an explorative and method searching approach, adjusting for multiple tests is not necessary. The results of the correlation analysis could be used an indication of possible detection criteria for a new model of detection. Another experiment would have to be conducted in order to evaluate the new model. Future research has a foundation to build a new model of detection from the results of H1, the explorative analysis, and the observations of verbal and non-verbal detections. The new model of detection could include verbal, non-verbal detection parameters, as well as eye movements.

The results concerning the second research question (H2) showed that children overall, during the unexpected event, do not attend the TA much (less than 25% all the trials). The fact that children look more at the TA when the TA is in charge, as opposed to when the child is in charge, could be interpreted as the child acknowledging the TA as a fellow player in the game, therefore the child could be said to display ToM capabilities. These results were in accordance with the conclusion from Axelsson et al. (2013) that preschoolers are able to engage in social interaction with a TA.

In this study, there were advantages in ecological validity of using a remote eye-tracker and eye-tracking as a measurement for detection. One of which is that you are able to conduct the experiment at the children's individual preschools, as opposed to a laboratory setting. The fact that a remote eye-tracker is not as intrusive as a head- or tower-mounted system resulted in that the children did not realize that their eyes were being recorded. Also, the experimenter did not have to disclose the unexpected event, which would have had to be necessary if the noticing of the unexpected event was measured by the pressing of a button. The use of a remote eye-tracker was deemed less intrusive than a head or tower mounted system, which in turn arguably led to more organic reactions from the children. One disadvantage was that the children were able to move around which decreased the tracking ratio. If the children would have been recorded with a tower mounted system, the tracking ratio would have been better. However, the willingness to participate and the number of trials the children would have been able to complete would have plummeted. The off-task, watering the garden, was a motivational boost for the children, and also an effective way of maintaining their willingness to participate and play more.

Conducting the calibration behind a cover story, looking at the calibration dots were necessary to unlock and start the game, was a great way of making the calibration a fun activity for the children. Furthermore, it was very hard to calibrate the children due to them moving their heads and not being able to sit still. Trueswell, Sekerina, Hill and Logrip (1999) also reported difficulties in calibrating children which eventually led to participant attrition. Trueswell solved the calibration problem with a head mounted display by calibrating on a short adult with small eyes and then made adjustments for the child before recording. In this study, gently holding the child's head during calibration had a positive effect decreasing the deviation and possibly also decreased the participant attrition, which is something future research might want to replicate.

Manually coding the eye-tracking data worked well and was a great time saver for this study. Coding the data manually gave me time to conduct a statistical analysis and finish the study. Although, the major drawback of having conducted the data analysis manually was that if you wanted to analyze another AOI or add a four second limit, all the data must be processed again. Therefore, due to limited time, only the look back AOI was controlled with a four second limit. Another drawback with manual coding was that I only got one measurement, AOI hits, out of the analysis. If an analysis would have been conducted by adding dynamic AOIs to the eye-tracking data and analyzing it automatically, data such as dwell time on the AOIs, and a reaction measure for when the unexpected event was detected could have been added to the statistical analysis. Unfortunately, analyzing it automatically was not an option when the decision of which method to use had to be made. In parts, because of the time it would have taken to add individual dynamic AOIs to every participant, and my, at the time, limited experience in programing scripts to handle the data.

In hindsight, if I would have had the opportunity to redesign the study and more time, a number of different experimental design choices would have been made. For instance, I would not have coded the fixations and AOI hits manually. Instead, I would have implemented the dynamic AOIs in order to process the eye-tracking data automatically. I would have also used progression derived from the tablet sessions as a measure of the children's baseline level of number sense. I would also perhaps add an additional operationalized number sense exercise to compare with the progression measure and the performance in the eye-tracking. Although, comparing a progression measure to traditional operationalized tests might be a whole new study. The present design does not control for WM. Therefore, a WM test should be added. In the present study, there could be children with WM difficulties who do not react to the unexpected event, yet have a high level of number sense. In addition, a ToM test should have been included, in order to control for ToM. In the present study, it could be the case that only the ones with fully developed ToM capabilities are the ones who interact socially with the TA (looking at the TA). I would have also recorded the verbal and non-verbal detections in addition to the eye-tracking measurements. Furthermore, a number of different design choices within the experiment program would have been made. For instance, not using randomly generated correct answer for the tasks, instead have all the children play the same amount of exercises with branch one as the correct answer, branch two etc. In an attempt to counteract order effects, the exercises could be counterbalanced with a *Latin square*. Having all the children play the exercises, with the same "correct answer" would have given me the opportunity to analyze if detection occurred more often on some branches than others. Another addition would be that the children would have to play the same amount of trials, instead of having them play as many as they wanted. Another different design choice would have been making the control exercises (1a, 2a) use the same number range as the manipulation 1-6 instead of 1-9. This would eliminate the difference in difficulty between the experimental and control condition.

Conclusion and future research

This study could be seen as a groundbreaking introduction to the usage of unexpected events and noticing the unexpected event as novel way of letting children expose their level of number sense. The proposed model of noticing, a look back, did not account for the whole notion of detecting an unexpected event. However, this study emphasizes that a better model of noticing could be constructed by combining measurements of verbal, non-verbal detections, as well as eye movements such as look back and look at button. Future research could learn from this study and examine the possibility of creating a better model of noticing. One idea might be to create the model of detection based on the way experts (adults or children that have the sufficient number sense) detects the unexpected event and then compare that to developing children. With simple tricks it is not as hard as it sounds to keep preschoolers interested while conducting an eye-tracking experiment. Overall, using an educational game as an operationalized exercise to test number sense worked well. It also had the additional advantage of increasing the children's motivation and willingness to participate in the experiment. Even though the children do not attend the TA much during an unexpected event, this study showed that children attend the TA more when the TA is in charge and this could be interpreted as displaying ToM capabilities. Further research on children's interaction with TA is needed, and it seems that the research on preschooler's ability to engage with educational games and TAs only begun.

Acknowledgements

I would like to thank all of the wonderful children from the preschools who participated and made conducting an eye-tracking experiment an easy and enjoyable task.

A very special thank you to my supervisors Professor Agneta Gulz and Dr. Richard Andersson for the guidance, valuable comments, and support all the way through the process of writing this master thesis.

I would also like to thank the Humanities Lab and Kenneth Holmqvist for letting me use the remote eye-tracker.

Many thanks to Magnus Haake, Rasmus Bååth, and Emil Boman for all the help with statistics, programming, and data management.

Another special thanks to Martin Lingonblad and Kalle Palm for all the memorable moments and support while writing in the basement.

Last but not least, I would like to thank Andreas Londos for proofreading and Fanny Johansson for supporting me every step of the way!

References

Axelsson, A., Anderberg, E., & Haake, M. (2013, January). Can preschoolers profit from a teachable agent based play-and-learn game in mathematics?. In *Artificial Intelligence in Education* (pp. 289-298). Springer Berlin: Heidelberg.

Biswas, G., Katzlberger, T., Bransford, J., & Schwartz, D. (2001). Extending intelligent learning environments with teachable agents to enhance learning. In *Artificial Intelligence in Education* (pp. 389-397).

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*(3-4), 363-392.

Case, R., Okamoto, Y., (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the society for research in child development*, i-295.

Case, R. (1998, April). A psychological model of number sense and its development. In *annual meeting of the American Educational Research Association, San Diego*.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59-68.

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention*, *30*(2), 3-14.

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, *18*(4), 334-352.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471-533.

Clarke, B., & Shinn, M. R. (2004). A Preliminary Investigation Into the Identification and Development of Early Mathematics Curriculum-Based Measurement. *School Psychology Review*.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive development*, *9*(4), 377-395.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., ... & Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761-773.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215.

Dehaene, S. (1997). The number sense. New York: Oxford University Press.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision research*, *36*(12), 1827-1837.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 455-470.

Garnham, W. A., & Perner, J. (2001). Actions really do speak louder than words-but only implicitly: Young children's understanding of false belief in action. *British Journal Of Developmental Psychology*, *19*(3), 413 -432.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, *1*(1), 20-23.

Gersten, R., & Chard, D. (1999). Number Sense Rethinking Arithmetic Instruction for Students with Mathematical Disabilities. *The Journal of special education*, *33*(1), 18-28.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist*, *50*(1), 24-37.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of learning disabilities*, *37*(1), 4-15.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science, 14*(1), 79-106.

Green, H. J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta psychologica*, *125*(3), 257-278.

Griffin, S.A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.). *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25-49). Cambridge, MA: MIT Press.

Griffin, S. (2004). Building number sense with Number Worlds: A mathematics program for young children. *Early childhood research quarterly*, *19*(1), 173-180.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, *7*(11), 498-504.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Moreno, R., & Mayer, R. E. (2005). Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *Journal of educational psychology*, *97*(1), 117.

Gulz, A., Haake, M., & Silvervarg, A. (2011, January). Extending a teachable agent with a social conversation module–effects on student experiences and learning. In *Artificial intelligence in education* (pp. 106-114). Springer Berlin: Heidelberg.

Lindström, P., Gulz, A., Haake, M., & Sjödén, B. (2011). Matching and mismatching between the pedagogical design principles of a math game and the actual practices of play. *Journal of Computer Assisted Learning*, *27*(1), 90-102.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense large-number discrimination in human infants. *Psychological Science*, *14*(5), 396-401.

Pareto, L., Arvemo, T., Dahl, Y., Haake, M., & Gulz, A. (2011, January). A teachable-agent arithmetic game's effects on mathematics understanding, attitude and self-efficacy. In *Artificial Intelligence in Education* (pp. 247-255). Springer Berlin: Heidelberg.

Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: evaluating math comprehension and motivation. *Educational technology research and development*, *60*(5), 723-751.

Pellegrino, J. W., & Goldman, S. R. (1987). Information processing and elementary mathematics. *Journal of Learning Disabilities*, *20*(1), 23-32.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(04), 515-526.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics*, *11*(2), 129-158.

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., ... & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, *23*(3), 409-422.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129-184.

Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: Interactivity in the teachable agent paradigm. *Learning with animation: Research and implications for design*, 114-140.

Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology*, *16*(3), 53-81.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89-134.

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10-22.