

Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge

Tim van Marlen^{a,*}, Margot van Wermeskerken^a, Halszka Jarodzka^{b,c}, Tamara van Gog^a

^a Department of Education, Utrecht University, The Netherlands

^b Welten Institute, Open University of the Netherlands, The Netherlands

^c Humanities Laboratory, Lund University, Sweden

ARTICLE INFO

Keywords:

Example-based learning
Eye tracking
Attention cueing
Eye movement modeling examples

ABSTRACT

Eye movement modeling examples (EMME) are video modeling examples with the model's eye movements superimposed. Thus far, EMME on problem-solving tasks seem to be effective for guiding students' attention, but this does not translate into higher learning outcomes. We therefore investigated the role of ambiguity of the verbal explanation and prior knowledge in the effectiveness of EMME on geometry problems. In Experiment 1, 57 university students observed EMME or regular video modeling examples (ME) with ambiguous verbal explanations. Eye-tracking data revealed that –as in prior research with unambiguous explanations– EMME successfully guided students' attention but did not improve test performance, possibly due to students' high prior knowledge. Therefore, Experiment 2, was conducted with 108 secondary education students who had less prior knowledge, using a 2 (EMME/ME) × 2 (ambiguous/unambiguous explanations) between-subjects design. Verbal ambiguity did not affect learning, but students in the EMME conditions outperformed those in the ME conditions.

1. Introduction

Video modeling examples in which a model demonstrates and explains how to perform a task (e.g., “how to” tutorial videos on YouTube), are widely used in formal and informal learning settings. Such videos lie in the tradition of example-based learning, which is an effective and efficient way of learning, provided that the examples are well-designed (Renkl, 2014; Van Gog & Rummel, 2010). It has been proposed that, depending on the task at hand, the design of screen-recording video examples could be further improved by showing learners what the model is looking at, by displaying the model's eye movements (Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). Displaying a visualization of the model's eye movements (e.g., fixations represented as a circle or dot) is expected to guide learners' attention to what the model is looking at in that moment, which should make it easier to understand and learn from the demonstration and verbal explanation. Several studies have found beneficial effects of such “eye movement modeling examples” (EMME) on attention guidance and found enhanced learning of classification tasks (Jarodzka et al., 2012; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013) and enhanced integration of text and pictures during reading (Mason, Pluchino, & Tornatora, 2015a; 2015b). Thus far, however, EMME on problem-solving tasks seem to be effective for guiding learners' attention, but this

does not translate into higher learning outcomes (i.e., higher performance on the test problems) compared to the no EMME control condition (Van Marlen, Van Wermeskerken, Jarodzka, & Van Gog, 2016). This discrepancy between the studies regarding the effectiveness of EMME might be related to the extent to which the verbal explanation accompanying the EMME is clear (i.e., unambiguous) to the participants. For instance, studies using classification tasks in which EMME were found to be effective (Jarodzka et al., 2012, 2013), specific jargon was used; when learners do not yet know the jargon, this increases the usefulness of visual guidance. In contrast, in a study on learning to solve a puzzle problem, it was clear from the verbal explanation what object the model was looking at, and the visual guidance provided by EMME was not useful for learning (Van Gog et al., 2009). The present study addressed two potential explanations for this lack of effect of attention guidance on learning procedural problem-solving from EMME: ambiguity of the verbal explanation and prior knowledge.

1.1. Eye movement modeling examples

Multimedia materials provide a combination of verbal and pictorial information, which according to the *dual-coding theory* (Clark & Paivio, 1991; Paivio, 1986) are processed in separate auditory and visual channels. According to the *Cognitive Theory of Multimedia Learning*

* Corresponding author. Department of Education, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, The Netherlands.

E-mail address: t.v.a.vanmarlen@uu.nl (T. van Marlen).

(Mayer, 2014), learners first need to attend the relevant verbal and pictorial information (selection). After selecting the relevant verbal and pictorial information learners *organize* this information into coherent mental representations, and *integrate* the verbal and pictorial mental representations with each other and with available prior knowledge (Mayer, 2014). In dynamic learning materials like video modeling examples, one challenge for the selection of information lies in the transience of the material. If the learner does not attend to the right information at the right moment, it is no longer available for processing (i.e., organization and integration) and learning is hindered (Ayres & Paas, 2007). One reason why learners might not be able to attend to the right information at the right time in a video modeling example, is that it is likely that there is a discrepancy in what the expert model and the novice learner are attending to, and that the verbal explanation provided by the model may not be sufficiently clear to rapidly guide the learner's attention to what the expert is referring to.

The discrepancy between experts' and novices' attention allocation has been shown in different eye-tracking studies. Experts often attend to task-relevant information relatively longer and faster while paying less attention to task-irrelevant information than novices (Charness, Reingold, Pomplun, & Stampe, 2001; Van Gog, Paas, & Van Merriënboer, 2005; Van Meeuwen et al., 2014; Wolff, Jarodzka, Van den Bogert, & Boshuizen, 2016). This expertise effect has also been demonstrated within participants as a result of task experience (Blair, Watson, & Meier, 2009; Canham & Hegarty, 2010; Haider & Frensch, 1999; Hegarty, Canham, & Fabrikant, 2010). This difference in attention allocation might cause learners to miss the information the model is attending to, unless the model's verbal explanation would be sufficiently clear to rapidly guide learners' attention to the right information at the right time.

It has been proposed that one way to reduce the discrepancy between the model's and the learner's attention allocation would be to show the learner what the model is attending to, by displaying a visualization of the model's eye movements (e.g., as a dot or circle; Van Gog et al., 2009). In such eye movement modeling examples (EMME), the learner is not only shown how the model is performing the task (by means of a screen recording of the model's computer screen), but also where the model was looking while performing the task. By showing the eye movements of the model the visual attention of the learners is guided and synchronized with the model to create a state of *joint attention* (i.e., *joint attention* is the phenomenon characterized as automatically looking at an object someone else is looking at; Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Frischen, Bayliss, & Tipper, 2007), thus helping the learners attend to the relevant information at the right time which, in turn, can be expected to improve learning.

However, findings regarding the effectiveness of EMME for learning are mixed. Whereas some studies have found beneficial effects on learning classification tasks (Jarodzka et al., 2012, 2013) and learning text and picture integration (Mason et al., 2015b, 2015a), EMME in which problem-solving tasks are demonstrated seem to be effective for guiding learners' attention (Van Marlen, Van Wermeskerken, Jarodzka et al., 2016), but this does not translate into higher learning outcomes (Van Marlen, Van Wermeskerken, Jarodzka et al., 2016; see also; Van Gog et al., 2009). One possible reason for these mixed findings might lie in the extent to which the model's verbal explanation is sufficient to rapidly guide learners' attention to what the model is referring to. EMME might be most effective for learning when the model's verbalizations are ambiguous.

1.2. The role of verbal ambiguity in the effectiveness of EMME

When the verbal explanation in a modeling example contains ambiguous verbal referents, it will not be immediately clear to the learner what the model is referring to. Ambiguity of verbal referents can originate from different sources. For instance, experiments in cognitive science have shown that ambiguity can arise due to the visual context

(e.g., multiple objects that the referent might refer to; Allopenna, Magnuson, & Tanenhaus, 1998; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995), the lack of specificity of the verbal information (e.g., referring to a target without verbal location descriptions; Louwerse & Bangerter, 2010), or a lack of prior knowledge (e.g., lack of background knowledge about a to be inspected and discussed painting; Richardson, Dale, & Kirkham, 2007; Experiment 2). When verbal referents are ambiguous for any of those reasons, it will take listeners more time to locate the relevant (i.e., target) information, if they are able to locate it at all (Louwerse & Bangerter, 2010; Van Marlen, Van Wermeskerken, & Van Gog, 2018).

These studies about the effects of verbal ambiguity on the speed and accuracy with which referents are located, suggest that the attention guidance provided by EMME might be most needed and most effective for learning when the model's verbal explanation is ambiguous for learners. Providing clear verbal explanations might not always be possible depending on the task and source of the ambiguity. For instance, in classification tasks providing a clear verbal description of a complex visual shape denoted by a jargon term, might be quite difficult. In the classroom, teachers/instructors can resolve this problem by using available non-verbal cues (e.g., looking or pointing at the part of the task they are discussing) that will disambiguate their verbal message. However, in digital video instructions these non-verbal cues are not necessarily present. Thus, it is likely that verbal explanations in some circumstances are not sufficient and have to be accompanied by non-verbal cues that align the learners' attention with that of the model. EMME do this by showing the learner what the model is looking at, at any given moment, which may resolve potential ambiguities in the model's verbal explanation. Hence, the discrepancy in results regarding the effectiveness of EMME might be due to whether verbal explanations are perceived as ambiguous without further guidance of an EMME. Indeed, there is some tentative evidence suggesting that this is the case: It is likely that verbal referents were ambiguous for the learners in the studies that found positive effects of EMME on learning classification tasks (Jarodzka et al., 2012, 2013; Vitak, Ingram, Duchowski, Ellis, & Gramopadhye, 2012).

For instance, in one study learners had to learn to classify fish locomotion patterns and were shown videos of different fish while an expert gave verbal explanations about their locomotion pattern (Jarodzka et al., 2013). In his verbal explanation, the expert was using terms like 'the dorsolateral fin', which can be ambiguous when the learner does not yet know what that is. In this case, seeing the expert's eye movements (i.e., seeing what he is looking at) would help to attend to the right information at the right time. Similarly, in the study by Vitak et al. (2012) learners had to classify cells on histological slides. The learners were shown video examples with or without the expert model's eye movements superimposed onto the example slides while listening to a verbal explanation. The expert referred to certain cells with terms like "there's" or "this one". Results indicated that the expert's eye movements were helpful in disambiguating the verbal referents as indicated by fewer classification errors on subsequent test tasks. Also, search behavior was more efficient for learners in the EMME condition as those learners needed less time to classify the cells and made fewer fixations on subsequent test tasks.

In studies on problem-solving tasks, in which the verbal referents were likely unambiguous (e.g., referring to "angle A" of a geometry problem), EMME had no beneficial effects on learning (Van Marlen, Van Wermeskerken, Jarodzka et al., 2016) or a negative effect (Van Gog et al., 2009). In the study by Van Marlen, Van Wermeskerken, Jarodzka et al. (2016), participants were shown modeling examples with or without the model's eye movements superimposed while also hearing verbal explanations about how to solve geometry problems (Exp. 2). Although EMME were effective for *more rapidly* guiding attention towards the information the model referred to (i.e., shorter times to first fixations and –probably as a consequence– longer fixation of referents), there was no difference in how many referents were fixated between the

EMME and regular modeling example condition. This suggests that the verbal information was already sufficient in guiding the visual attention towards the relevant information, which might explain why EMME did not result in better performance (Van Marlen, Van Wermeskerken, Jarodzka et al., 2016). That is, when the verbal referents are sufficient to follow and understand the model's demonstration and explanation, the attention guidance provided by the eye movements would be unnecessary, and research regarding the redundancy effect shows that displaying unnecessary information does not enhance learning but instead can be detrimental for learning (Kalyuga & Sweller, 2014).

In sum, the (un)ambiguity of the model's verbal explanation might determine the (in)effectiveness of EMME compared to regular video modeling examples in which the model's eye movements are not visualized. Therefore, Experiment 1 investigated whether verbal ambiguity indeed plays a role in the effectiveness of EMME.

2. Experiment 1

To investigate whether EMME would not only be effective for attention guidance but also for learning when the verbal explanation is ambiguous, Experiment 1 replicated the study design by Van Marlen, Van Wermeskerken, Jarodzka et al. (2016) but using EMME in which the referents were made more ambiguous (i.e., in the sense that they were less specific; e.g., “this angle” instead of “angle A”). A pilot study (Van Marlen, Van Wermeskerken, Boven et al., 2016) showed that participants studying ambiguous EMME fixated the referred information more often than participants studying ambiguous modeling examples (ME) without the model's eye movements and there seemed to be a (non-significant) trend in test performance in favor of the ambiguous EMME condition. Based on the eye-tracking results of both the pilot study and previous studies demonstrating that EMME can guide attention (Jarodzka et al., 2013; Van Marlen, Van Wermeskerken, Jarodzka et al., 2016; Mason et al., 2015b, 2015a), we hypothesized that ambiguous EMME would guide learners' attention (H1), resulting in verbal referents being fixated more often (H1a), faster (H1b), and longer (H1c) than in the ambiguous ME condition. In addition, we hypothesized that the attentional guidance would also foster learning, resulting in higher accuracy (H2a) and faster solving speeds (H2b) on test problems in the ambiguous EMME than in the ambiguous ME condition.

2.1. Methods

2.1.1. Participants and design

Participants were 57 Dutch university students, enrolled at a Faculty of Social Sciences ($M_{age} = 21.98$, $SD = 2.82$; 18–33 years, 21 male). They were assigned to either the EMME ($n = 28$) or ME condition ($n = 29$) at random. Participants had normal or corrected to normal vision and received a €5.00 reward for their participation.

2.1.2. Materials and apparatus

2.1.2.1. Eye tracking equipment. A SMI RED250 eye tracker (SensoMotoric Instruments, GmbH) with a sampling rate of 250 Hz was used to record the eye movements of the participants. The experiment was created and presented with the software package Experiment Center 3.4.165 (SensoMotoric Instruments, GmbH). The experiment was run on a monitor with a resolution of 1680×1050 pixels and a refresh rate of 60 Hz. Within the Experiment Center software the screen recording function was used for measuring the participants' eye movements during example study (i.e., EMME or ME) and problem solving. ActionScript 3.0 was used to program the geometry problems and Flash Adobe CC was used for the presentation of the geometry problems.

2.1.2.2. Pretest. To check whether there were no prior knowledge differences between conditions a pretest was administered. It

consisted of two open questions (e.g., Angle A is equal to: ...) and three multiple-choice questions about a geometry figure. The questions tested knowledge of corresponding angles, alternating angles, straight lines, and triangles.

2.1.2.3. Geometry problems. The program *Geogebra* (www.geogebra.org) was used to create the geometry problems. The problems consisted of line drawings (i.e., black lines on top of a white background) of triangles and parallel lines which combined different geometrical principles regarding parallel lines (i.e., Z-rule or alternating angle rule; F-rule or corresponding angle rule). Above each geometry figure the problem statement was provided (e.g., “How many degrees is angle B?”), which required four angles (i.e., four sub steps) to be solved. All problems contained 13 angles coded A, B, C, etc. The values of some angles were provided in the problem and unknown values of angles were indicated with a question mark. Two geometry problems were created for the modeling examples together with two isomorphic versions of these problems (i.e., exact same layout of the line drawings but with different numbers). Additionally, four transfer problems were created which had different visual layouts and numbers than the isomorphic problems. Due to the different visual layouts of the transfer problems, participants had to search for the correct starting point and order to solve the problems; for the isomorphic problems, in contrast, participants had observed the correct starting point and order in the modeling examples. To facilitate calculations (and to prevent participants from making errors), a digital calculator was placed right next of the problem statements for all geometry problems. For all problems, the rectangular area of the line drawings containing the problem statement, the calculator and the geometry figure had a width ranging from 1092 to 1351 pixels and a height ranging from 787 to 847 pixels (see Fig. 1).

2.1.2.4. Video modeling examples. SMI Experiment Center 3.4.165 was used to present the experiment and to record the eye movements of the model for the modeling examples. SMI BeGaze 3.4.52 was used to visualize the eye movements and to create the video examples. In both conditions a female model explained verbally how to solve the problem, using ambiguous references to the angles e.g., “Now that you know *this* angle, you can calculate the *other* angle.” (see Appendix for a screen capture of a video modeling example along with the transcript of the verbal explanation). In all modeling examples the model began by searching for the location of the angle mentioned in the problem statement. Once the angle was localized, the model started working backwards until a starting point to solve the problem was found. Then, the model started to determine the angle of the starting point and explained each solution step until the angle of the problem statement was determined. Both conditions were provided with the same verbal explanations along with the answers to each solution step (i.e., angles) that became visible during the problem-solving process. In the EMME condition, participants additionally saw the model's eye movements superimposed onto the modeling example. The eye movements were visualized as a blue translucent dot with a 30-pixel diameter and were created using the Bee Swarm utility in SMI BeGaze. The length of the modeling example videos was the same across conditions (with one video lasting 122 s and the other video lasting 131 s).

2.1.3. Procedure

All participants were tested in individual sessions of ca. 25 min. As participants entered the lab, they were briefly given a general overview of the experiment. Then participants were presented with the pretest. Once the participants finished the pretest they were seated in front of the eye tracker at approximately 57 cm distance of the monitor by means of a forehead and chin-rest. After a five-point calibration (including four-point validation) procedure, participants were instructed that they were about to see video examples about how to solve

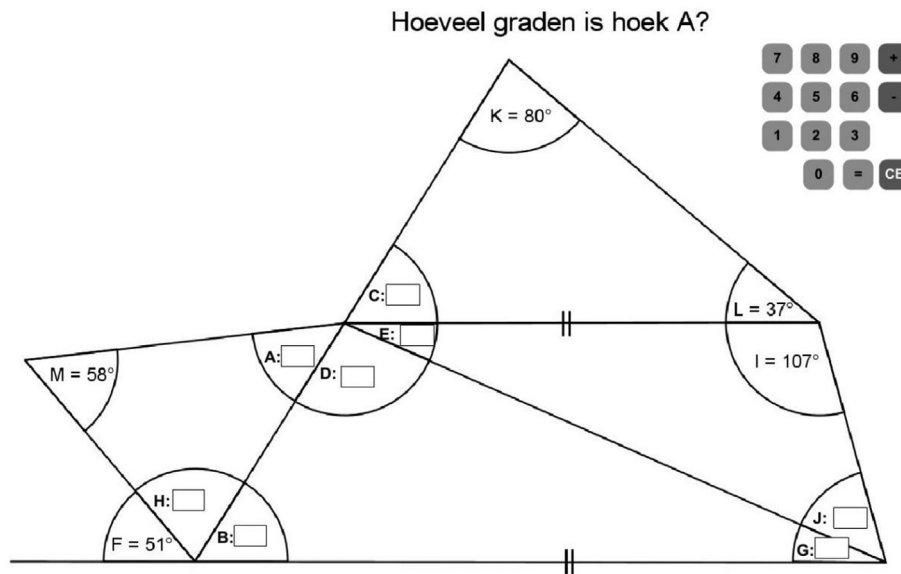


Fig. 1. An example of a geometry problem used both in Experiment 1 and Experiment 2. On top is the problem statement, (translated from Dutch) “How many degrees is angle A?”.

geometry problems and that each video demonstrated the correct solution to the geometry problem. In the EMME condition, participants were also instructed that they would see the model's eye movements, visualized as a moving blue translucent dot. To familiarize participants with such a visualization of a model's eye movements, participants in the EMME condition were shown a short EMME illustration video, in which the model was inspecting an image of a living room. Then, participants of both conditions were instructed that they would be presented with a similar problem after each video example and that they should solve this problem as rapidly and as accurately as possible. After these instructions, participants were presented with the two example-problem pairs. Subsequently, after re-calibration of the eye tracker, participants solved the four transfer problems. The order of the video examples with the matching isomorphic problems were counterbalanced across participants while the transfer problems were presented in a fixed order. During each problem-solving task, participants' eye movements, performance, and response times were recorded.

2.1.4. Data analysis

2.1.4.1. Prior knowledge. Participants could score one point for each correctly answered pretest question (i.e., max. score = 5 points).

2.1.4.2. Eye tracking measures. To investigate whether EMME was beneficial for guiding participants' attention towards the right information in the geometry figure at the right time, we determined the onset time of the verbal referents in the narration of the verbal explanations (e.g., in the sentence “Now that you know *this* angle, you can calculate the *other* angle.” the verbal referents are the words ‘this’ and ‘other’; each video modeling example contained 15 referents). For each verbal referent an area of interest (AoI) was constructed around the corresponding angle in the geometry figure (e.g., angle D or E). We then determined the proportion of fixations on the AoI corresponding with the verbal referents by dividing the number of fixated AoIs of the verbal referents by the total number of referents (*proportion fixated*). In addition, we determined how much time participants needed to first fixate the AoI of the verbal referent (*time lag*) after the onset of the verbal referent, and how long participants fixated the AoI of the verbal referent (*fixation duration*). Fixations (i.e., peak velocity $\leq 40^\circ/\text{s}$ and fixation duration ≥ 100 ms; cf. Jarodzka et al., 2013; Litchfield, Ball,

Donovan, Manning, & Crawford, 2010) were only included in the analyses if the fixation was within the time window of 1500 ms after the onset time of the verbal referent (cf. Dahan & Tanenhaus, 2005). Based on research findings suggesting that programming and initiating eye movements triggered by language input cost approximately 100 ms or longer (Altmann, 2011), fixations occurring within the first 100 ms were excluded from all eye tracking analyses.

Three participants (ME condition) were excluded from all eye movement data analyses because of poor calibration measures (i.e., deviation > 1 deg for x-axis or y-axis; x-axis deviation: $M = 0.46^\circ$; $SD = 0.17^\circ$; y-axis deviation: $M = 0.41^\circ$; $SD = 0.15^\circ$). One other participant (EMME condition) was identified as an outlier due to having very long time lags (z-score > 2.5). In addition, three participants (EMME condition $n = 2$, ME condition $n = 1$) were identified as outliers due to having very long fixation durations (z-score > 2.5).

2.1.4.3. Performance. One point was given for a geometry problem if all four steps had been performed correctly. Performance accuracy was computed by summing the number of points obtained and dividing that sum by the number of tasks, to obtain the proportion correctly solved problems; this was done separately for isomorphic and transfer problems. Two participants were identified as outliers for the analysis of isomorphic test problem accuracy and two participants were identified as outliers for the analysis of transfer test problem accuracy because of poor performance scores (z -score < -2.5). Performance speed was computed by averaging the response times on the correctly solved problems, separately for isomorphic and transfer problems. One participant was identified as an outlier for the response times isomorphic analysis and one participant was identified as an outlier for the response times transfer analysis because of very high response times (z -score > 2.5).

2.1.4.4. Statistical analyses. After the exclusion of all participants who were outliers on one of the performance and eye-tracking measures, the final sample consisted of forty-four participants ($n = 22$ EMME condition; $n = 22$ ME condition) and this sample was used for all the analyses reported. The data were analyzed with non-parametric Mann-Whitney U tests due to violation of the normality assumption with the exception of the time to first fixation analysis, which was analyzed with

Table 1

Mean (and SD) and median (and range) of performance and eye tracking measures of the eye movement modeling example (EMME) and modeling example (ME) conditions in experiment 1.

	EMME <i>M</i> (SD) <i>Mdn</i> (range)	ME <i>M</i> (SD) <i>Mdn</i> (range)
Eye Tracking Measures		
Proportion of Fixations (<i>n</i> = 44)	0.46 (0.13) 0.53 (0.47)	0.32 (0.12) 0.30 (0.47)
Time Lag (in ms; <i>n</i> = 44)	563.44 (111.21) 563.55 (440.74)	667.90 (102.18) 679.49 (391.57)
Fixation Duration (in ms; <i>n</i> = 44)	499.03 (187.51) 465.94 (699.83)	418.95 (159.91) 361.13 (678.29)
Proportion Correct		
Isomorphic (<i>n</i> = 44)	0.93 (0.18) 1.00 (0.50)	0.91 (0.20) 1.00 (0.50)
Transfer (<i>n</i> = 44)	0.86 (0.18) 1.00 (0.50)	0.88 (0.17) 1.00 (0.50)
Response Times (s)		
Isomorphic (<i>n</i> = 44)	71.43 (17.32) 67.16 (65.13)	61.54 (12.96) 61.00 (62.34)
Transfer (<i>n</i> = 44)	110.43 (36.93) 101.42 (127.86)	96.19 (27.32) 89.14 (113.08)

a *t*-test. We used *r* as a measure of effect size, with $r = .10$, $r = .30$, $r = .50$, representing small, medium, and large effects respectively (Cohen, 1988). We additionally conducted Bayesian analyses¹ with JASP (version 0.8.6.0; jasp-stats.org; JASP Team, 2018). One of the advantages of Bayesian analyses is that it expresses how much more likely the alternative or null hypothesis is given the obtained data instead of just rejecting the null hypothesis (see for an overview article: Wagenmakers et al., 2018). How much more likely the obtained data are under one of the hypotheses compared to the other hypothesis is expressed as a Bayes factor (BF). For instance, a $BF_{10} = 8.00$ indicates that the obtained data are eight times more likely under the alternative hypothesis than the null hypothesis. For each main analysis, we added the BF.

2.2. Results

We first checked for differences in prior knowledge. Prior knowledge was quite high in both conditions, and there was a small but significant difference in prior knowledge between the EMME condition ($M = 4.00$ out of 5; $SD = 1.07$) and the ME condition ($M = 4.59$ out of 5; $SD = 0.80$), $U = 159.00$, $z = -2.20$, $p = .028$, $r = .33$.

2.2.1. Eye tracking measures

See Table 1 for the means and standard deviations for the eye-tracking measures used in the following analyses. To address our first hypothesis, that ambiguous EMME would guide learners' attention to the relevant information in the video modeling examples (H1), a Mann-Whitney *U* test was conducted on the proportion of fixated verbal referents and fixation duration on referents, and a *t*-test on the time to first fixation on referents. The analysis on the proportion of fixated verbal referents (H1a) revealed that, in line with our hypothesis, participants in the EMME condition fixated the referents more often than participants in the ME condition, $U = 97.50$, $z = -3.42$, $p = .001$, $r = .52$, $BF_{10} = 78.633$. The Bayes factor indicates that the observed data are 78.6 times more likely under the alternative hypothesis (i.e., the conditions differ in the proportions of fixations) than under the null hypothesis (i.e., the conditions do not differ in proportions of fixations). In terms of the time (in ms) required to first fixate the verbal referents after onset (H1b), participants in the EMME condition were –as expected– significantly faster than participants in the ME condition, $t(42) = 3.24$, $p = .002$, $r = .49$, $BF_{10} = 15.694$. The Bayes factor indicates that the observed data are 15.6 times more likely under the alternative hypothesis (i.e., the conditions differ in the time to first fixation) than under the null hypothesis. Contrary to our hypothesis (H1c), however, there was no significant difference in fixation duration (in ms) between participants in the EMME condition and participants in the ME condition, $U = 170.00$, $z = -1.69$, $p = .091$, $r = .25$, $BF_{10} = 0.751$. The Bayes factor indicates that the observed data are less

than one time more likely under the alternative hypothesis (i.e., the conditions differ in the fixation duration) than under the null hypothesis.

2.2.2. Performance

See Table 1 for the means and standard deviations of the performance measures. To address our second hypothesis that seeing an EMME would result in enhanced performance accuracy (H2a) and speed (H2b) compared to the ME condition, non-parametric Mann-Whitney *U* tests were conducted. These revealed that neither accuracy on the isomorphic problems, $U = 231.00$, $z = -.41$, $p = .684$, $r = .06$, $BF_{01} = 3.147$, nor accuracy on the transfer problems, $U = 237.50$, $z = -.12$, $p = .904$, $r = .02$, $BF_{01} = 3.300$, differed significantly between conditions. The Bayes factors for the proportion correct for both the isomorphic and transfer problems indicates that the observed data are more than three times more likely under the null hypothesis (i.e., no difference between conditions) than under the alternative hypothesis. Regarding the response times for the isomorphic problems the results indicate that participants were faster in the ME condition than in the EMME condition, $U = 153.00$, $z = -2.09$, $p = .037$, $r = .32$, $BF_{10} = 1.806$. However, for the transfer problems there were no differences in response times between the conditions, $U = 191.00$, $z = -1.20$, $p = .231$, $r = .18$, $BF_{01} = 1.445$. The Bayes factor for the response times for the isomorphic problems indicates that the observed data are 1.8 times more likely under the alternative hypothesis (i.e., the conditions differ in the response times) than under the null hypothesis. However, for the transfer problems the Bayes factor indicates that the observed data are more than 1.4 times more likely under the null hypothesis (i.e., no difference between conditions) than under the alternative hypothesis.

2.3. Discussion

In line with our first hypothesis, results of the eye-tracking analyses revealed that seeing the model's eye movements in verbally ambiguous modeling examples (EMME) guided learners' attention more often (H1a) and faster (H1b) towards the verbally referred information than verbally ambiguous modeling examples that did not display the model's eye movements (ME). However, and in contrast to our second hypothesis, this did not result in better (H2a) performance on the geometry test problems in the EMME vs. the ME condition. Regarding response times no support was found for our hypothesis (H2b) as we found that learners of the ME condition were faster in solving the isomorphic problems, however for the transfer problems the solving speed did no longer differ between conditions. Although we had expected EMME to be more effective under conditions of verbal ambiguity, the results are largely in line with Van Marlen, Van Wermeskerken, Jarodzka et al. (2016) who used examples with unambiguous explanations. This seems to suggest that the ambiguity of the model's verbal explanation (in terms of being unspecific) might not play a role in the effectiveness of EMME.

¹ We would like to thank an anonymous reviewer for this suggestion.

It would be premature to draw that conclusion, however, as a possible explanation for the current findings might lie in the high amount of prior knowledge (i.e., average score of 4.1–4.6 out of 5 in both conditions) and the (resulting) overall high performance on the test problems in both conditions. Under conditions of high prior knowledge, the cognitive load imposed by the learning task is lower (Sweller, Ayres, & Kalyuga, 2011), and learners in the condition without attention guidance will probably be able to accommodate the additional load imposed by having to search for the verbal referents, without losing track of the model's demonstration and explanation. This explanation has also been offered for mixed findings in research on (other forms of) visual cueing (Van Gog, 2014) and is in line with findings from a recent meta-analysis showing that the effectiveness of attention guidance by means of visual cues is moderated by the amount of prior knowledge a learner has (Richter, Scheiter, & Eitel, 2016): learners with low levels of prior knowledge generally benefitted more from visual cues in multimedia learning materials than learners with higher levels of prior knowledge. However, note that even advanced learners or experts can benefit from EMME provided that there is still room for improvement in their performance. This was shown in a recent study by Gegenfurtner, Lehtinen, Jarodzka, and Säljö (2017) in which experts and novices performed a diagnostic task on CT and PET medical images before and after watching an EMME of an expert verbally explaining how to perform the task. It was found that even though experts had high prior knowledge, they benefited from seeing the EMME, which suggests that the effectiveness is not necessarily limited to lower levels of prior knowledge but might also depend on the extent to which there is room for improvement.

Therefore, Experiment 2 investigated the effects of ambiguous EMME and ME in a lower prior knowledge sample of secondary education students who have not yet been taught the F and Z principles in geometry.

3. Experiment 2

Next to the ambiguous conditions of Experiment 1, we also included unambiguous EMME and ME in Experiment 2, because these had not yet been tested in a lower prior knowledge sample. Thus, Experiment 2 had a 2 (EMME vs. ME) \times 2 (ambiguous vs. unambiguous verbal explanations) between-subjects design and was conducted with secondary education students. Because this experiment was conducted in a regular classroom setting, it did not involve eye tracking but only assessed test performance. Based on previous research showing beneficial effects of studying EMME (Jarodzka et al., 2012, 2013) and assuming the results of Experiment 1 were indeed due to the high prior knowledge, it was hypothesized that studying EMME would result in better test performance than studying ME (H1), that examples with unambiguous verbal explanations would lead to better performance than examples with ambiguous explanations (H2), and that the beneficial effects of EMME on performance would be larger in the ambiguous verbal explanation condition (interaction effect, H3).

3.1. Methods

3.1.1. Participants and design

Initially 142 first-year students from five classes of a Dutch school for secondary education participated in Experiment 2. However, due to software difficulties the experiment did not run properly in one class ($n = 30$). In addition, four participants did not provide informed consent to use their data. This resulted in a final sample of 108 students ($M_{age} = 12.05$, $SD = 0.46$, 53 male). Participants were randomly assigned to one of the four conditions resulting from a 2 (modeling example: EMME vs. ME) \times 2 (verbal explanation: ambiguous vs. unambiguous) between-subjects design: ambiguous EMME ($n = 28$), unambiguous EMME ($n = 26$), ambiguous ME ($n = 26$), and unambiguous ME ($n = 28$).

3.1.2. Materials

The experiment was created and conducted with the online questionnaire software Qualtrics (www.qualtrics.com). The pretest and the geometry test problems (isomorphic and transfer) were the same as those used in Experiment 1. The modeling examples (EMME and ME) with the ambiguous verbal explanations were exactly the same as in Experiment 1; the modeling examples with unambiguous verbal explanations were created by replacing the audio file with ambiguous verbal referents (e.g., “this angle”) by an audio file (recorded by the same model) with unambiguous verbal referents (e.g., “angle A”).

3.1.3. Procedure

The experiment, which was conducted during math class lasting approximately 50 min, was run in a computer classroom at participants' school. Participants were assigned to one of the conditions at random in advance of the session. There was a sheet of paper next to every computer with the name of the participant and the corresponding participant number to ensure that participants would work with the correct version of the computer program. Participants first received general and practical instructions about the experiment and were given the opportunity to ask clarification questions. Participants were then instructed to start the computer program and work through the program at their own pace. The program first asked students for their consent to use their performance data for research purposes. Then the pretest followed. After the pretest, all participants were presented with brief definitions and example images of the different types of angles used in the experiment.² Then participants were instructed that they were about to see video examples that would teach them how to solve geometry problems. In the EMME conditions, participants were also instructed that they would see the model's eye movements, visualized as a moving blue translucent dot (and they were shown a screenshot illustrating this; cf. Experiment 1). Then, participants in all conditions were instructed that after each video example they first had to indicate on a 5-point rating scale ranging from 1 (*very unclear*) to 5 (*very clear*) how clear they thought the verbal instructions were in the video example. Participants were additionally instructed that after rating the video example that they would be presented with a similar problem after each video example and that they should solve this problem as rapidly and as accurately as possible. After these instructions, participants were presented with the two example-problem pairs in the format of their assigned condition. Subsequently, participants solved the four transfer problems. The order of the video examples with the matching isomorphic problems were counterbalanced across participants and the transfer problems were presented in random order. During each problem-solving task participants' performance was recorded.

3.1.4. Data analysis

3.1.4.1. Prior knowledge. See Experiment 1.

3.1.4.2. Video ratings. The ratings of how clear learners perceived the verbal instructions of the video examples to be, were averaged across the two videos for each participant.

3.1.4.3. Performance. See Experiment 1 for details regarding the calculation of performance accuracy. Twenty-one participants were unable to finish all the transfer problems before the end of the math lesson and therefore the data of these participants were excluded, leaving 87 participants for the transfer performance analysis.

² This was done to prevent a bottom effect, as participants had not yet been taught the relevant geometry principles. Participants in all conditions received the same information.

Table 2

Mean (and SD) and median (and range) of the prior knowledge test, video ratings, and the proportion correct of the solved geometry problems for the eye movement modeling example (EMME) conditions and the modeling example (ME) conditions in experiment 2.

	EMME		ME	
	Ambiguous Instructions <i>M</i> (SD) <i>Mdn</i> (range)	Unambiguous Instructions <i>M</i> (SD) <i>Mdn</i> (range)	Ambiguous Instructions <i>M</i> (SD) <i>Mdn</i> (range)	Unambiguous Instructions <i>M</i> (SD) <i>Mdn</i> (range)
Prior Knowledge (1–5; <i>n</i> = 108)	1.93 (1.30) 2.00 (5.00)	1.88 (1.24) 2.00 (5.00)	1.77 (1.21) 2.00 (4.00)	2.07 (1.18) 2.00 (5.00)
Video Ratings (1–5; <i>n</i> = 108)	3.79 (0.75) 4.00 (3.00)	3.65 (0.80) 3.50 (3.00)	2.92 (0.90) 2.50 (3.50)	3.59 (0.72) 3.50 (3.50)
Proportion Correct				
Isomorphic (<i>n</i> = 108)	0.59 (0.41) 0.50 (1.00)	0.54 (0.40) 0.50 (1.00)	0.19 (0.32) 0.00 (1.00)	0.38 (0.35) 0.50 (1.00)
Transfer (<i>n</i> = 87)	0.28 (0.36) 0.00 (1.00)	0.30 (0.31) 0.25 (0.75)	0.11 (0.27) 0.00 (1.00)	0.21 (0.28) 0.00 (0.75)

3.2. Results

The prior knowledge, video ratings and performance data are presented in Table 2. The data were analyzed with 2 (modeling example: EMME vs. ME) \times 2 (instruction: ambiguous vs. unambiguous verbal referents) ANOVAs, and partial eta squared is reported as a measure of effect size, with $\eta_p^2 = .01$, $\eta_p^2 = .06$, $\eta_p^2 = .14$, representing small, medium, and large effects respectively (Cohen, 1988). In addition, we performed the Bayesian ANOVAs and reported the inclusion Bayes Factor (BF_{inc}). The BF_{inc} estimates the likelihood of the model if it contains the effect. For instance, when reporting a main effect of modeling example a $BF_{inc} = 7.500$, would indicate that the model including the effect of modeling example is 7.5 times more likely than the null model without an effect of modeling example. A check on the pretest scores revealed no main effect of the type of modeling example, $F(1, 104) < 1.00$, $p = .954$, no main effect of instruction, $F(1, 104) < 1.00$, $p = .589$, and no interaction, $F(1, 104) < 1.00$, $p = .469$.

3.2.1. Video ratings

To test whether participants perceived the verbal referents as less clear in the ambiguous conditions and to test whether the ratings of the verbal referents of the modeling examples are also affected by the presence of the model's eye movements a 2×2 ANOVA³ with the video ratings as dependent variable was conducted. This analysis revealed a main effect modeling example, $F(1, 104) = 9.21$, $p = .003$, $\eta_p^2 = .08$, $BF_{inc} = 15.950$, no significant effect of instruction, $F(1, 104) = 3.06$, $p = .083$, $\eta_p^2 = .03$, $BF_{inc} = 2.453$, and a significant interaction, $F(1, 104) = 6.82$, $p = .010$, $\eta_p^2 = .06$, $BF_{inc} = 6.927$. The results of the inclusion Bayes Factor suggest that the model including an effect of modeling example is almost 16 times more likely than the null model. Simple effects analysis revealed that the interaction was caused by lower video ratings for the ME ambiguous condition compared with the EMME ambiguous condition, $F(1, 104) = 15.94$, $p < .001$, $r = .38$, $BF_{10} = 75.981$. The Bayes factor indicates that the observed data are almost 76 times more likely under the alternative hypothesis (i.e., the ambiguous conditions differ) than under the null hypothesis. The video ratings did not differ between the ME unambiguous condition and the EMME unambiguous condition, $F(1, 104) = 0.09$, $p = .766$, $r = .03$, $BF_{10} = 0.286$. The Bayes factor indicates that the observed data are only 0.286 times more likely under the alternative hypothesis (i.e., the unambiguous conditions differ) than under the null hypothesis.

³ Because the ME unambiguous instruction condition did not meet the normality assumption, we additionally conducted non-parametric Kruskal-Wallis tests to examine whether the main effects from the 2×2 ANOVA would hold. This revealed the same effect for modeling example, $H(1) = 7.18$, $p = .007$, and instruction, $H(1) = 1.67$, $p = .197$.

3.2.2. Performance

To examine the effectiveness of EMME and also to examine whether this was affected by the ambiguity of verbal referents, a 2×2 ANOVA⁴ was conducted. For the isomorphic problems this analysis revealed a main effect of modeling example, $F(1, 104) = 15.36$, $p < .001$, $\eta_p^2 = .13$, $BF_{inc} = 98.329$, no main effect of instruction, $F(1, 104) < 1.00$, $p = .359$, $\eta_p^2 = .01$, $BF_{inc} = 0.351$ and no interaction, $F(1, 104) = 2.67$, $p = .106$, $\eta_p^2 = .03$, $BF_{inc} = 0.696$. This analysis indicates that learners in the EMME conditions performed better on the isomorphic problems compared with learners in the ME conditions, however the ambiguity of the verbal referents did not affect the performance nor the effectiveness of EMME relative to ME. The Bayes inclusion factor suggest that the model including an effect of modeling example is almost 98 times more likely than the null model. Similar results were found for the transfer problems as this analysis revealed a main effect of modeling example, $F(1, 83) = 4.13$, $p = .045$, $\eta_p^2 = .05$, $BF_{inc} = 0.916$, no main effect of instruction, $F(1, 83) < 1.00$, $p = .364$, $\eta_p^2 = .01$, $BF_{inc} = 0.247$ and no interaction, $F(1, 83) < 1.00$, $p = .543$, $\eta_p^2 < .01$, $BF_{inc} = 0.185$ (see Table 2). This indicates that performance was only affected by the type of modeling examples with learners in the EMME conditions outperforming the learners in the ME conditions. The Bayes inclusion factor suggest that the model including an effect of modeling example is roughly one time more likely than the null model.

3.3. Discussion

The aim of Experiment 2 was to examine the effects of EMME and explanation ambiguity on learning to solve a procedural task in secondary education students. In line with our hypothesis (H1), EMME yielded higher learning outcomes than ME. However, and contrary to our hypothesis (H2), ambiguity of the verbal explanation did not affect learning outcomes, either directly or in interaction with the type of modeling example. We will discuss these results in more detail in the next section.

⁴ Due to the non-normal distribution of the performance data, we additionally conducted non-parametric Kruskal-Wallis tests to examine whether the main effects from the 2×2 ANOVA would hold. For the isomorphic problems, we found that EMME outperformed the ME condition, $H(1) = 12.89$, $p < .001$, and there was no difference between the ambiguous instruction condition and the unambiguous condition, $H(1) = 0.68$, $p = .410$. Similar results were found for the transfer problems, EMME outperformed the ME condition, $H(1) = 4.15$, $p = .042$, and there was no difference between ambiguous instruction condition and the unambiguous condition, $H(1) = 1.38$, $p = .241$. In sum, the results of the non-parametric tests are in line with the results regarding the main effects of the 2×2 ANOVA.

4. General discussion

In two experiments, we investigated the role of the ambiguity of the model's verbal explanation and students' prior knowledge in the effectiveness of EMME for learning a procedural geometry task. Experiment 1 examined the effects of EMME on attention allocation and learning when the verbal explanation was ambiguous (i.e., unspecific). This experiment revealed that EMME helps to guide learners' visual attention towards the verbal referents faster (H1b) and more often (H1a) compared to a modeling example that did not show the model's eye movements. However, this attentional guidance by EMME did not result in higher accuracy (H2a) or consistently faster (H2b) geometry problem solving compared to ME, possibly due to the high prior knowledge (i.e., ceiling effect). Consequently, Experiment 2 investigated the effects of EMME and explanation ambiguity on learning to solve a procedural geometry task in secondary education students who had limited prior knowledge.

Experiment 2 revealed that secondary education students in the EMME conditions outperformed their counterparts in the ME conditions (H1) on both isomorphic and transfer problems. This is an interesting and important finding, as it extends the tasks to which EMME are applicable. There was some evidence that EMME can be effective for learning classification tasks (Jarodzka et al., 2012, 2013; Sridharan, Bailey, McNamara, & Grimm, 2012; Vitak et al., 2012) and strategy learning (Mason et al., 2015b, 2015a), but to our knowledge, this is the first study to demonstrate that EMME can be effective to enhance learning of a procedural problem-solving task. Earlier studies found no beneficial effect of EMME for learning a procedural geometry problem-solving task in university students (Van Marlen, Van Wermeskerken, Jarodzka et al., 2016) and a study using a procedural puzzle problem even found a negative effect on learning when the EMME was accompanied by a verbal explanation (Van Gog et al., 2009). The present study demonstrates that EMME can be effective for learning a procedural problem-solving task, but suggests that this is only the case if prior knowledge is low. Findings by Van Marlen, Van Wermeskerken, Jarodzka et al. (2016) and from Experiment 1 show that for university students, who had relatively high prior knowledge, EMME successfully guided their attention but this did not affect their learning outcomes. Presumably, students who have more working memory capacity at their disposal, for instance because they have more prior knowledge, are able to accommodate the additional working memory load imposed by having to search for the verbal referents, without losing track of the model's demonstration and explanation. So even if students who do not receive EMME may take a little more time to locate the referent, this does not hamper their learning. For the secondary education students in Experiment 2, who had lower prior knowledge than the university students from Experiment 1, the guidance provided by EMME did foster learning.

Interestingly, the ambiguity of the verbal explanation did not significantly affect learning outcomes in Experiment 2 (H2). Learners did perceive the verbal explanations in the video examples to be clearer when they were accompanied by the model's eye movements, especially when the verbal explanation was ambiguous. This interaction effect on perceived clarity seems to be mirrored by the test performance data

shown in Table 2, with the difference between the mean scores of the EMME and ME conditions being larger when the explanation was ambiguous than when it was unambiguous. However, there was no significant interaction effect on performance. There are several possible explanations for this lack of effect of verbal ambiguity. First, in Experiment 2, it is possible that given the substantial variation in test performance (which is not uncommon in the classroom), the sample size was too low to be able to detect a significant interaction. Another potential explanation might be that the ambiguity manipulation was too subtle. Our ambiguity manipulation consisted of using unspecific rather than specific referents (e.g., “that angle” instead of “angle A”). Although it may not be immediately clear for learners where to look based on such unspecific referents, they may be less detrimental for learning than the use of referents that learners will not know the meaning of when they cannot see what the model is looking at (e.g., the use of jargon like “dorsolateral fin”; Jarodzka et al., 2013). Moreover, the fact that the model could be observed performing the problem-solving steps by clicking and typing, would have synchronized learners' attention intermittently throughout the example. So even if they lost track due to the unspecific referents in the no EMME condition, this would be resolved once the model performed a step.

Therefore, it would be interesting in future research to replicate Experiment 2 with a larger sample size and to extend it to different problem-solving tasks (with/without physical actions on the computer screen) and other sources of ambiguity of the model's verbal explanation. In addition, it would also be interesting to directly manipulate prior knowledge within a single experiment. A limitation of the present study is that we only manipulated prior knowledge between experiments, and therefore cannot directly verify that this is indeed the factor that moderates the effectiveness of EMME. In addition, future research might explore whether other factors like working memory capacity, spatial ability, and reasoning ability also moderate the effectiveness of EMME. In general, further research on EMME (and other forms of cueing) in which prior knowledge is manipulated is important considering that the literature shows on the one hand that people with low prior knowledge benefit most of the visual cues (Richter et al., 2016), while on the other hand, a recent study suggests that for certain tasks, experts may profit more from EMME than novices (Gegenfurtner et al., 2017).

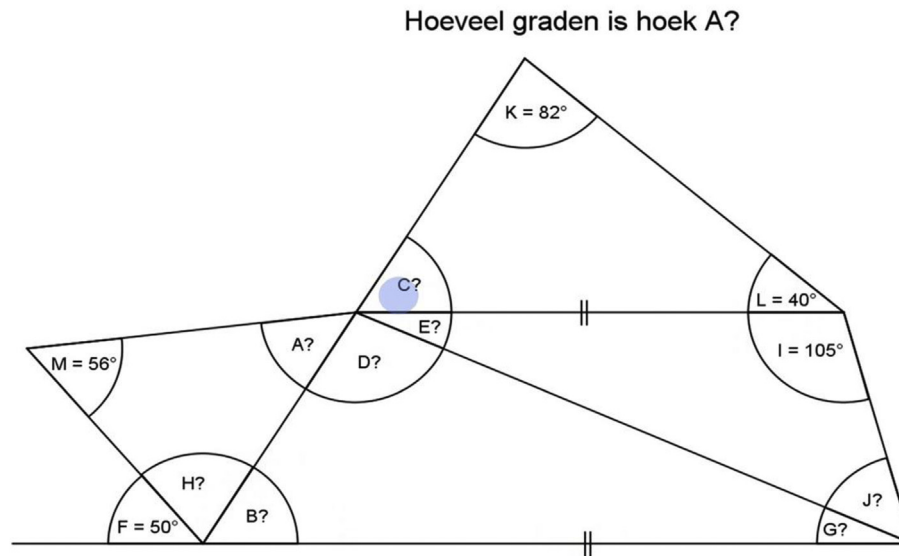
To conclude, the present study shows that –in contrast to prior findings– EMME can be effective in enhancing learning of a procedural problem-solving task, at least when prior knowledge is low. This is also relevant for educational practice, as our study demonstrates that EMME –once created– can easily be implemented in secondary education classrooms.

Acknowledgements

This research was funded by a Vidi grant (# 452-11-006) from the Netherlands Organization for Scientific Research (NWO). The authors would like to thank Susan Ravensbergen for assistance with data collection. In addition, we would like to thank all the staff members and students of Emmauscollege for participating.

Appendix

Example of a transcript of the verbal explanation in a modeling example in Experiment 1 and Experiment 2.



Screenshot of a modeling example (EMME condition) in Experiment 1 and Experiment 2 with the blue dot representing the location of the model's gaze. The following verbal instruction was used during the modeling example (translated from Dutch): "The question is, how many degrees is angle A? You start by searching for this angle. Angle A is part of a triangle. A triangle contains a total of 180 degrees. If two of the three angles are known in a triangle, you can calculate the third angle. You calculate the third angle by subtracting the two angles from 180 degrees. You cannot calculate it right now, because the second angle is unknown. The second angle is part of a straight line. A straight line contains a total of 180 degrees. You can calculate the unknown angle in a straight line by subtracting all known angle from 180 degrees. However, because the right angle is also unknown, it is not possible to calculate this angle no. You cannot calculate the right angle directly, but it can be derived from this angle, because these are equal. This can be seen by this sign, which indicates that the lines are parallel and thus have the same angle. Because of the parallel lines, you can derive by means of the corresponding angle principle that these angles are equal. This angle is unknown for now but can be calculated. This angle equals 180 degrees minus the known angles, equals 58 degrees. Now this angle is known, you know that the other angle, by means of the corresponding angle principle, also equals 58 degrees. With this angle known, you can now calculate angle A. Angle A equals 180 degrees minus the known angles, equals 52 degrees. So angle A is 52 degrees."

References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439. <https://doi.org/10.1006/jmla.1997.2558>.
- Altman, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, 137, 190–200. <https://doi.org/10.1016/j.actpsy.2010.09.009>.
- Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology*, 21, 811–820. <https://doi.org/10.1002/acp.1351>.
- Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, 112, 330–336. <https://doi.org/10.1016/j.cognition.2009.04.008>.
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106, 1465–1477. <https://doi.org/10.1016/j.cognition.2007.05.012>.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20, 155–166. <https://doi.org/10.1016/j.learninstruc.2009.02.014>.
- Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, 29, 1146–1152. <https://doi.org/10.3758/BF03206384>.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3, 149–210.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12, 453–459. <https://doi.org/10.3758/BF03193787>.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436. <https://doi.org/10.1007/BF02143160>.
- Frishen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133, 694–724. <https://doi.org/10.1037/0033-2909.133.4.694>.
- Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, 113, 212–225. <https://doi.org/10.1016/j.compedu.2017.06.001>.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 172–190.
- Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 37–53. <https://doi.org/10.1037/a0017683>.
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., et al. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40, 813–827. <http://doi.org/10.1007/s11251-012-9218-5>.
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70. <http://doi.org/10.1016/j.learninstruc.2012.11.004>.
- JASP Team (2018). JASP (version 0.8.6.0) [computer software]. Retrieved from <https://jasp-stats.org>.
- Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 247–262). (2nd ed.). New York: Cambridge University Press.
- Litchfield, D., Ball, L., Donovan, T., Manning, D., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, 16, 251–262. <https://doi.org/10.1037/a0020082>.
- Louwerse, M. M., & Bangerter, A. (2010). Effects of ambiguous gestures and language on the time course of reference resolution. *Cognitive Science*, 34, 1517–1529. <https://doi.org/10.1111/j.1551-6709.2010.01135.x>.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2015a). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, 41, 172–187. <http://doi.org/10.1016/j.cedpsych.2015.01.004>.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2015b). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.12271> Advance online publication.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.). *The*

- Cambridge handbook of multimedia learning (pp. 43–71). (2nd ed.). New York: Cambridge University Press.
- Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford, UK: Oxford University Press.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. <http://doi.org/10.1111/cogs.12086>.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is co-ordination. *Psychological Science*, 18, 407–413. <https://doi.org/10.1111/j.1467-9280.2007.01914.x>.
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36. <https://doi.org/10.1016/j.edurev.2015.12.003>.
- Sridharan, S., Bailey, R., McNamara, A., & Grimm, C. (2012). *Proceedings of the 2012 Symposium on eye tracking research & Applications ETRA '12*, 75–82. Santa Barbara, CA: ACM. <https://doi.org/10.1145/1743666.1743737>.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 263–278). (2nd ed.). New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369>.
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25, 785–791. <http://doi.org/10.1016/j.chb.2009.02.007>.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, 19, 205–221. <https://doi.org/10.1002/acp.1112>.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22, 155–174. <https://doi.org/10.1007/s10648-010-9134-7>.
- Van Marlen, T., Van Wermeskerken, M., Boven, L., Ketelaar, J., van Trijp, K., Vernooij, S., & Van Gog, T. (2016a). Are eye movement modeling examples most useful when verbal instructions are ambiguous? *Paper presented at the EARLI Special Interest Group 6 and 7* Dijon, France.
- Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016b). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computer in Human Behavior*, 65, 448–459. <https://doi.org/10.1016/j.chb.2016.08.041>.
- Van Marlen, T., Van Wermeskerken, & Van Gog, T. (2018). *Effects of visual complexity and ambiguity of verbal instructions on target identification*. Manuscript submitted for publication.
- Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., De Bock, J. J. R. P., & Van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32, 10–21. <https://doi.org/10.1016/j.learninstruc.2014.01.004>.
- Vitak, S. A., Ingram, J. E., Duchowski, A. T., Ellis, S., & Gramopadhye, A. K. (2012). Gaze-augmented think-aloud as an aid to learning. *Proceedings of ACM CHI 2012 conference on human factors in computing systems*, 2991–3000. Austin, Texas: ACM Press.
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25, 35–37. <https://doi.org/10.3758/s13423-017-1343-3>.
- Wolff, C. E., Jarodzka, H., Van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, 44, 243–265. <https://doi.org/10.1007/s11251-016-9367-z>.