Noticing an unexpected event in Magical Garden - with a Teachable Agent using Eye-Tracking

Upptäcka något oväntat i Magiska Trädgården

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

Noticing 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 have been presented. What often has been concluded is that children are capable of more than what was thought. The focus of this study is, children´s developing number sense.

New technology has introduced a genre of educational games which utilize the captivating powers of computer games and they have shown a positive effect on learning and motivation. A new method of measuring number sense is proposed and tested: Detecting an "unexpected event" in the educational game Magical Garden. The unexpected event was constructed in a way that only children with sufficient number sense would notice the unexpected event. The pedagogical instrument teachable agent (TA) is a part of Magical Garden´s design.

42 preschoolers participated in this study which used eye-tracking method of capturing possible detections of the unexpected event, as well as measuring the interaction between the children and the TA during the unexpected event.

The results showed that children attend the TA significantly more when the TA was in charge of the decisions in the game. Detecting an unexpected event could be promising method of measuring number sense. However, the model of detection used in this study was not comprehensive.

.1 Introduction

How would the world look like if you did not understood numbers, relations or quantities? That is a very hard thing to even imagining because numbers, relations and quantities are such a fundamental part of human cognition and thinking. Although, this is the reality for many young children, children have to 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 and without intervention this could lead to failing early math (Griffin, Case, & Siegler, 1994) and developing learning disabilities (Gersten, 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 an effect on both learning and motivation (Schwartz, 2004; Moreno, 2005).

In this thesis an "unexpected event" is introduced to the educational game Magical Garden and is thought to act as a method of testing number sense capabilities in preschoolers. The unexpected event was constructed in a way that only children who have sufficient number sense would notice the unexpected event. This method of testing number sense is groundbreaking and has not been done before. The measurement used to confirm that a child had noticed the unexpected event was with eye-tracking technology. A proposed model of noticing will be presented and tested.

Magical Garden utilizes a teachable agent as a pedagogical instrument. Question has been raised whether preschoolers can benefit from a teachable agent, due to the fact that most preschoolers have not a fully developed theory of mind. Haake et al. (2013) was pioneering in researching preschooler's relation to a teachable agent. Their results indicated that preschoolers can interact socially with a teachable agent and they requested further research in the relationship between preschoolers and teachable agents. In this thesis I aim to investigate the interaction between the teachable agent and the children during the unexpected event and hopefully add to the growing body of knowledge regarding the interaction between teachable agents and preschoolers.

The structure of the thesis

First in the introduction, a report on previous research in the fields of number sense, learning and teachable agents will set the stage for the thesis. Then, noticing something unexpected is proposed as a way of examining children's level of number sense and the method of using eye-tracking technology to capture the act of noticing the unexpected. An "unexpected event" is introduced as an experimental manipulation and a model of noticing it is proposed. The introduction is concluded with 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. The results are presented and discussed. The thesis in concluded with a conclusion and future research.

Number sense and learning

In this 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). Another important skill when learning is, to be able focus the attention towards what is important. Gelman emphasizes the importance of attending the right things in order to gather more knowledge and build up the cognitive skill set (Gelman, 1990). Number sense is the ability to understand numbers in a broader sense, not only as numbers but also as quantity, as something you can manipulate, something flexible, and a way of looking at the world. 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 the child is 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 represented in the understanding that five could be represented in a lot of different ways, so 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. 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 understand that specific quantities can also represent general amounts. Number sense starts to develop in the childhood. Gersten et al. argue that number sense should be viewed as a skill which could be trained and taught as opposed to something innate and static (Gersten, 1999). Additionally, Lipton describes in his research that infants already have a basic understanding of quantity and ability to discriminating different quantities, the precision of discriminating increases over development (Lipton, 2003). 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 to the left and bigger to the right (Dehaene, 1997). 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 the numbers when she can recite numbers. According to Carey (2004), children often do not understand what parents tell them due to the fact that they cannot attach the words to concepts of the numbers (Carey, 2004). The first things children learn about numbers is the order of numbers and to recite it 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 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, 2011). Giving semantical content to the unknown number "three" is done by taking its order in consideration, understanding that "three" is after "two", and that "three is one more than two". Each step, 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 senses depending on their development of number sense. Gersten argues that the process of learning the concepts of numbers is an automatic gateway into mathematics and the key to solving basic arithmetics (Gersten, 1999). All of the concepts presented above; order, cardinality 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 operationalize and measure number sense e.g. counting, number identification, quantity discrimination, and missing number. 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 argued that these counting exercises are a good way of measuring number sense. Clarke and Shinn (2004) conducted an experiment testing number sense based on exercises testing number identification with the exercise identify a given number between 0-20, quantity discrimination and mental number line with the exercise which one of two numbers is the bigger one, and missing number with the exercise fill in the "blank" number which is missing in a sequence. Okamoto and Case (1996) created the number knowledge test which is regarded to be an exhaustive test for assessing the child´s conceptual knowledge of numbers (Chard et al, 2005). It consists of a number of exercises, like the ones presented above, on three levels with increasing difficulty forming a perfect Guttman scale. A child's level of number sense was determined when they could not complete two of the same task. 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 and there is no one way of doing it. Although there are different exercises they all are related to number sense.

What are the consequences if a child does not develop and increase 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, 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 tested was to systematically train children with math facts, in order to automatize their thinking procedure (Pellegrino, Goldman, 1987). This type of intervention worked poorly, it was both unpleasant for the children and had the opposite effect children lost interest in math (Gersten, 1999). The importance of play as a pedagogical tool for teaching and to get children motivated in their acquisition for new abilities has been known for many years. Play and playing is by Geary considered a primary way of adding meat to the conceptual skeleton in form of new abilities and knowledge. Also different types of games with numbers and play involving numbers have been found all over the world (Geary, 1995). Griffin, Case, and

Siegler suggest 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 (Griffin, Case, & Siegler, 1994). Gee describes that when the play and fun in learning disappears, motivation and educational opportunities plummets (Gee, 2003). In this introduction to number sense, basic concepts of number sense has been unveiled, as well as different ways to operationalize number sense with exercises, display the tight connection between number sense and mathematics, and emphasize the problems which will follow with a lack of number sense.

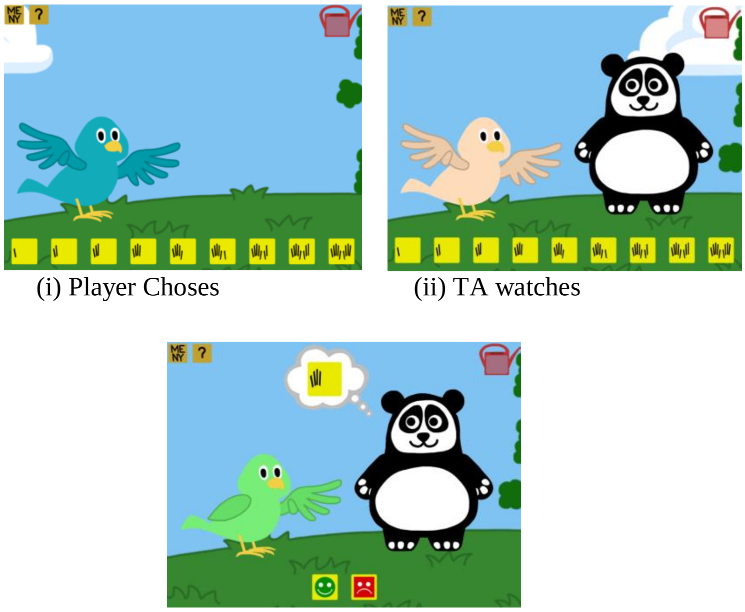
Magical Garden

In this and the coming section I will present the educational game Magical Garden and argue that it is an operationalized dynamic exercise and could measure number sense capabilities. After this I will introduce the concept of a *Teachable-Agent* (TA) and discuss its addition to the Magical Garden.

In this study, children ages 4-6 will play the educational game Magical Garden which is designed for 4-6 year olds 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 in to take the tree-elevator to their parents (see figure x). The second mini-game is “Bee Flight”, where the mission is to help a bumblebee in find nectar from the right flower. The third mini-game is “Lizard”, where the mission is to help a chameleon shoot its tongue to hit ants on a tree. The fourth mini-game is “Balloon”, where the mission is to help a woodlouse go on a treasure hunt in 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.

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 & 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, in order to teach mathematics or number sense. This is meant to be a fun activity for the children and a goal to keep the children engage and willing to play more.

Each of the mini-games follows the same structure. There are three different modes of gameplay (i) the player choses which button to press, (ii) the TA watches the player chose which button to press or (iii) the TA thinks of a button (display a proposal for which button it might be in a thought-bubble) and asks the player if the button it thinks about is the correct button. 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 thinks that the TAs proposal is wrong. If the player presses the red button, the player has to show the TA which the correct answer is by pressing the correct button. Every game is built up by these three modes (see figure X). Each mode is played three times in each mini-game. Depending on the performance in previous mini-games, the difficulty 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 is added to the gameplay.



(i) Player Choses

(ii) TA watches

(iii) TA thinks and the player agree or not agree

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

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 in an elevator made of a bucket. The baby bird shows the player which branch it wants to go to with its feathers and the task is to press the yellow button with the same number of fingers the bird is showing (i). The baby bird jumps into the tree-elevator and the elevator displays, in symbolic number, current branch on the front of the bucket. When the elevator stops at the correct branch, the bird jumps off and is welcomed by its parent and 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, “this is not my parent”, and that the baby bird lives further down or higher up. 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 (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 (iii) (see figure X). Three more birds are helped this way and when all the birds have been saved the TA and the birds cheer and say “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 mention above, the lift buttons represents the amount of objects. The baby bird’s feathers also represent an amount. Finally, the tree is represented like a scale where the correct branch is the value of the scale (see figure 2). The combination of these three representations[1] in the game creates an operationalized exercise such as number identification, being able to make the abstraction that quantities can be represented in different ways and the order of numbers both in the elevator buttons and the branches of the tree. The exercises are combined by elements of training number sense capabilities and therefore, failure to complete an exercise, in the game would be an indication of not fully understanding the branch of number sense which currently is trained or tested. Therefore, Magical Garden is a well suited platform to train number sense because it utilizes key concepts and have operationalized exercises in the form of mini-games.

Teachable-Agents

Magical Garden is also a Teachable-Agent-based game. As previously mentioned the TA is a part of the (ii) and (iii) game mode. A Teachable-Agent is a computer application represented as a digital student. Teachable agents were introduced by Schwartz and Biswas to the paradigm of intelligent learning environments in an effort to enhance student’s motivation and learning (Biswas, 2001). A TA is trained and taught by the user and cannot learn in any other way. Therefore, a TA is a reflection of what it has been taught. The TA is able to display what it has been taught which is thought to motivate and give the student an opportunity to reflect upon it (Biswas, 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 in the act of tutoring, the tutor as well as the tutee gain understanding and benefits from tutoring in a form of social learning (Chi, 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). 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 teachable agents has mainly been carried out on undergraduate students (Schwartz, 2007; Biswas, 2001) and 8-14 year old children (Chase et al 2009; Gulz, Haake & Silvervarg, 2011; Pareto et al, 2011; Pareto et al, 2012; Lindström et al, 2011). Other than Axelsson, Anderberg, & Haake (2013) preschoolers, age 4-6, have not been extensively tested in their interaction with a TA. One question that arises is whether preschoolers can even benefit from the pedagogical help the TA is offering. One argument against the idea that a TA would be a strong pedagogical aid for preschoolers is that preschoolers do not have fully developed theory of mind (ToM). The premise of interacting with a TA is to create an environment of social learning and because preschoolers are not able to do this they cannot benefit from a TA. Premack and Woodruff (1978) described ToM as the most important factor to understand and to act socially. 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. 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 is thus diluted and meaningless. ToM has traditionally been measured by a false belief test. 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). Does this mean that preschoolers cannot utilize a TA without a fully developed ToM? 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 derived from anticipating eye-movements. Axelsson, Anderberg, & Haake (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 a TA and no be disturbed by its presences. This led them to conclude that it is possible for preschoolers to interacting socially with a TA. Axelsson, Anderberg, & Haake (2013) emphasized the importance of further research on preschoolers. In this study I wish to delve in to and examine if there is something indicative of children, acknowledging the TAs role as another player in the game, thus revealing ToM capabilities. How do the children along the TA when the TA is in charge of the decision making in the game? Could this be a situation where the children have an opportunity to acknowledge the TA as another player in the game and also show ToM capabilities?

With these introducing sections I have started to set the

[1] The baby bird’s feathers indicating which level of the tree is home and that the lift buttons connection to the different levels of the tree and pressing the lift button with the same amount as the baby bird is showing will take it to that level of the tree.

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, learning the concepts of the representations, and play. Number sense can be operationalized and measured by using exercises that assess the basic aspects of number sense. It is also possible to sieve out children with a lacking number sense using operationalized exercises. 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. Below I will discuss why noticing an unexpected event would be indicative of developed number sense. Thereafter, I will also discuss and describe how the noticing could be captured and measured.

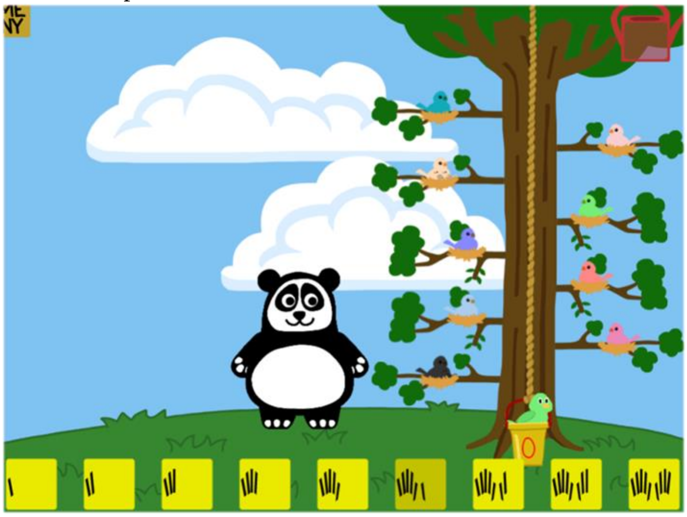


Figure 2. Displays the mini-game Bird Hero with all of its components the tree with nine different branches, the baby bird, the elevator, the buttons, and the Teachable agent Panders.

Noticing

How does one notice anything at all? Noticing is a private experience and one way of defining it is if a verbal report could have been given (Schmidt, 1990). Schmidt describes that there is a distinction between perceiving information and noticing information. One must actively attend to something in order to be able to notice it. Therefore, noticing is according to Schmidt (1990) the basic indication of being aware of something. To be aware and act upon the awareness, information must be processed and manipulated by the brain. Corbetta and Shulman (2002) describe two systems with functional differences that work together to guide the attention in a visual environment. 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 luminesce. The top-down system depends on endogenous or cognitive stimuli such as knowledge, expectation, and goals. I most cases, attention is driven by a combination of bottom-up and top-down stimulus. If the unexpected event is disclosed by bottom-up stimuli it is not sound to give credit to number sense. In order to conclude that it depends on number sense noticing must be evoked by top-down attention, be driven by previous knowledge, expectations, and mismatch with the intended goal. Could working memory (WM) be the reason for noticing or not noticing an unexpected event? WM is 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 should be credited to number senses. The child has to have the sufficient level of number sense, e.g. know that five feathers represent the fifth branch, and use the information kept in mind by the WM in order to notice an unexpected event. Could a detection of something unexpected, be indicative a developed number sense? If the design of the unexpected allows a detection to be made by top-down control, it seems possible. The exercise also has to be operationalized in a way to test developing number sense. Then if a detection is made, a mismatch noticed in a visual environment by top-down control, how could it 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. What if someone is shy or uncomfortable talking to the experimenter? 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 the unexpected event out loud. Another possible way of manifest a detection would be to let the children press a button to indicate that there has been a manipulation. Pressing a button would also be a clear indication that could easily be measured. An issue with pressing a button as an indicator for noticing is that the unexpected event has to be disclosed in order for the children to know when and why they are supposed to press the button. Yet another problem is that the children have to actually follow the instructions and remember to press the button when noticing the manipulation. Both of these presented ways of manifest noticing, runs in to trouble by not catching all of the actual noticing of the unexpected event. Smith (2012) presented a way of manifest noticing as increased visual attention and a way of measuring increased visual attention via 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 of previous knowledge (Henderson, 2003). Deuble and Schinerder (1996) presented evidence for the hypothesis of the coupling of eye-movements and the visual attention towards a target object; therefore 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 which is consistent with the hypothesis that attention and eye-movements are tightly connected at neural level.

A detection of the unexpected event with increased visual attention must then be based on top-down controlled attention to the mismatch and not to bottom-up effects for it to be for fill the requirement of noticing. The close connection between top-down control and the control of eye-movements was presented, which gives grounds for investigate eye-tracking as an online-measurement for the detection on a dynamic unexpected event. Before the proposed model for detection is introduced I will provide a short introduction to eye-tracking.

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 which 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, depending on its amplitude it can be faster or slower than 30-40 ms. The eye makes on average three saccades per second. During a saccade the visual input which is obtained is 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, 2011). When the eye is “still”[2] and the center of gaze is directed towards an object this is called a fixation. During a fixation, visual input is collected and actively processed. A fixation differs in length and could be several seconds long. Holmqvist et al (2011) stated that when a fixation is measured, the actual measure you are after 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 of the eye. From these recordings, the raw data, the movements of the eye, are categorized as fixations, saccades, and blinks by the eye-tracker. 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, and the spatial relations between the pupil and the corneal reflections center. As mention above, the close coupling with visual attention and eye-movements give grounds for using eye-tracking to capture underlying cognitive processes (Deuble, Schinerder 1996; Corbetta et al., 1998; Holmqvist et al 2011). From the recorded eye-movement, a lot of measures can be calculated 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, and 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, 2011). The tower-mounted systems often fixate the head with a chinrest as well as present the stimulus on a screen in front of the participants. Therefore the tower-mounted system can have a higher sampling rate, because it can assume the eye is captured by the camera instead of having to utilize 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 you move around and look at different things in an environment such as the supermarket. Therefore the head-mounted system struggles with defining the stimulus and to accurately calculate the gaze estimation. The recorded eye movements have to often be manually coded on to a gaze map frame by frame which is a tedious process. The remote eye-tracker´s sampling rate is slightly lower than the immobile tower-mounted systems and higher than the mobile head-mounted systems. When attached to a monitor it could be used out in the field as well as in a laboratory setting. For good measure the participant has to be around 70 cm away from the eye-tracker and there is a possibility to move the head a little bit without losing connection to the eye-tracker (Holmqvist, 2011). An advantage of the remote eye-tracker is that it is not as intrusive and apparent that the eyes are being recorded as if the participant has to wear a helmet or rest their chin on a chinrest.

Eye-tracking has diagnostic role because it provides an objective and quantitative measure of the eye-movements which could reveal the underlying cognitive processes (Duchowski, 2002). As a tool, eye-tracking can be used in a 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 investigate reading exercises. In contrast to, mathematical tasks were little progress has been made in eye-movements research. One of the few previous studies on eye-tracking with mathematical task is Green et al (2007) who conducted a study on adult’s eye-movements when adding three-digit numbers and they showed the use of strategic eye-movements. Another study is Schneider et al (2008) who presented a validation of eye-movements as a measure for developing number sense in children. They examined an exercise called mental number line (mention above) 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 measure of developing number sense both have validity and utility (Schneider et al, 2008). Hitherto, no one has captured eye-movements during dynamic mathematical exercises such as the unexpected event. Using eye-tracking to capture the noticing of an unexpected event in Magical Garden would then be pioneering work.

What type of eye-movements would then be indicative of noticing? As mentioned above, the method of eye-tracking provides a number of different possible measurements. In the next section, I will introduce an unexpected event as an experimental manipulation to Magical Garden and noticing it would be a way of sieving out the different levels of num-

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

ber sense in children. Thereafter, a novel model of detection will be proposed which could be captured by eye-tracking.

Unexpected event

What is "the unexpected event" and could noticing it expose children´s level of number sense? An unexpected event has to contain some crucial elements in order for noticing the unexpected event to be indicative of a developed number sense. Firstly, it has to evoke a response from the children of a specific group; in this case only the children who fully understand the exercise (have a developed number sense). Secondly, it has to be measurable. Thirdly, the way of noticing the unexpected event has to have a meaningful and logical explanation tied the behavior. In this study, the unexpected event is incorporated in the mini-game Bird Hero. The unexpected event is that:

The elevator continues past the intended branch and gets stuck in the tree top[3]

During regular gameplay the elevator always moves, as expected, to the level chosen by the player. The expectation that the elevator will move to the chosen level is created by the top-down system. The expectation depends on the knowledge of which number the baby bird is showing with its feathers, the lift button´s number and the branch of the tree they correspond to. Therefore the unexpected event is only unexpected for those who expected the elevator to go to the branch corresponding to the elevator button. There are no bottom-up salient or unexpected stimuli exposing the unexpected event before it gets stuck in the tree top. The only thing that changes is that the elevator keeps going up. If a player does not know the elevator was supposed to stop at a specific branch, it is still moving towards the branch when it in fact has already passed the chosen branch. This separates the ones that know from the ones that do not understand or have sufficient developed number sense. Noticing, therefore evoking a response would indicate a correct answer. Not noticing, therefore not evoking a response would indicate an incorrect answer. The children will understand that something went wrong when the unexpected event is revealed as 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 and comes 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 and the elevator moves down to the correct branch. The reasons for adding the service bird at the end of the unexpected event was primarily as a precaution so the child will not get upset or wonder if the game is broken and explicitly say that something went wrong. With the unexpected event the children are given the opportunity to notice/detect the manipulation before it reaches the tree top.

[3] 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 and the player must try again.

Model of detection

Where will the children look if they notice the manipulation? As previously proposed a detection of the unexpected event could be manifested as increased visual attention. This increased visual attention must be towards an area, or as a fixation on an AOI, that has logical and meaningful explanation as to 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 has passed the correct branch.

The reasoning behind this proposal was that attention would be given towards the correct branch during an unexpected event only if it is directed there by top-down attention such as previous knowledge (knowing which branch was the correct branch). Children who do not have the knowledge of which the correct branch is (either a lack of working memory or number sense) will not notice that something is wrong with the elevator, and should not have any reason to perform "look back". They would most likely follow the elevator and wait for it to stop. The movement of the elevator would be the most bottom-up salient object in the scene and therefore draw the most attention (Corbetta, Shulman, 2002). According to Henderson (2003) even though the movement of the elevator it is a strong salient feature, visual salient objects are be less attended to during an active task on a meaningful scene. In this case, the unexpected event is a meaningful scene to the child if the child knows which branch was the correct one. Also, noticing the unexpected event would be considered an active task which would then negate the salient feature of the moving elevator. It should not be counted as a "look back" if the child is looking at the correct branch when the elevator is on the correct branch and then continues looking at the branch when the elevator has passed. The child needs to have moved her eyes from the branch in order to make a "look back" during the unexpected event. The model is then that the children will attend the correct branch with a "look back". The proposed model is novel and although the reasoning behind the model seems sound it has to be tested. By testing the model it will be determined whether it is a good way of measuring developing number sense. It certainly could be the case that this proposed model is not the best model for testing noticing of an unexpected event. Therefore, in addition to the proposed model, other eye-movements will be collected and analyzed such as looking in anticipation and looking at the buttons. 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 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 iPads at their preschools during dedicated sessions a number of times each week for three weeks. Every mini-game played on the iPads was logged and the 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 played during the eye-tracking experiment alternated between containing an unexpected event and not containing an unexpected event (control condition). The children played the game at least five times and a maximum of ten times. The 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 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 of concerning procedure, experimental design, as well as the exact measurements follows in the method section.

Research questions

(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?

(H3) From an explorative perspective examine the measured eye-movements; see how they correlate with each other and with performance?

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. 42 children (21 girls, *M=4.6*, *SD=0.72*) participated in this study.

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 screen a 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 iPads. An experimenter introduced the game to the children, one at the time. The children played the game individually on iPads at Kindergarten or school 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 iPads to children during the dedicated lessons. Each child logged in on an individual account with 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 Kindergarten had several iPads 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 and 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 respectively 12 children). The average duration of the eye-tracking experiment including calibration was 30 minutes. A teacher was present during the experiment for some children, sitting on a chair a few meters away, but most of the experiments were conducted alone with the experimenter. To start the experiment a calibration and validation of both eyes were made. 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 saying that looking at the dot and following it was the password unlock the game and to start playing. The experimenter gently held the head still of children who were unable to keep their head still, to ensure an acceptable calibration. The calibration method was an 8-point with a black circle on silver background. The calibrations points were accepted by the experimenter hitting space bar. The validation was the standard 4-point with a white dot on a silver background and accepting calibration points automatic when fixation was stable. In this study a re-calibration was performed if a child deviated more than 1.7° on either eye. 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. 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 motivation 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 when child in charge and TA watching, (1b) experimental condition when child in charge and TA watching, (2a) control condition when TA in charge and child watching, and (2b) experimental condition when TA in charge and child watching (see table 1). The tasks were in the same order for each trial (1a, 1b, 2a, 2b).

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

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 the (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 branch indicated by the pressing of the button. 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 occur, the elevator moving passed the chosen branch and crashes in the tree top. The thought behind adding the service bird at the end of the unexpected event was primarily a way of not frightening the children and explicitly say that "something went wrong ", as a precaution so the child won't get upset or wondering if the game is broken. This would also act as a cut off point for the time frame of some AOIs. If the child answered incorrectly the standard feedback will be given. In (2a) and (2b) the TA was in charge and 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 control condition (2a) the elevator would move to the intended branch. In the experimental condition (2b) the unexpected event would occur. After completing the 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 press the iPad´s screen to indicate the buttons during the iPad 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 as a break to preserve energy. Most of the children wanted to completely fill the watering can, complete two trials, so they could water more once they water. After completing at least five trials the children were thanked for participating and given a diploma.

The 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 in the original version (1:2014-11-26). The first task (1a) was the same as game mode (ii) expect 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 in Paint.net.

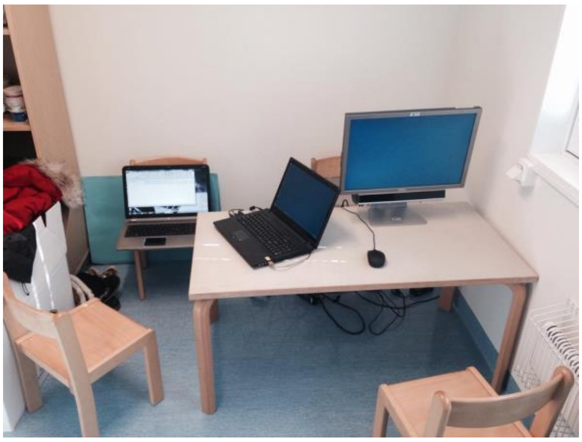


Figure 3. Picture of the eye-tracking set-up at the preschool.

Measurements and data analysis

In this study two performance measures were constructed, one from the iPad sessions and one from the eye-tracking experiment for each child. As in the procedure and design of the experiment, all of game data from the iPad sessions was logged. A mini-game consisted of the three gameplay modes (i, ii, and iii). 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. Therefore, the total performance on Magical Garden for each child was calculated from the total number of correct answers divided by number of tries from all of the finished mini-games during the iPad session. The performance measurement from the games played on the special version of Magical Garden during the eye-tracking experiment was based on the performance of the child from exercise 1a and 1b. The number of tries per trial were two plus the number of incorrect answers. Total performance was therefore calculated from the total number of correct answers divided by the number of tries. 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. Therefore it had to be manually calculated from the recorded videos of the experiment.

The eye-tracking measurement that was used was AOI hit. Holmqvist defined an AOI hit as following: "AOI hit, which states for a raw sample or a fixation that its coordinate value is inside the AOI" (Holmqvist, p. 189, 2011). Having AOI hit as the measure for attending the AOI is sound because the interest is if the child makes a "look back" or not, or if the child uses their top-down attention to indicate that they have noticed the unexpected event or not. Therefore, a common measure like dwell time on the AOI is not used. 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 e.g. fixation durations summarized by Rötting (2001; in Holmqvist, 2011) ranged from 60-120ms and Granka et al (2008; in Holmqvist, 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 will reduce the variance of the sample and the AOI hits that are recorded are times when the child has attended the AOI. As mention 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. Also in the table the exact location of the AOIs on the screen are presented, and their specific time frame. A "look back" (AOI 2) was the proposed as the model of detection of the unexpected event. The reason the unexpected event only was in the range 1-6, was that the children should have time to make a "look back" or another eye-movement before the elevator reached the tree top.

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

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 is a miss (coded as a 0). For example during exercise 1b of the first trial, the AOI "look back" could either have a hit or miss. Missing data was only coded as NA, when the gaze cursor was flickering or moving all over the screen thus give an incorrect and not reliable data. If missing data (no flickering and the gaze cursor had been missing from the screen) was due the child not looking at the screen it was coded as a miss (0), the same as not looking at the AOI. The reason for coding missing data (when the child was not looking at the screen) as 0 instead of NA was that 0 should be the same as not attending the AOIs. If the child is not looking at the screen she is not attending the AOIs and if missing data would have been coded NA that information would have been lost and the results would have indicated a higher level of attention 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.

Table 2. Displays 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[45].
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 arrive at the correct branch.

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 way the fixations was coded in this study was conducted by manually by looking frame by frame and reported if the gaze cursor stays "still" fixated on a AOI (see above for the fixation level). Holmqvist (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, 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 have had to be added to every participant because they 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 and add 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, accurate measurement on dwell time, the measurement used in this study AOI hits allowed a 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 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 ration presented in Table 3 is a bit misleading it is the tracking ratio of the whole experiment. The children attended the monitor during the unexpected event to a larger extent than they did during the rest of the experiment.

[4] The definition of "one branch under the correct branch" is when the number on the branch is showing on the bucket.
[5] 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' 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 iPad	10.35 (3.81)
Performance iPad (%)	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)

Looking at the descriptive statistics, from Table 3, the iPad 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 iPad sessions. The performance from the iPad 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 iPad 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 performance form the iPad session could predict the performance in the eye-tracking experiment. There was high performance overall. A possible celling effect could be seen with nine children scoring 100% on the eye-tracking version of Magical Garden (see figure 4).

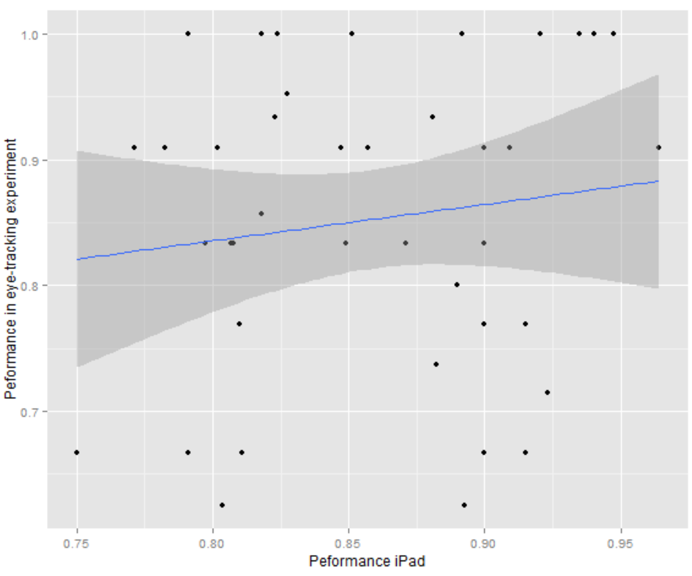


Figure 5. Scatterplot of performance from the iPad sessions as a function of the performance in the eye-tracking experiment (n = 40). A line was added to display the regression, the shadow represent the standard error.

Research question 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).

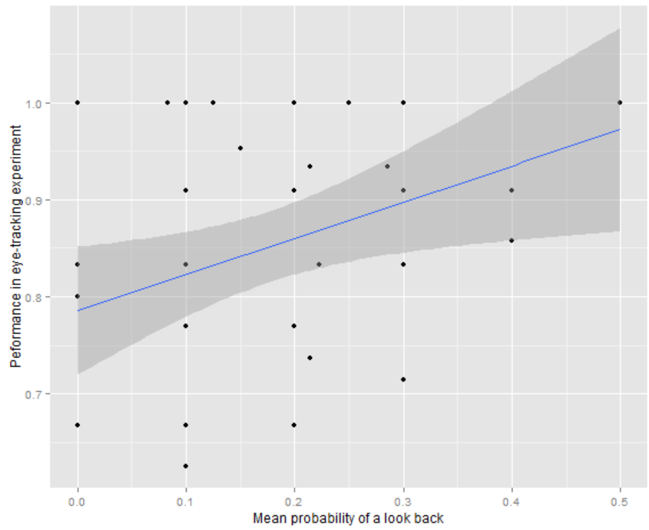


Figure 6. Scatterplot of performance from the eye-tracking experiment as a function of the mean probability of a "look back" (n = 40). A line was added to display the regression (0.37), the shadow represent the standard error. Dots may overlap.

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". The reason for this was that every level should then have the same length of time frame. 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.

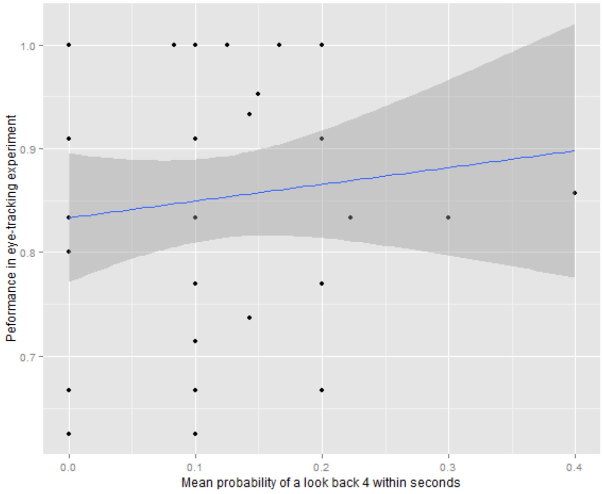


Figure 7. Scatterplot of performance from the eye-tracking experiment as a function of the mean probability of a look back within 4 seconds (n = 40). A line was added to display the regression (0.16), the shadow represent the standard error.

Research question 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 0.244 (*SD*= 0.262), which means that the TA was attended to in a fourth of the trials. Figure 8 shows overall the mean look at TA when it is divided 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 0.172 (*SD*= 0.2) and when the TA was in charge the children the mean looked at the TA was 0.316 (*SD*= 0.29) (see figure 8).

In order to negate the possibility of a few children alone increasing the total mean, the 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 0.143 (*SD* = 0.248), 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].

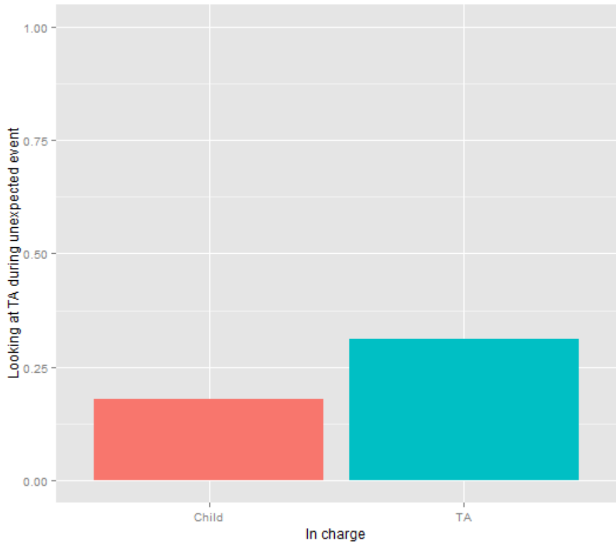


Figure 8. Bar plot of mean number of looking at the TA during an unexpected event combined for all participants (n=40), divided up by in charge (Child or TA).

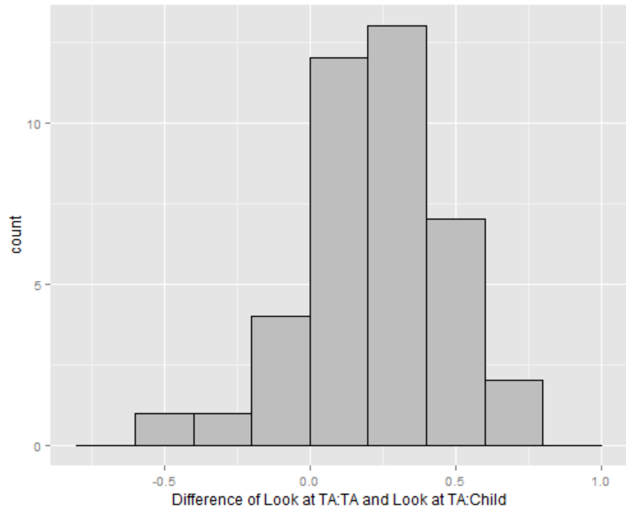


Figure 9. Histogram of all the participants (n=40) the difference of mean look at TA when TA was in charge and mean look at TA when the child was in charge.

Research question 3

From an explorative perspective examine the measured eye-movements; see how they correlate with each other and with performance?

In order to look at how 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 presented in Table 4.

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

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 ranging between 4.5–6.5 years of age if they noticed an unexpected event when playing the educational game Magical Garden with a teachable agent. Magical Garden was argued to be a suitable operationalized exercise for testing number sense. The unexpected event was designed to only call attention from the children who had a sufficient level of number sense in order to understand that an unexpected event had happened. The children had been familiarized with Magical Garden (without the unexpected event) by playing it on iPads for three weeks before the eye-tracking experiment. A model for noticing the unexpected event was proposed, namely that the children would look back at the correct branch when an unexpected event occurred. The first research question H1, hypothesized that a looking back would predict performance. The second research question H2 asked whether children would even attend the TA during an unexpected event. The final research question was to investigated, from an explorative point of view, how and if the captured eye-movements measures correlated with each other and the performance measure.

The results were the following. Concerning research question H1, the results showed that the model proposed for noticing an unexpected event, 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 research question (H2), the children did look at the TA during an unexpected event and there was found 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 then if the child was in charge. In the third and explorative research question (H3), a strong correlation ($r = .5$) was found between "look at the button "and performance in the eye-tracking experiment. Additional medium correlations was found between "look at button" and "anticipate" ($r = .41$), as well as between "look at TA" and "look at button" ($r = .31$), and between "look at TA" and "look back 4 sec" ($r = .39$).

Are the results reasonable and how should they be interpreted? Firstly, concerning H1, the model "look back" without the four second limit showed that there was a significant correlation with performance. This indicates that the model of detection was not completely off target. However, the "look back" model does not seem to explain everything that happened when the child noticed an unexpected event. When the four second limit was added to the "look back" model, 35% of the look backs was now not defined as look backs and no significant correlation was found with performance. A reason for this result could be that there was not enough power left to break through the looming celling effect of performance. The effect of the four second limit meant that 35 % of the originally defined look backs happened after four seconds. One question that arose during the analysis of the data was why the children performed a "look back" sometimes only after four seconds? One reason for this could be that the children discovered that something was wrong only when the elevator was a great enough the distance away from the correct branch, and were therefore were too late to be considered a "look back" with the four second limit. Another explanation for why the children performed a "look back" after four seconds might be that looking back at the correct branch is not the only way to indicate detection of the unexpected event. During the eye-tracking experiment, both verbal report and non-verbal detections of the unexpected events was common, although they were not officially documented because the experimental design was focused on eye-movements. The most common non-verbal response to the unexpected event was for the child to turn and look at the experimenter. When the child looked at the experimenter, the eye-tracker could not track the eyes which lead to a decrease in documented detections of the unexpected event. This could be an explanation to why some "look backs" occurs first after four seconds. The correlation analysis conducted across all collected eye-movements showed that the child looking at the button correlated strongly ($r=.5$) with her performance. This further indicates that the children noticed and detected the unexpected events in different ways. The correlation analysis was not adjusted for multiple tests. From an explorative point of view using the results, from the not adjusted correlation analysis, in search of a new model for detection this is not an issue. One would have to conduct another experiment to test the new model. It makes sense logically that looking at the button should positively correlated with performance, because the chosen button contains information about which branch the elevator was supposed to go to. In this context, the child looking at the button makes sense if they start to suspect that something is not right. Also, anticipating eye-movements correlated with 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 suspicion of that something is wrong make sense. The fact that preschoolers even make anticipatory eye-movements is worth emphasizing. It is harder to explain why there should be a positive correlation between “look at TA” and “look back 4sec”/”look at button”. From the results of H1 and H3 it seems that future research have been given a foundation to build a new model of detection constructed by verbal and 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 that much (less 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 and therefore the child could be said to display ToM capabilities. These results are in accordance with the conclusion of Haake et al (2013) that preschoolers are able to engage in social interaction with a TA.

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 in at the children’s preschools not in 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 had have 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 then 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 decreases the tracking ratio. If the children would have been recorded with a tower mounted system the tracking ratio may had been better. Although, the willingness to participate and the number of trials the children would have been able to complete would have plummeted, because of having to stay still. The off-task, watering the garden, was a motivational boost for the children, 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 head and not being able to sit still. Trueswell (1999) 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 on the child before recording. In this study, gently holding the children’s head during calibration had a positive effect and decreased the deviation and possibly also decreased the participant attrition, which is something future research might want to adapt.

The way the performance measure from the iPad session did not correlate with the performance from the eye-tracking experiment. The reason for this might be that, the performance measures 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 a better or more representative way of calculating a longitudinal measure of number sense would have been progression instead of average performance over all the exercises.

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

If I would have had the opportunity to re-design the study, a lot of different experimental design choices would have been made. Firstly, I would not have used randomly generated correct answer for the exercises, instead have all the children play the same amount of exercises with branch one as the correct answer, branch two etc. The order in which the exercises would have been presented could have been counterbalanced with a latin square. This would have also given me the opportunity to analyze and control for if “look backs” or the preferred model of detection occurred more often on some branches than others. Secondly, have the control exercises use the same number range as the manipulation (1-6). Thirdly, have all the children play the same amount of trials. Fourthly, the present design does not control for WM, so a WM test should have been added. In the present study there could be children with WM difficulties who does not react to the unexpected event, yet have a high level of number sense. Fifthly, a ToM test should have been included in order to control for that it is not only children who have fully developed ToM capabilities that are the ones who look more at the TA when the TA is in charge. Sixthly, record the verbal and non-verbal detections, in order to create a model of detection. Finally, I would not have coded the fixations and AOI hits manually, and tried to implement dynamic AOIs in order to process the eye-tracking data with a script instead.

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 explain the whole notion of detecting an unexpected event. However, this study emphasizes that a better model of noticing could be constructed by combining 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. With simple tricks of keeping the interested conducting an eye-tracking experiment on preschoolers is not as hard as it sounds. Using an educational game to operationalized exercises testing number sense worked overall well. It also had the additional advantage of the increasing the children´s motivation and willingness to participate in the experiment. Even though, the children do not attend the TA that much during an unexpected event, this study showed that children attend the TA more when the TA is in charge and this could be interpret 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 has just started.

Acknowledgements

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

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

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

Thanks also go to Magnus Haake, Emil Boman and Rasmus Bååth 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 proof reading and Fanny Johansson for supporting me at 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.

Berch, D. B. (2005). Making sense of number sense implications for children with mathematical disabilities. *Journal of learning disabilities*, *38*(4), 333-339.

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

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.

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

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.

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

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.

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.

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 learning1. *Applied linguistics*, *11*(2), 129-158.

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.

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.

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.