Computers & Education 72 (2014) 323–327

Contents lists available at ScienceDirect

Computers & Education

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

Learning from video modeling examples: Effects of seeing the human model's face

Tamara van Gog*, Ilse Verveer, Lise Verveer

Institute of Psychology, Erasmus University Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 21 July 2013
Received in revised form
6 December 2013
Accepted 9 December 2013

Keywords:
Instructional design
Example-based learning
Observational learning
Attention allocation
Eye tracking

ABSTRACT

Video modeling examples in which a human(-like) model shows learners how to perform a task are increasingly used in education, as they have become very easy to create and distribute in e-learning environments. However, little is known about design guidelines to optimize learning from video modeling examples. Given that the human face is known to capture observers' attention, the question addressed in this study is whether seeing the model's face in the video would help or hinder learning. Participants twice studied a video modeling example in which a problem-solving task was demonstrated and in which the model's face was either visible or not, and after each view they attempted to solve the problem themselves. Their eye movement data were recorded while watching the video examples. Results show better performance after observing the second example in the condition that did see the model's face.

1. Introduction

Example-based learning is a highly effective instructional technique, especially for novice learners (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2013; Sweller, Van Merriënboer & Paas, 1998; Van Gog & Rummel, 2010). Two different types of examples can be distinguished: worked examples in which a solution procedure is shown in the form of written text, and modeling examples in which a solution procedure is demonstrated and/or explained to learners by a human(-like) model (Van Gog & Rummel, 2010). Modeling examples can be observed in 'live' face-to-face educational contexts (e.g., Bjerrum, Hilberg, Van Gog, Charles, & Eika, 2013), on (digital) video (e.g., Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013; Schunk, Hanson, & Cox, 1987), as a digital video recording of the model's screen (e.g., McLaren, Lim, & Koedinger, 2008; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009), or as an animation in which the model could be represented by an animated agent (see Wouters, Paas, & Van Merriënboer, 2008).

Nowadays, video modeling examples are being increasingly used as they have become easy to create and distribute in e-learning environments. However, we still know relatively little about how such video examples should be designed to optimally enhance students' learning. One design aspect in which a lot of variation among video modeling examples is evident, is whether and how they show the human model. Some videos show the model almost entirely (i.e., torso and head; e.g., Schunk et al., 1987), some only partly (e.g., zooming in on the hands; e.g., Groenendijk et al., 2013), and screen-recording videos do not show the model at all (i.e., only what the model is doing on the computer screen, possibly with a voice over; e.g., McLaren et al., 2008; Van Gog et al., 2009). The question addressed in this study is whether or not seeing the model's face while studying video modeling examples would form a distraction that might hinder learning.

It may appear strange to think of another human's face as a potential distractor, but eye-tracking research on social interactions has shown that when we look at someone who is providing a spoken explanation of an event, we tend to look primarily at that person's face (as much as 95% of the time; Gullberg & Holmqvist, 2006). It seems this finding cannot be attributed to a kind of 'politeness' or 'social norm' in social interaction, as even on small-screen videos, where the speaker cannot see the observer and therefore cannot be offended by the observer's viewing behavior, approximately 90% of observers' fixations were on the speaker's face, and this increased to 94% when the video was life size (Gullberg & Holmqvist, 2006).

In some cases, video can even lead to increased looking at others compared to live situations. For example, in a situation in which social interaction was not required, it was found that participants gazed much more often at a confederate of the researchers who was seen sitting

* Corresponding author. Institute of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Tel.: +31 1040889041.
E-mail address: vangog@fsw.eur.nl (T. van Gog).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2013.12.004

in a waiting room on a video monitor, than when the confederate was sitting in the same room with the participant (Laidlaw, Foulsham, Kuhn, & Kingstone, 2011). In an interactive collaboration task, not using eye tracking but a video-based gaze-coding procedure, Doherty-Sneddon et al. (1997) found that participants who could see each other eye to eye on video, looked at each other's face 56% more than participants who could see each other live, or participants who saw each other on video but without being able to make eye contact.[1]

Further evidence that faces are attention magnets and that this is not just the case in live social situations, comes from a study with human-like animated agents that are often used in animations or computer-based tutoring environments. Learners have been found to look at faces of such agents as they would at real humans (Louwerse, Graesser, McNamara, & Lu, 2009). That is, they look primarily at the agent's face, and in the presence of multiple agents, look at the speaking agent's face, which may draw attention away from those aspects of the learning environment that the agent is speaking about, which, in turn, might hamper learning. Even though effects on learning were not assessed by Louwerse et al. (2009), this capturing of attention by the agent's face seems a plausible explanation for the mixed findings regarding effectiveness of animated pedagogical agents that were described by Moreno (2005).

However, it has not yet been investigated whether learners who are engaging in social learning task, that is, learning by observing video modeling examples, would show similar viewing patterns. For instance, in the study by Gullberg and Holmqvist (2006), next to watching the speaker's face more than 90% of the time, it was found that participants paid hardly any attention to the speaker's gestures, despite the fact that gestures often provide a useful information source for listeners (e.g., Ping & Goldin-Meadow, 2008). However, in contrast to viewing a video of someone relating an event, observing video modeling examples is done with a specific goal, that is, learning to perform the task oneself. Moreover, the model will not just be making co-speech gestures, but is engaging in a demonstration, which implies movements made by the model that are likely to attract attention. Especially if the model's gaze is directed towards an object prior to manipulating it, it is likely that learners would reflexively follow the model's gaze (for studies on gaze-following behavior, see e.g., Driver et al., 1999; Gregory & Hodgson, 2012; Kuhn, Tatler, Findlay, & Cole, 2008; for a review, see e.g., Langton, Watt, & Bruce, 2000).

Nevertheless, if learners would indeed attend to the model's face for a substantial amount of time, this could imply that they are not attending sufficiently or timely to the task that the model is demonstrating. In that case, video modeling examples that show the model only partly, that is, only the demonstration (which can then receive full attention), might be more effective for learning. The present study addressed this question, by comparing the effects on learning of observing a video modeling example in which the model's face is visible or not visible, and recording participants' eye movements while viewing the videos.

2. Method

2.1. Participants and design

Participants were 43 adults (10 male; age $M = 22$, $SD = 8.80$); the majority (35) were psychology students. They participated for course credit or a monetary reward of 3.5 Euro (approximately 4.5 US Dollar at the time of writing). Participants were randomly assigned to one of two conditions: 1) the model's face is visible in the video modeling example ($n = 21$) or 2) the model's face is not visible in the example ($n = 22$).

2.2. Materials

2.2.1. Problem-solving task and video modeling examples

The learning task was a real-object version of a computer-based problem-solving task that circulates on the Internet, called Frog Leap. In the computer-based version of the problem-solving task, three green frogs are sitting on stones on one side of the river, three brown frogs on the other side, with one empty stone in the middle and the goal (reachable in 15 steps) is to have them switch sides (see Van Gog, 2011; Van Gog et al., 2009). For this study, a version of the task was created using real objects, consisting of three plastic yellow fishes and three plastic green seals.

The video modeling example showed a female model seated behind a table on which the objects were placed, who demonstrated the problem-solving procedure while verbally explaining it. The video had a duration of 142 s, and was recorded with a JVC Everio HD camcorder on a tripod. The video for the condition that did see the model's face, showed the model's torso and head (see Fig. 1). It had a size of 720×580 pixels and was shown on a 1600×1200 screen, against a black background. The video and was divided into three Areas of Interest (AoI): 'head' (720×250), 'middle' (720×85; no useful information), and 'demonstration' (720×245; where the objects were located and hand movements were made). The video for the condition that did not see the model's face was created by removing the AoI 'head' (see Fig. 2). As a consequence the size of the videos was not equal between the two conditions (720×580 in the face visible condition compared to 720×330 in the face not visible condition), but note that the size of the 'demonstration' (and 'middle') area was equal between conditions, which was important to be able to meaningfully compare the eye movements on the AoI 'demonstration' between both conditions.

Immediately after studying the modeling example, participants were provided with the same plastic objects and had to attempt to solve the problem themselves. This example-problem solving sequence was repeated twice. Then, a transfer task was administered, which consisted of the same problem, but with one object added on each side (i.e., 4 fishes and 4 seals), which no longer allows one to copy the observed procedure from the examples (15 steps) but can be solved in 24 steps if the problem-solving procedure is really understood.

2.2.2. Eye-tracking equipment

The eye tracker that was used was a Tobii 2150 (50 Hz sampling rate) with ClearView software. Participants were seated approximately 70 cm from the screen. A 9-point calibration procedure was applied prior to studying the example both times.

[1] Intriguingly, while live interaction resulted in more efficient verbal communication, seeing each other eye to eye on video led to less efficient verbal communication than not being able to see each other at all (i.e., audio only). It should be noted though, that this finding was in contrast with results of Boyle, Anderson, and Newlands (1994) and that this had no effect on collaborative task performance.