It's all a matter of expectancy: how expectancies about task execution perspective influence learning through video modeling examples

Jonge, T.H.S. de; Koning, B.B. de; Wermeskerken, M. van

Learning and Instruction 105 (2026) 102422

Contents lists available at ScienceDirect

Learning and Instruction

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

It's all a matter of expectancy: How expectancies about task execution perspective influence learning through video modeling examples

Tristan H.S. de Jonge [a,c,*], Bjorn B. de Koning [b], Margot van Wermeskerken [a]

[a] Department of Education, Utrecht University, the Netherlands
[b] Department of Education and Child Studies, Erasmus University Rotterdam, the Netherlands
[c] Department of Education and Child Studies, Leiden University, the Netherlands

ARTICLE INFO

Keywords:
Video modeling examples
Perspective effect
Multimedia learning
Video learning
First-person perspective
Third-person perspective

STRUCTURED ABSTRACT

Background: The perspective effect in video modeling suggests that learning is enhanced when instructional videos are recorded from a first-person angle (presenter's viewpoint) rather than a third-person angle (external viewpoint). Two explanations have been proposed: the Embodied Resonance Hypothesis, which attributes this advantage to greater mirror neuron activation during first-person observation, and the Expectation Alignment Hypothesis, which suggests that learning is impaired when there is a mismatch between the recording angle and expectations about the task angle.
Aims: This study aimed to investigate the mechanisms underlying the perspective effect by examining the recording angle, task angle, and expected task angle in observational learning.
Sample: A total of 128 undergraduate psychology students participated in the study.
Methods: Participants viewed two video modeling examples demonstrating circuit board assemblies, recorded either from a first-person or third-person angle. Their expectations regarding the angle from which they would perform the task were manipulated beforehand (first-person vs. third-person). Subsequently, participants replicated the assemblies from varying angles (first-person vs. third-person). Task performance and cognitive load were outcome variables.
Results: Results supported both hypotheses. First-person videos led to better learning outcomes overall, supporting the Embodied Resonance Hypothesis. In addition, performance was impaired and cognitive load increased when the expected task angle did not align with the recording angle, supporting the Expectation Alignment Hypothesis.
Conclusions: The perspective effect in video learning appears to result from both an inherent advantage of first-person and by expected misalignment with the task angle. These insights offer practical implications for optimizing instructional video design.

1. Introduction

Video as an instructional medium has seen a remarkable surge in educational settings (Lin & Yu, 2023). A type of instructional video that is particularly helpful for learning procedural knowledge are video modeling examples, i.e., videos in which a task is demonstrated by a presenter (Van Gog & Rummel, 2010). Over the past decade, research has focused on optimizing video modeling examples by investigating design choices such as presenter characteristics like gender, age, and expertise (Hoogerheide et al., 2016) and the visibility of the presenter's face or hands (De Koning et al., 2022; Kizilcec et al., 2015; Ouwehand et al., 2015; Van Gog et al., 2013; Van Wermeskerken et al., 2018). Another key attribute of video modeling examples is the angle from which they are recorded, with recent studies comparing a first-person angle to a third-person angle (Allé et al., 2023; De Koning et al., 2022; Fiorella et al., 2017; Leng et al., 2024).

A first-person recording angle captures the video modeling example from the viewpoint of the person performing the task, while the third-person angle captures the task from an external viewpoint (i.e., from the position of someone observing across from the presenter). Studies comparing the effects of these two recording angles on learning have favored video modeling examples recorded from a first-person angle (e.

* Corresponding author. Department of Education, Utrecht University, P.O. Box 80140, 3508 CS, Utrecht, the Netherlands.
E-mail address: hstdjonge@gmail.com (T.H.S. de Jonge).

https://doi.org/10.1016/j.learninstruc.2026.102422
Received 10 November 2025; Received in revised form 23 March 2026; Accepted 12 June 2026
Available online 17 June 2026

g., De Koning et al., 2022; Fiorella et al., 2017; Garland & Sanchez, 2013; Leng et al., 2024), as they found them leading to better performance compared to those recorded from a third-person angle. This advantage is often referred to as the *perspective effect* (Fiorella et al., 2017; Leng et al., 2024) – the idea that learning is more effective when procedural information is observed from the natural first-person perspective compared to the third-person perspective.[1]

However, the conclusions drawn from current literature are limited by the fact that performance was assessed on tasks that were all completed from a first-person angle. This raises a critical issue, as it remains unclear whether the perspective effect is caused by an inherent advantage of the first-person recording angle, or because a first-person recording angle often aligns with the task angle (as was the case in prior studies). Conversely, it is often impossible to complete tasks from a third-person angle, which results in an inherent misalignment between third-person recordings and the task angle. The most profound effect of this misalignment could already stem from mere expectations about the task angle: The video might be more difficult to process if the recording angle is different from the angle that the task is expected to be executed from.

This issue is particularly relevant in cases where learners perform tasks from different angles, such as when tying a bowtie – where the movements differ depending on whether one is tying their own or someone else's bowtie. In such cases, recording from a first-person angle might not inherently yield better learning outcomes. On a practical note, creating third-person recordings also requires less specialized equipment (as the camera can be simply placed in front of the model) and they can capture much contextual information (Debarba et al., 2017), such as a blackboard behind the presenter. Hence, a clearer understanding of the effects of recording angles is necessary to refine guidelines on how to optimize video modeling examples for learning.

Therefore, the present study investigates the effects of recording angle in video modeling examples, the angle from which participants expect to perform a replication task (expected angle), and the angle from which they actually perform this replication task (task angle) on learning and cognitive load. By examining the effects of these factors and the alignment between them, the present study aims to clarify the mechanisms that underlie the perspective effect: an inherent benefit of the first-person recording angle or a benefit that depends on expectations about the task angle.

1.1. The effects of recording angle

Although early work comparing first-person and third-person recording angles can be traced back to the beginning of the 21st century (Ruby & Decety, 2001; Vogeley & Fink, 2003; Vogt et al., 2003), experimental investigations comparing their effect in video modeling examples are more recent. Fiorella et al. (2017) compared first-person and third-person video modeling examples in a memory task that involved the assembly of an electrical circuit. They found that participants who watched the first-person video outperformed those who watched the third-person video in replicating the circuit configuration, but cognitive load while watching or performing the task did not differ between conditions.

Similarly, first-person and third-person recording angles were compared by Garland and Sanchez (2013) and De Koning et al. (2022), utilizing a knot-tying task to assess their effects on motor learning. Garland and Sanchez (2013) found that participants who watched first-person videos completed the replication task faster and made fewer errors compared to those who watched third-person videos. However, this perspective effect was absent for static representations (images). De Koning et al. (2022) also found a benefit of a first-person recording angle over a third-person recording angle, with higher accuracy in recognizing and replicating knots. Additionally, they examined the presence of the presenter's hands and discovered that the perspective effect was only present when hands were visible.

Although the above-mentioned studies provide behavioral evidence for learning differences between first-person and third-person video modeling examples, neuroscientific research on recording angles offers deeper insights into the cognitive processes that may underlie this effect. Using functional Magnetic Resonance Imaging (fMRI), Jackson et al. (2006) found greater activation in the sensorimotor cortex when participants viewed hand and foot movements recorded from a first-person recording angle as opposed to a third-person recording angle. More recently, Leng et al. (2024) combined an experimental design with functional Near-Infrared Spectroscopy (fNIRS) to investigate the effects of recording angle when participants watched origami video modeling examples. Results showed that, compared to third-person recordings, first-person recordings led to better and faster performance on the origami task. Their fNIRS data showed that, during video observation, the group who watched the first-person recordings, compared to the group who watched the third-person recordings, exhibited higher cortical activation in the right dorsolateral prefrontal cortex, which is involved in cognitive control and working memory. This suggests that, compared to third-person recordings, first-person recordings elicit more learning-relevant cognitive processes, such as memory storage and attention.

Overall, these studies point to an advantage of the first-person recording angle for procedural tasks. However, this does not necessarily imply that third-person recordings are inherently disadvantageous. Research also highlights potential benefits of a third-person recording angle. For instance, gaze provides important social cues and information about what a model is currently focusing on (Van Wermeskerken et al., 2018), yet such cues are typically absent in first-person recordings. Third-person recording angles are also associated with a more observer-oriented (i.e., allocentric) mode of processing (Allé et al., 2023; Ciorli & Pia, 2023), which may be advantageous in situations where an observer-based understanding of an action is required rather than direct motor engagement. Furthermore, third-person recording angles often capture more of the surrounding context and can support spatial awareness (Gorisse et al., 2017), which may be particularly important when environmental information is relevant to successful task performance. Optimizing video modeling examples therefore requires a more nuanced understanding of the perspective effect and a clearer account of why it occurs and under which conditions it holds. However, existing research does not converge on a single explanation, and the effect may be attributable to different underlying mechanisms.

1.2. Embodied Resonance Hypothesis

When observing others, such as in a video modeling example, individuals construct a mental representation that captures the procedural intricacies of the observed action (Jeannerod & Frak, 1999). Experimental and neuroimaging studies have demonstrated that the motor mechanisms involved in action observation substantially overlap with those engaged during action execution (e.g., Torriero et al., 2007). This occurs due to a widespread mirror neuron system that internally simulates observed actions (Rizzolatti & Craighero, 2004), a process known as *embodied simulation* (Wilson, 2002).

Prior research has put forward the mirror neuron system to explain the perspective effect (De Koning et al., 2022; Jackson et al., 2006; Leng et al., 2024; Vogeley & Fink, 2003). This explanation will be referred to as the *Embodied Resonance Hypothesis* (ERH), as it attributes the

[1] In the present study, *perspective* refers to the viewpoint from which an action is observed. *Angle* refers to the physical spatial instantiation of that perspective, namely the concrete position from which a video is recorded or a task is performed. For instance, although actions are inherently executed from a first-person perspective, a task can be physically oriented in a way that spatially corresponds to a third-person recording angle.

first-person recording benefits to the enhanced resonance between observed actions and the observer's own motor system.

The mirror neuron system operates by activating motor representations and linking newly observed behaviors to these representations (Buccino et al., 2004; Vogt et al., 2003). The more closely an observed action resembles an existing motor representation, the stronger that representation is activated, making it easier for the observed action to be integrated into memory (Buccino et al., 2004; Meltzoff & Decety, 2003; Rizzolatti, 2005). Actions observed from a first-person perspective align more closely with an individual's pre-existing motor schemas than actions observed from a third-person perspective (Jackson et al., 2006; Mayer et al., 2020) because one's own actions are exclusively performed in first-person, and motor memory thus predominantly stores first-person experiences. Due to this stronger neurological link to action execution – compared to third-person observations, which differ from one's own motor schemas – actions observed from a first-person perspective are simulated more strongly in the brain and are more effectively integrated into memory, leading to enhanced processing and recall of the actions. This is evidenced by greater activation of the mirror neuron system (Jackson et al., 2006; Watanabe et al., 2013) and brain regions related to observation and motor preparation (Leng et al., 2024) when observing actions recorded from a first-person angle compared to a third-person angle.

1.3. Expectation Alignment Hypothesis

While the ERH posits the perspective effect as a relatively fixed phenomenon – suggesting a consistent advantage for first-person recordings – the present study offers an alternative explanation that presents a more dynamic account: The Expectation Alignment Hypothesis (EAH) suggests that the benefit of a particular recording angle depends on how well it aligns with the viewer's expectations about the task angle. Here, expectation reflects anticipatory sensorimotor predictions about how the action will be executed at the time of performance, which may lead viewers to mentally transform a third-person recording in order to replicate it from their own first-person perspective. These spatial transformations occur as the brain processes and encodes information in anticipation of future use (Lohmann et al., 2017; Spratling, 2016), which typically involves replication from a first-person angle, as actions are naturally executed from one's own point of view.

There are various forms of mental spatial manipulation, but two are particularly relevant to the perspective effect: *mental rotation*, where an object's orientation is mentally manipulated; and *perspective taking*, where an individual mentally shifts their viewpoint to simulate an alternative spatial perspective (Hegarty & Waller, 2004). Such transformations induce a significant degree of cognitive load (i.e., cognitive effort; Bauer et al., 2021; Bochynska et al., 2020), which reduces the available working memory capacity. This means that less of the limited working memory is available for effective processing and encoding of the observed information, which may lead to poorer task performance.

Thus, according to the EAH, the learning benefits of first-person video modeling examples arise because mental transformations are only required when viewing third-person recordings, as individuals typically expect to perform tasks from a first-person angle. However, along the reasoning of the EAH, a similar phenomenon would occur when watching a first-person recording while expecting to replicate the task from a third-person angle, suggesting that there is no inherent advantage to a first-person recording angle.

1.3.1. Expectation and spatial processing

The idea that mental transformation plays a role in the perspective effect is not unique to the EAH. Numerous studies have linked the perspective effect to spatial processing, supported by neuroimaging evidence showing that observing third-person recordings increases activity in brain regions involved in spatial processing, such as the precuneus, posterior cingulate cortex, frontopolar cortex, and inferior parietal cortex (Leng et al., 2024; Ruby & Decety, 2001; Vogeley & Fink, 2003; Watanabe et al., 2011). While these studies often attribute the perspective effect at least partly to the spatial manipulations required for third-person recordings, they typically consider this a matter of transfer – where stored third-person information must be adjusted during task performance to align with a first-person task. The EAH proposed in the present study extends this reasoning by suggesting that these spatial transformations may already occur during video observation, driven solely by expectations about the angle from which the task is expected to be performed.

In this context, individual differences in spatial ability may also be relevant. Mental rotation ability reflects the efficiency with which individuals can manipulate spatial representations (Hegarty & Waller, 2004). If misalignment between recording angle and expected task angle induces spatial transformations during observation, individuals with lower spatial ability may experience greater cognitive demands when required to mentally align these perspectives. Consequently, assuming the EAH holds, potential performance impairments due to misalignment may be more pronounced for individuals with lower mental rotation ability than for individuals with higher mental rotation ability. However, prior research on video modeling examples has not consistently demonstrated moderating effects of incorporated spatial ability measures (e.g., perspective taking; Fiorella et al., 2017).

More generally, the role of expectation in learning has been explored in other contexts, such as how a test expectancy or teaching expectancy can enhance learning outcomes (Fiorella & Mayer, 2013; Liang et al., 2024). This means that certain task expectations trigger preparatory processes, often influencing learning regardless of whether the expectation is ultimately met. In a similar vein, the EAH suggests that misaligned task expectations increase cognitive demands even in the absence of actual misalignment. For instance, imagine a third-person video modeling example in which an instructor visually demonstrates the steps of applying a tourniquet to their own arm (e.g., threading the strap, tightening the windlass). When trainees watch the video, they may implicitly expect that they will later replicate the task from a first-person angle. This expectation can prompt them to mentally transform each step, thus increasing cognitive load. Even if the trainees are later tasked with applying the tourniquet to a partner's arm – a scenario that does not require the same spatial transformations since the steps are spatially similar to those shown in the video – their performance may still be impaired due to the mental transformation triggered by their initial expectation during the observation phase.

Presumably, the expectation of having to perform tasks from a first-person angle arises by default, which explains why the perspective effect appears in studies where expectation was not deliberately manipulated (e.g., De Koning et al., 2022; Fiorella et al., 2017). This expectation likely stems from the physical reality that people can only perform actions from their own first-person perspective; for instance, one cannot perform a dance from a third-person angle. The fact that prior studies neither tested third-person task performance nor induced the expectation of third-person task performance may therefore have introduced a confound between recording angle and task expectation, such that first-person recordings were systematically aligned with participants' default expectations, whereas third-person recordings were systematically misaligned. Consequently, the observed benefits of first-person recordings may have reflected this alignment rather than an inherent advantage of the first-person recording angle itself.

Nonetheless, there are cases where a third-person angle can be replicated. In assembly tasks, for example, an object can be interacted with from an opposite angle, making the interaction similar to what is seen in a third-person recording, thus recreating a third-person task angle. These cases offer an opportunity to better understand the mechanisms responsible for the perspective effect: The ERH would suggest an inherent advantage of the first-person recording angle, regardless of the task angle or expectations about it, whereas the EAH would favor the recording angle that matches the expectation about the task angle.

1.4. The present study

The present study examined the effects of recording angle in video modeling examples, with a specific focus on how alignment between the recording angle and respondents' expectations about the task angle influences learning. It presented participants with video modeling examples in which a circuit assembly task was recorded from different angles (recording angle: first-person vs. third-person). The physical circuit on which participants were to reconstruct the circuit was positioned next to the monitor that displayed the video, with its orientation intended to shape their expectations about the angle from which they would perform the task (expected angle: first-person vs. third-person). After the video, participants recreated the circuit as accurately as possible, but the orientation of the circuit varied depending on the assigned task angle (first-person vs. third-person), which did or did not align with the expected angle (unbeknownst to the participant). By varying these three angles (see Fig. 1), the study aimed to disentangle the Embodied Resonance Hypothesis (ERH) and the Expectation Alignment Hypothesis (EAH).

Embodied Resonance Hypothesis: Observing behavior recorded from a first-person angle receives more support from the mirror neuron system compared to observing from a third-person angle. The ERH would be substantiated by a main effect of recording angle, regardless of the expected or task angle. Specifically, compared to third-person recordings, first-person recordings would lead to better task performance and lower cognitive load during observation and during task execution.

Expectation Alignment Hypothesis: Alignment between the recording angle and expected angle avoids the need for mental transformations during observation and thus enhances performance relative to conditions in which the angles are misaligned. The EAH would be supported by interaction effects between the recording angle and the expected angle. Specifically, when the recording angle matches the expected angle, it should lead to better task performance and lower cognitive load during observation, compared to when there is a mismatch.

Task angle was primarily used as a control variable. While it does not play a role in the ERH – which merely concerns the angle of observation – it may be relevant for the EAH. Specifically, task angle may influence performance only in conditions where the recording and expected angles are aligned: If the recording angle and expected angle align but the task angle differs, the mismatching task angle is unexpected and unprepared for. This prompts participants to still perform mental transformations, albeit during task execution. This would impair performance and increase cognitive load during the task. In contrast, when there is already a misalignment between the recording and expected angle, participants would preemptively perform mental transformations during observation and the encoded memory representation already incorporates the transformation. It would then make little difference whether or not the task angle aligns with the recording angle.

While the two hypotheses are proposed as independent and non-interfering, both may be active simultaneously and hold to varying degrees. If both operate in parallel, their effects may be additive, such that a first-person recording could still confer an embodied advantage even under perfect alignment between angles. The present study thus examines to what extent the perspective effect holds when the alternative explanation is statistically controlled.

Lastly, in line with previous studies that included a measure of spatial ability (Fiorella et al., 2017; Leng et al., 2024), the present study explored whether mental rotation ability moderated the hypothesized effects. This was based on the assumption that the mental transformation required for the incorporated task involves rotating an object (Hegarty & Waller, 2004).

2. Methods

2.1. Participants

A total of 128 undergraduate psychology students from Erasmus University Rotterdam (EUR) participated in the study. These students (19 males, 107 females, and two with another gender identity) were between 18 and 33 years old ($M = 20.65$, $SD = 2.66$).

The experiment was conducted in two language versions: Dutch (73 Dutch students) and English (55 international students with moderate to native English proficiency). Participation occurred through self-selection using the EUR research administration system, which incentivized participation with 1 h of research credits (a curricular requirement).

A simulation-based power analysis for mixed-design ANOVAs (Lakens & Caldwell, 2021) indicated that 100 to 132 participants were required to detect moderate main and interaction effects. The analysis assumed effect sizes ranging between $\eta^2 = 0.02$ and $\eta^2 = 0.08$, based on prior findings (De Koning et al., 2022; Fiorella et al., 2017), with 80% power and a significance level of 0.05.

The study was conducted in accordance with ethical standards and received approval from the Faculty Ethics Review Board of Utrecht University. All participants received an information letter in advance and gave informed consent at the start of the study.

2.2. Design

The study had a 2 (recording angle; first-person vs. third-person) × 2 (expected angle; first-person vs. third-person) × 2 (task angle; first-person vs. third-person) mixed design with recording and expected angles as between-subjects factors and task angle as within-subjects factor.

Participants performed two trials and therefore two task versions (i.e., circuit configurations) were used. The assignment of conditions and order of task versions was fully counterbalanced (see Appendix A).

2.3. Materials

The materials and procedure were meticulously designed and rigorously tested. A detailed account of the design decisions, underlying rationale, and steps taken to establish the reliability and validity of the measures can be found in the legitimation report stored on an Open Science Framework (OSF) page for this project.[2]

2.3.1. Task

The learning task involved recreating a configuration on an electrical circuit board (Fig. 2) from memory, based on a video modeling example. It utilized an educational tool that allowed users to experiment with electrical components to activate a device (e.g., a speaker or a fan). This circuit was located next to the monitor on which participants watched the video and was either oriented to resemble a first-person or third-person angle. To induce expectations about the task angle, participants were informed that they would perform the task from this orientation. If the circuit was oriented opposite to the video, they were explicitly told that the task would have to be executed in opposite orientation relative to the video. The study incorporated two different circuit configurations: one activating a speaker and another powering a fan. Both configurations included a battery pack and a selection of 12 other components (Fig. 2, right). However, participants were provided with a selection of 17 components, some of which served as deliberate distractors. The battery pack and either the speaker or fan component were fixed in place at the start of the trials (Fig. 2, left), to provide cues about the circuits' orientation.

[2] Link to OSF page: https://osf.io/gpvew/overview?view_only=312a27bc5d3f4ffbb44642b99c6e757e.

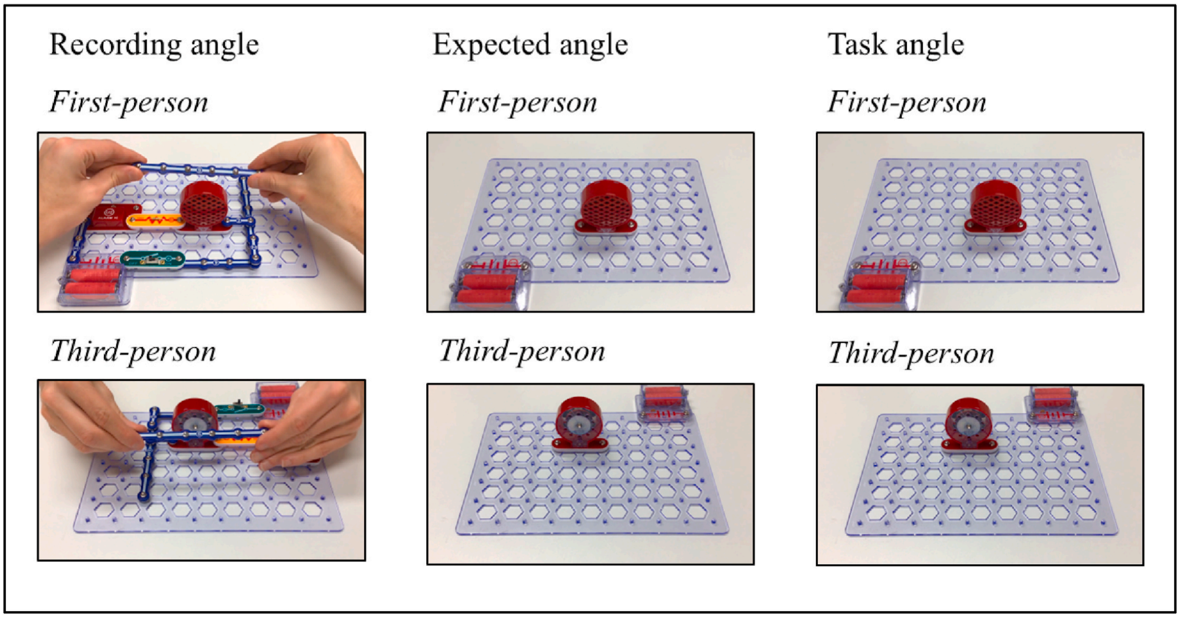

Fig. 1. Factors and conditions
Note. For recording angle, snapshots from the video are depicted. For the expected angle and task angle, the images depict the circuit board at the start of the task.

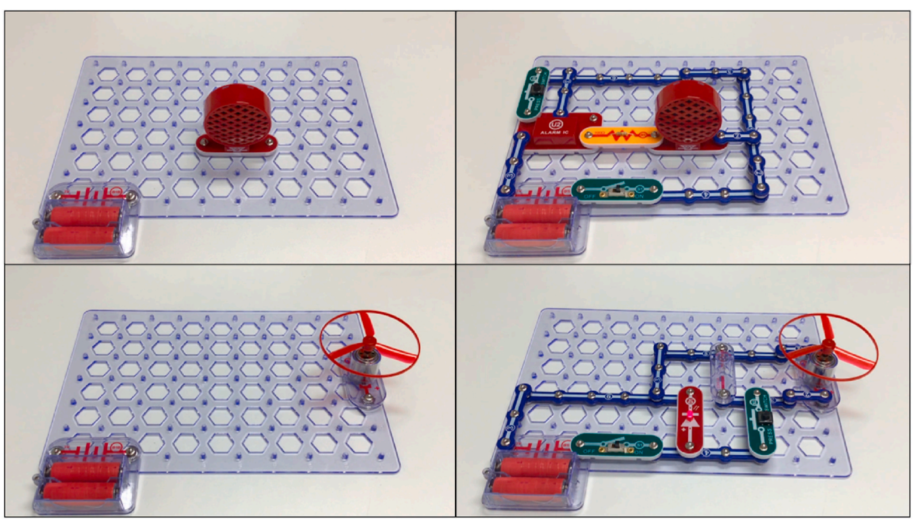

Fig. 2. Circuit configurations.

To assess task performance, participants' final configurations were scored using a flowchart (see OSF page). Scores were based on four criteria: correct use of component, precise placement on the circuit, orientation, and relative placement (i.e., connection to adjacent components). Each criterion contributed 0.25 points, totaling one point per component. Connector pieces (blue components in Fig. 2), where orientation was not applicable, were scored across three criteria (0.33 points each). The maximum total score per configuration was 12 points.

All 256 entries were graded by a single rater. To evaluate interrater reliability, a second rater independently assessed both attempts of the first 20 participants. As the number of scoring options varied per component, the quadratic weighted kappa was calculated per component. These values were then averaged across all components to provide an overall reliability estimate, $\kappa = 0.98$, indicating near-perfect agreement (McHugh, 2012).

2.3.2. Video modeling examples

The video modeling examples demonstrated the assembly of the electrical circuits. In total, eight videos were created (first-person and third-person versions for both configurations in two language versions), each lasting 155 s. They were displayed on a 24.1-inch LCD monitor (1920 × 1200 resolution), occupying the central 825 × 616 pixels of the screen.

All videos were recorded with a blank background, showing only the circuit board and the presenter's hands. The videos also included audio, including raw audio (e.g., the clicking noise when components were attached) and voice-over narration describing each step of the process. To ensure applicability across different recording angles, the narration avoided directional terms like left/right or top/bottom. An example statement is: *"Then, place an on/off switch horizontally between the battery pack and the connector piece with a length of four."*

2.3.3. Instruments

Except for the assembly task, the entire experiment was conducted digitally in Qualtrics, which included the video modeling examples, consent form, questionnaires, and other materials.[3]

2.3.3.1. Prior knowledge. Participants' prior knowledge was assessed through self-report. This included a question about their experience with the specific circuit kit used in the study, and eight yes/no items adapted from Fiorella et al. (2017). For instance, "I know the difference between serial and parallel circuits." The sum of items answered "yes" was used as the prior knowledge measure. Internal consistency of this self-report of prior knowledge was acceptable (Cronbach's $\alpha = .60$).

2.3.3.2. Cognitive load. To measure cognitive load, participants were asked to rate the effort they invested while watching the video using a 9-point scale, with 1 denoting very little effort and 9 denoting very much effort (Paas, 1992). A similar item was presented after they completed the task, this time asking the effort exerted during the attempt.

2.3.3.3. Demographics. The experiment concluded with demographic questions, including participants' gender and age. In the English version of the study, an additional 5-point item assessed participants' English proficiency, ranging from novice (1) to distinguished levels (5).

2.3.3.4. Mental rotation ability. Mental rotation ability was measured with the validated Mental Rotation Test (MRT; Peters et al., 1995). Each trial presented a reference figure alongside four response options. Participants were required to identify the two images that, when rotated, would match the reference figure. Following four example trials, participants completed two sets of 12 trials. They were allotted 3 min per set to complete as many items as possible. One point was awarded if both target images were correctly identified, resulting in a maximum possible score of 24. The test demonstrated good internal consistency (Cronbach's $\alpha = .83$).

2.4. Procedure

The study was conducted in individual sessions that lasted approximately 40 min. Participants entered the room where a physical circuit setup was placed next to the monitor (Fig. 3). They received both verbal and digital textual instructions about the study before providing consent. Then, participants watched a video modeling example, which was recorded from either a first-person or third-person angle (i.e., recording angle). The physical circuit next to the monitor was also oriented to depict either a first-person or third-person angle, and participants were explicitly informed that they would perform the task from this angle (i.e., expected angle).

After watching the video modeling example, participants rated their perceived cognitive load while watching the video. They then recreated the configuration on the physical circuit, with the circuit oriented to resemble a first-person or third-person angle (i.e., task angle). This meant that, just before starting the task, the circuit was rotated 180° for half of the participants, while the other half performed the task without rotation. After completing the task, participants rated the cognitive load they experienced during the task.

Thereafter, the second trial commenced. Participants whose circuit was rotated after watching the video were ensured that no rotation would occur in the second trial. This was done to maintain unbiased expectation manipulation in the second trial,[4] as the rotation explicitly violated their expectations (i.e., they suddenly had to perform the task from an angle opposite to the one instructed).

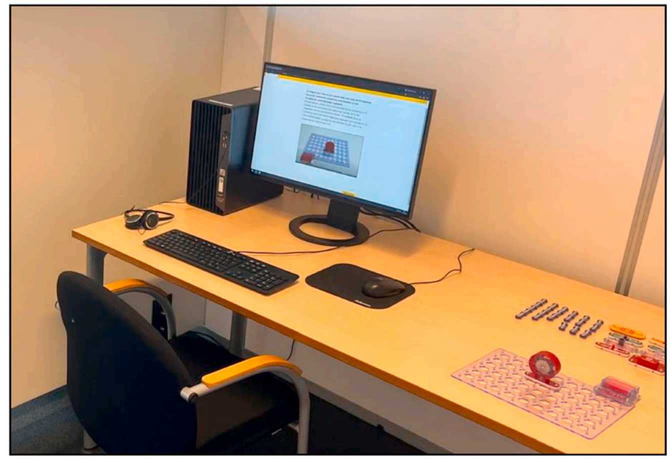

Fig. 3. Experimental setup.

The second trial was the same as the first trial but had two key differences. First, participants completed the task on a circuit with the opposite orientation (i.e., different task angle). Second, participants completed the remaining one of the two possible configurations. The whole procedure is shown in Fig. 4.

2.5. Data analysis

Statistical analyses were conducted with R Studio 4.5.0 (R Core Team, 2025). The R code, including the full list of used packages, is available on the OSF project page.

The main analyses comprised two steps. First, separate 2 (recording angle: first-person vs. third-person) × 2 (expected angle: first-person vs. third-person) × 2 (task angle: first-person vs. third-person) mixed-design ANOVAs were conducted for task performance, cognitive load during the video, and cognitive load during the task. Second, multiple linear mixed-effects models were run as post-hoc contrast tests. These contrasts were employed because, rather than testing all possible group comparisons, contrast tests focus on specific, theory-driven comparisons. This approach allowed for the comparison of matching conditions (first-person × first-person or third-person × third-person) versus mismatching conditions (first-person × third-person or third-person × first-person) in terms of their effects on task performance and cognitive load during the video and the task. The models were built using match dummy variables (dichotomous variables, coded as 1 if two conditions were the same and 0 if they differed). The models also included a three-way match dummy variable (coded as 1 when all angles were the same), if its inclusion accounted for a significant amount of additional variance.

This approach was chosen because the theoretical interest centered on angle alignment rather than on the effects of a particular expected or task angle. Operationalizing alignment through contrast-coded match variables allowed the testing of theoretically specified comparisons while reducing the number of estimated parameters, providing greater statistical sensitivity and power than omnibus interaction terms (Rosenthal et al., 1999). As higher-order interactions require

[3] The pilot study revealed individual differences in viewing behavior (e.g., considering the physical circuit's orientation or not). To capture these differences, two items were included at the end of the Qualtrics survey: 1) "While watching the video, did you consider the physical circuit next to you?" and 2) "Did you have a specific preference for the angle of the task?" However, these items did not correlate with any other variable and therefore were not included in the study.

[4] A linear model revealed no evidence that experiencing the rotation in Trial 1 weakened the effect of a match between the recording and expected angle in Trial 2 (interaction: $b = -1.09$, $SE = 0.83$, $t(124) = -1.32$, $p = .190$), suggesting that the expectation manipulation in the second trial remained unbiased.

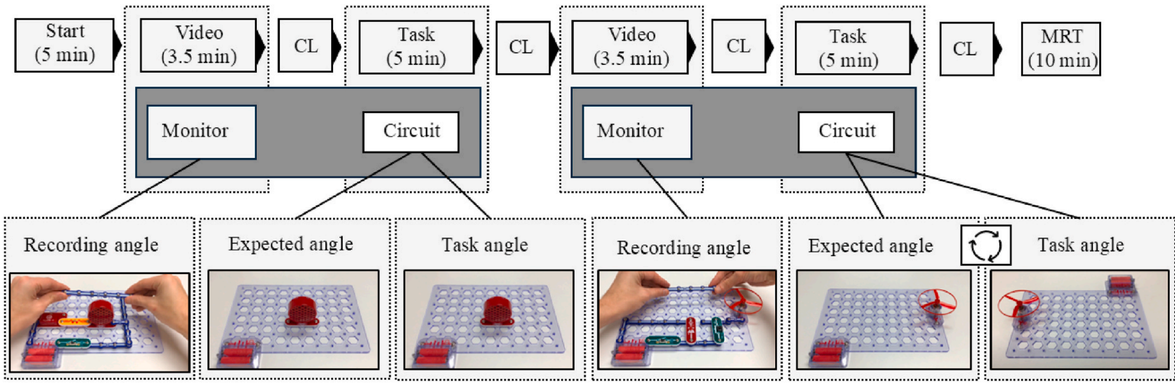

Fig. 4. *Procedure*
Note. "CL" refers to measuring cognitive load. Here, version 1 is depicted (see Appendix A).

substantially larger sample sizes to achieve adequate statistical power, the effect of full alignment was therefore reported only through the planned contrast rather than a three-way interaction.

Standardized regression coefficients in these models were interpreted as effect sizes, with values of 0.10, 0.30, and 0.50 considered small, medium, and large, respectively, following the guidelines proposed by Cohen (1988). In other analyses, Cohen's d was interpreted as effect size, with small, medium, and large effects corresponding to values of 0.20, 0.50, and 0.80, respectively.

2.5.1. *Statistical assumptions*

Inspection of histograms and Q-Q plots (observed residuals against expected normal values) revealed no severe violations of (multivariate) normality. Z-scores and studentized residuals (beyond $\pm$ 3.29) indicated no univariate outliers or influential cases. Lastly, plotting residuals against their fitted values revealed no severe heteroscedasticity in the compared groups.

3. Results

3.1. *Preliminary analyses*

Several comparisons were conducted to assess whether the analyses were unbiased by differences unrelated to the experimental manipulation. Table 1 presents the descriptive statistics for each configuration separately. Paired-samples t-tests comparing the two task configurations revealed that the fan configuration was more complex than the speaker configuration,[5] though the differences were moderately small. The fan configuration resulted in significantly lower task performance, $t(127) = -3.86$, $p < .001$, $d = 0.38$, and yielded more cognitive load, both while watching the video, $t(127) = 2.39$, $p = .019$, $d = 0.14$, and when completing the task, $t(127) = 3.00$, $p = .003$, $d = 0.35$.

Mixed-effects models were used to examine whether the order of the configurations (i.e., the speaker configuration first versus the fan configuration first) had any effect, but no such order effects were found (all $ps > 0.05$).

Lastly, the groups were compared on prior knowledge and mental rotation ability (MRT). Prior knowledge scores ranged from 0 to 9 ($M = 1.78$, $SD = 1.59$), and MRT scores ranged from 1 to 22 ($M = 9.91$, $SD = 4.56$). T-tests revealed that none of the groups compared in the main analyses (first-person vs. third-person, match vs. mismatch) differed on prior knowledge or MRT (all $ps > 0.05$).

3.2. *Main analyses*

3.2.1. *Task performance*

Means and standard deviations for each condition pairing are presented in Table 2. For task performance, the results of the mixed-design ANOVA revealed a significant main effect of recording angle, $F(1, 124) = 5.24$, $p = .024$, $\eta^2 = 0.02$, with higher performance when having watched first-person recordings ($M = 6.83$, $SD = 2.82$) compared to third-person recordings ($M = 6.08$, $SD = 3.01$). The analysis neither revealed a significant main effect of expected angle, $F(1, 124) = 1.69$, $p = .196$ (first-person $M = 6.24$, $SD = 3.27$; third-person $M = 6.67$, $SD = 2.55$), nor of task angle, $F(1, 124) = 0.44$, $p = .507$ (first-person $M = 6.37$, $SD = 2.77$; third-person $M = 6.55$, $SD = 3.10$). Significant interactions were found between recording and expected angle, $F(1, 124) = 79.20$, $p < .001$, $\eta^2 = 0.27$, recording and task angle, $F(1, 124) = 11.17$, $p = .001$, $\eta^2 = 0.04$, and expected and task angle, $F(1, 124) = 18.84$, $p < .001$, $\eta^2 = 0.06$.

Table 3 presents the results of the post-hoc linear mixed-effects models for task performance. The first column represents the base model, while the second column presents the model with the three-way match variable included. As the addition of the three-way match variable increased the explained variance ($\Delta R^2 = 0.04$, $p < .001$), the results of this model are reported.

The model revealed that participants who watched video modeling examples recorded from a third-person angle performed significantly worse on the task than those who watched them recorded from a first-person angle, $t(125) = -2.29$, $p = .024$. The effect of recording angle was moderately small.

In addition, the model revealed that a match between the recording angle and the expected angle yielded significantly better task performance compared to a mismatch, $t(125) = 8.74$, $p < .001$, corresponding to a moderate effect size. However, the model revealed that a match between the recording angle and the task angle did not yield significantly different task performance compared to a mismatch, $t(126) = -0.70$, $p = .483$. Furthermore, it was observed that when all angles were aligned – either all first-person or all third-person – task performance was significantly higher compared to when there was no full alignment, $t(126) = 4.33$, $p < .001$. The effect of this three-way alignment was moderately sized. Lastly, no significant effect was observed for a match between expected angle and task angle, $t(126) = -1.49$, $p = .139$ (absent in Table 3 as it was tested separately to avoid perfect collinearity). The effects of matching angles on task performance are depicted in the top panel of Fig. 5.

To evaluate the relative support for the ERH versus the EAH, the standardized coefficients from the final mixed-effects models were

[5] Due to discrepancies between the two configurations, the analyses were repeated using standardized outcome variables. Standardization involved subtracting the mean and dividing by the standard deviation within each configuration (Singer & Willett, 2003). The results were similar to those obtained with the original, non-standardized variables. Therefore, only the analyses based on the non-standardized variables are reported.

Table 1
Descriptive statistics per configuration.

	Speaker configuration				Fan configuration				Overall			
	M	*SD*	min	max	*M*	*SD*	min	max	*M*	*SD*	min	max
Task performance	7.01	2.85	0.00	12.00	5.90	2.93	0.33	11.75	6.46	2.94	0.00	12.00
Cognitive load (video)	5.09	2.16	1.00	9.00	5.39	2.15	1.00	9.00	5.24	2.15	1.00	9.00
Cognitive load (task)	5.98	2.08	1.00	9.00	6.66	1.82	2.00	9.00	6.32	1.98	1.00	9.00

Note. $N = 256$ ($n = 128$ per configuration).

Table 2
Means per condition pairing.

Factor			Task performance		Cognitive load			
Recording angle	Expected angle	Task angle			(video)		(task)	
			M	*SD*	*M*	*SD*	*M*	*SD*
first-person	first-person	first-person	8.83	2.49	4.56	2.02	5.72	2.26
first-person	first-person	third-person	7.33	2.84	4.69	2.38	6.44	2.11
first-person	third-person	first-person	5.57	2.42	5.62	1.84	6.00	1.95
first-person	third-person	third-person	5.61	2.19	5.84	1.83	6.81	1.79
third-person	first-person	first-person	4.66	2.38	5.72	2.11	6.53	1.85
third-person	first-person	third-person	4.16	2.97	5.88	2.14	6.56	1.63
third-person	third-person	first-person	6.40	1.97	4.88	2.17	6.88	1.72
third-person	third-person	third-person	9.10	1.85	4.72	2.36	5.66	2.25

Table 3
Linear mixed-effects models on task performance.

	Task performance (1)	Task performance (2)
Intercept	4.92***	5.51***
	(0.32)	(0.34)
	0.13	*0.13*
Recording angle (third-person)	−0.75*	−0.75*
	(0.33)	(0.33)
	−0.26	*−0.26*
Match (recording, expected)	2.92***	1.73***
	(0.33)	(0.43)
	0.50	*0.30*
Match (recording, task)	0.91**	−0.27
	(0.29)	(0.39)
	0.16	*−0.05*
Match (recording, expected, task)		2.37***
		(0.55)
		0.35
R^2	0.29	0.33

Note. $N = 256$ (2 trials per individual). Main entries are unstandardized regression coefficients, with entries in parentheses representing standard errors. Entries in italics are standardized regression coefficients.
*$p < .05$.
**$p < .01$.
***$p < .001$.

compared. For task performance, the effect of the ERH (recording angle) was relatively small (β = −0.26). In contrast, the effect of the EAH (expected angle) is distributed across two components due to the significant interaction effect (Hayes, 2013). Therefore, marginal effects were calculated to reflect the combined contribution of the main and interaction effects, yielding a standardized effect of β = 0.47. These findings show that expectation alignment with the recording angle explained substantially more variance in task performance than the recording angle itself, as reflected by a standardized effect that is approximately 1.8 times larger.

3.2.2. Video cognitive load

Regarding cognitive load participants experienced while watching the video, the mixed-design ANOVA showed no significant main effects: recording angle, $F(1, 124) = 0.11$, $p = .738$ (first-person $M = 5.18$, $SD = 2.08$; third-person $M = 5.30$, $SD = 2.23$); expected angle, $F(1, 124) = 0.02$, $p = .876$ (first-person $M = 5.21$, $SD = 2.22$; third-person $M = 5.27$, $SD = 2.09$); task angle, $F(1, 124) = 0.43$, $p = .514$ (first-person $M = 5.20$, $SD = 2.07$; third-person $M = 5.28$, $SD = 2.24$). However, the analysis yielded a significant interaction between recording angle and expected angle, $F(1, 124) = 9.08$, $p = .003$, $\eta^2 = 0.06$, but neither between recording angle and task angle, $F(1, 124) = 0.43$, $p = .514$, nor between expected angle and task angle, $F(1, 124) = 0.17$, $p = .678$.

Table 4 presents the results of the linear mixed-effects models for cognitive load during the video and cognitive load during the task. For cognitive load during the video, the addition of the three-way match variable did not explain a significant additional amount of variance ($\Delta R^2 < 0.01$, $p = .673$). Therefore, the model without the three-way match variable was reported. The model revealed that if there was a match between the recording angle and the expected angle, cognitive load while watching the video was significantly lower compared to if there was a mismatch, $t(125) = -3.03$, $p = .003$ (see Fig. 5, middle panel). This effect was moderately sized.

For cognitive load during the video, only the match between recording angle and expected angle was significant (β = −0.25), whereas recording angle itself showed no significant effect (β = 0.05). These findings show that the variance in cognitive load during video observation was explained by expectation alignment (EAH), but not by recording angle per se (ERH).

3.2.3. Task cognitive load

For cognitive load participants experienced during the task, the mixed-design ANOVA revealed no significant main effects: recording angle, $F(1, 124) = 0.39$, $p = .533$ (first-person $M = 6.24$, $SD = 2.05$; third-person $M = 6.41$, $SD = 1.91$); expected angle, $F(1, 124) = 0.01$, $p = .929$ (first-person $M = 6.31$, $SD = 1.98$; third-person $M = 6.34$, $SD = 1.99$); or task angle, $F(1, 124) = 0.15$, $p = .704$ (first-person $M = 6.28$, $SD = 1.98$; third-person $M = 6.37$, $SD = 1.98$). The analysis neither revealed a significant interaction between recording and expected angle, $F(1, 124) = 1.35$, $p = .248$, nor between expected and task angle, $F(1, 124) = 1.65$, $p = .202$. However, the analysis did reveal a significant interaction effect between recording angle and task angle, $F(1, 124) = 9.10$, $p = .003$, $\eta^2 = 0.03$.

In the post-hoc linear mixed effects models, the addition of the three-way match variable did not explain a significant additional amount of variance in cognitive load during the task ($\Delta R^2 < 0.01$, $p = .198$). Therefore, the model without the three-way match variable was

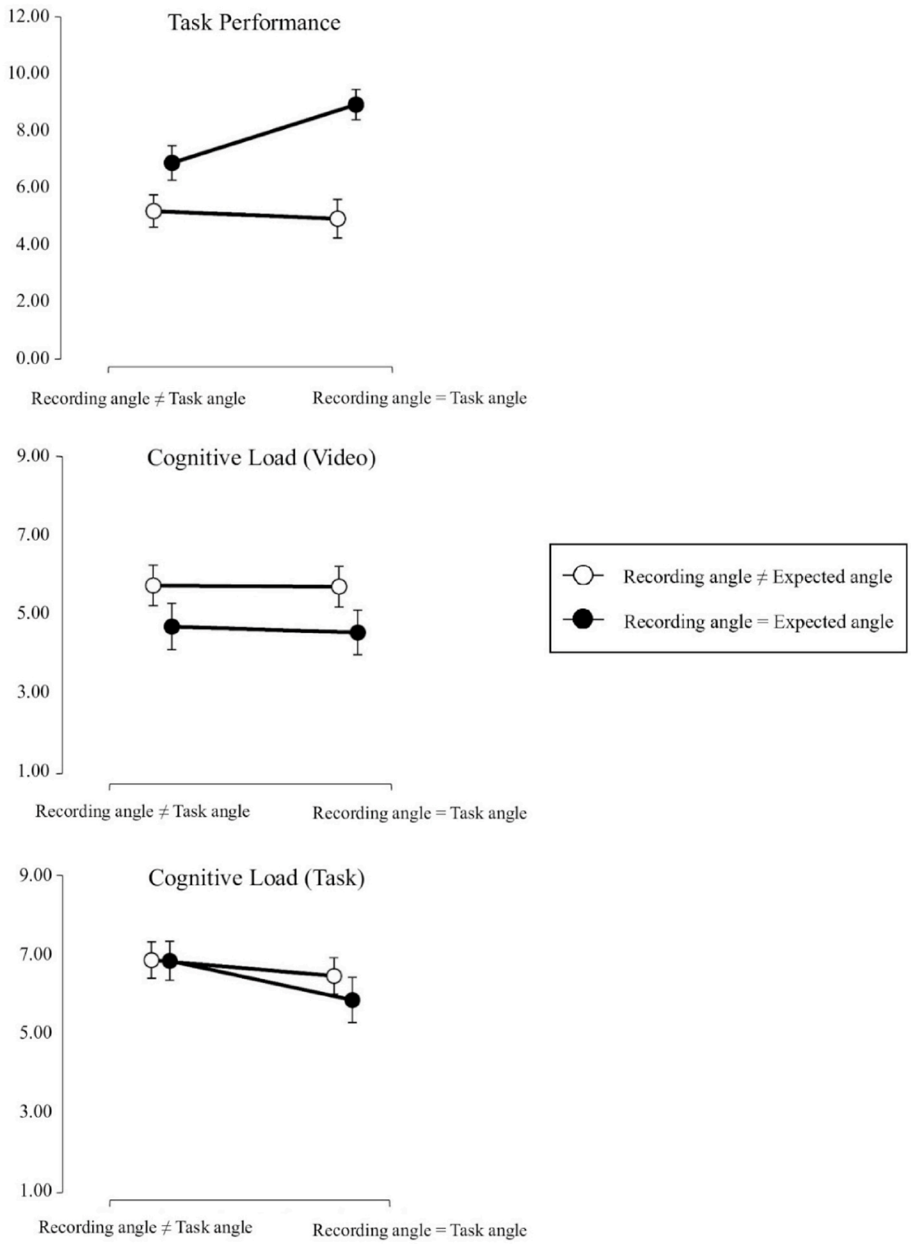

Fig. 5. Task performance and cognitive load across alignment versus misalignment
Note. Error bars represent standard errors of the mean.

reported. The model revealed that if there was a match between the recording angle and the expected angle, cognitive load while performing the task was significantly lower compared to if there was a mismatch, $t(125) = -2.99$, $p = .003$ (see Fig. 5, bottom panel). This effect was small; ($\beta = -0.17$).

3.3. Exploratory analyses

3.3.1. Knowing about misalignment

The design of the present study allowed for an exploration of the effects of informing participants in advance about a task angle that would not align with the recording angle. To investigate this, a subsample of 128 participants (who performed the task from a different angle than shown in the video) was analyzed. Half of them ($n = 64$) were informed about the misalignment beforehand, while the other half were not, as the circuit orientation changed just before task execution. Independent-samples t-tests revealed that those who were informed performed significantly worse than those who were not, $t(126) = -4.09$, $p < .001$, $d = -0.72$ (large effect, mean difference = −1.73), and reported higher cognitive load during the video, $t(126) = 2.67$, $p = .008$, $d = 0.72$ (medium effect, mean difference = 1.00). However, there was no significant difference in cognitive load during the task, $t(126) = -0.05$, $p = .962$.

3.3.2. Mental rotation ability

Appendix B presents the linear mixed-effects models of the main analysis again, this time including MRT (as main effect and as moderator for the alignment effects that were significant in the main analyses).

Adding MRT to the models led to a significant amount of additional variance being explained in task performance ($\Delta R^2 = 0.03$, $p = .028$) and in cognitive load during the task ($\Delta R^2 = 0.09$, $p < .001$), but not in cognitive load during the video ($\Delta R^2 = 0.02$, $p = .164$).

In the model for task performance, MRT did not have a significant main effect on task performance ($t(123) = -0.44$, $p = .659$). However,

Table 4
Linear mixed-effects models on cognitive load during the video and task.

	Cognitive load (Video)	Cognitive load (Task)
Intercept	5.75**	6.73**
	(0.31)	(0.25)
	−0.03	*−0.04*
Recording angle (third-person)	0.12	0.16
	(0.35)	(0.26)
	0.05	*0.08*
Match (recording, expected)	−1.05*	−0.30
	(0.35)	(0.26)
	−0.25	*−0.08*
Match (recording, task)	−0.09	−0.68*
	(0.13)	(0.23)
	−0.02	*−0.17*
R^2	0.06	0.04

Note. N = 256 (2 trials per individual). Main entries are unstandardized regression coefficients, with entries in parentheses representing standard errors. Entries in italics are standardized regression coefficients.
*$p < .01$.
**$p < .001$.

MRT significantly moderated the effect of alignment between the recording angle and the expected angle ($t(123) = 2.51$, $p = .013$, $SE = 0.08$, $\beta = 0.23$), with a moderately small effect size. Specifically, the higher their MRT ability, the more participants appeared to benefit from alignment between these angles. MRT did not significantly moderate the effect of a three-way match on task performance ($t(125) = -1.40$, $p = .164$).

In the model for cognitive load during the task, MRT had a significant main effect ($t(126) = -2.36$, $p = .020$, $SE = 0.04$, $\beta = -0.20$). On average, individuals with higher MRT scores experienced less cognitive load during the task (a moderately small effect). However, MRT did not significantly moderate the effect of alignment between the recording angle and the task angle ($t(126) = -1.57$, $p = .118$).

4. Discussion

The present study explored the effect of recording angle in video modeling examples, building on prior research that has consistently shown first-person recordings to be more effective for learning than third-person recordings (De Koning et al., 2022; Fiorella et al., 2017; Garland & Sanchez, 2013; Leng et al., 2024). This phenomenon, known as the perspective effect (cf. Fiorella et al., 2017), has been attributed to different underlying mechanisms, but a definitive explanation has yet to emerge. The present study addresses this gap by distinguishing and empirically evaluating two competing hypotheses for the perspective effect.

The Embodied Resonance Hypothesis (ERH) suggests that the mirror neuron system preferentially processes a first-person perspective, leading to an inherent learning benefit when observing first-person recordings. Alternatively, the Expectation Alignment Hypothesis (EAH) proposes that the perspective effect is the result of an anticipated need for mental transformations required to replicate actions observed from a third-person angle, as tasks are typically performed from a first-person angle. In particular, the EAH suggests that merely expecting a misaligned task angle already prompts these mental transformations. If the EAH holds, the disadvantage of third-person recordings compared to first-person recordings would diminish when individuals also expect to perform the task from a third-person angle.

To investigate these competing explanations for the perspective effect, the present study employed a mixed-design approach: It manipulated the recording angle of the video modeling example (first-person vs. third-person), the angle from which participants expected to perform the task (expected angle: first-person vs. third-person), and the angle from which participants actually executed the task (task angle: first-person vs. third-person). Through examination of their combined effects on task performance and self-reported cognitive load, the present study revealed that both the ERH and the EAH hold merit.

4.1. Evidence supporting the Embodied Resonance Hypothesis

The current findings suggest that watching a video modeling example recorded from a first-person angle resulted in better task performance compared to watching a video modeling example recorded from a third-person angle. As recording angle was not measured in isolation, this finding shows an *inherent* advantage of the first-person recording angle (i.e., advantage regardless of the expected angle or task angle) and serves as evidence for the ERH. This evidence aligns particularly well with neuroscientific research that suggests that a first-person recording angle elicits greater activation of the mirror neuron system and other brain areas related to action execution (Jackson et al., 2006; Leng et al., 2024). However, as the present study did not directly measure neural activity, the findings should be interpreted as behavioral support for the ERH rather than neuroscientific evidence of it.

No differences between the recording angles were found in cognitive load, either when watching the video or during the task, suggesting that cognitive load did not play a role in the ERH. This finding stands somewhat in contrast to the ERH, as enhanced mirror neuron activation should facilitate more automatic and less effortful understanding of actions observed from a first-person angle (Jackson et al., 2006). Nevertheless, failing to find an effect of recording angle on cognitive load is a recurring pattern in research (Fiorella et al., 2017; Leng et al., 2024). A possible explanation is the reliance on self-report measures to assess cognitive load. Since mirror neuron activation occurs primarily at an implicit level (Chong & Mattingley, 2008), participants may have been unaware of any reduction in cognitive load resulting from enhanced mirror neuron support.

4.2. Evidence supporting the Expectation Alignment Hypothesis

In addition to the main effect of recording angle, there were effects of recording angle that depended on its alignment with the expected angle and task angle, as suggested by the EAH: When participants expected to perform the task from an angle opposite to the recording angle, they expended significantly more effort processing the video and were less capable of replicating the circuit configuration (i.e., lower task performance). Thus, expecting to perform the task from the same angle as the video led to better performance, regardless of whether that was first- or third-person.

Furthermore, in isolation, actually performing the task from an angle opposite to the recording angle did not impact task performance, though it increased cognitive load during the task. Yet, in cases where participants did not anticipate misalignment (recording angle = expected angle ≠ task angle), performing the task from an opposite angle *did* impair task performance.

According to the EAH, these effects of angle alignment occur because participants spatially transform the observed information, which increases cognitive load and leads to errors. Following the EAH, when participants knew in advance that the task would require a different angle, they prepared for the misalignment while watching the video and thus conducted the mental transformation at that stage. This would explain why cognitive load increased during the video but not during the task; the observed behavior was already stored in memory in its rotated form. Conversely, if there was no expectation to perform the task from a rotated angle, no mental transformations occurred when watching the video, and cognitive load did not increase during the video. However, if the task then had to be performed from a rotated angle, mental transformations were still necessary – though they occurred during the task instead. This led to impaired performance and higher cognitive load during the task.

This explanation also accounts for the three-way alignment effect on performance: Performance was best when all angles aligned, as no transformations were needed at all. Based on the same reasoning, one would expect a similar pattern for cognitive load during the task, which was not the case. Although this effect did not reach significance, it may be due to limited power, as the data visually points to a small reduction in cognitive load when all angles align (Fig. 5). This may warrant further investigation.

The evidence for the EAH aligns with earlier studies that linked the third-person perspective to brain regions associated with spatial processing and memory (Leng et al., 2024; Ruby & Decety, 2001; Watanabe et al., 2013), highlighting the central role of mental transformations and the burden they place on working memory. Building on these prior studies, the present study showed that mental transformations can occur even while watching first-person recordings – if participants expect to perform the task from a third-person angle.

The evidence found in the present study for the EAH differs from previous findings in that it revealed not only effects on task performance, but also effects on cognitive load – where earlier studies did not (De Koning et al., 2022; Fiorella et al., 2017; Leng et al., 2024). Leng et al. (2024) attributed the absence of an effect in their study to the task being too easy, as both they and Fiorella et al. (2017) found that the perspective effect depends on task difficulty. The additional cognitive effort required for mental rotation may indeed be more noticeable in tasks with higher intrinsic cognitive load, as these are more likely to approach or exceed working memory capacity (Paas & Sweller, 2021). In the present study, cognitive load was relatively high and task performance relatively low, making this a plausible explanation for the discrepancy between the present findings and those of previous studies.

4.3. *What drives the perspective effect?*

Overall, the present study found support for both the ERH and the EAH, suggesting that both play a role in the perspective effect proposed by previous studies (e.g., De Koning et al., 2022; Fiorella et al., 2017; Leng et al., 2024). In the present study, the effects related to the EAH explained more variance in performance compared to the moderately small main effect of recording angle (ERH). This suggests that the EAH contributes more to the perspective effect than the ERH, though both mechanisms seem to play a role.

The current findings neither directly support nor challenge the perspective effect. Rather, they refine the concept by showing that it can be divided into a "pure" perspective effect (reflecting the ERH) and a perspective effect caused by expectation alignment (EAH). Prior findings on the perspective effect likely contain variance attributable to both the ERH and to the fact that tasks were (or were expected to be) executed from a first-person angle, creating alignment with first-person recordings (EAH).

While the EAH seems to be the more pronounced contributor in the present study, the relative weight of the EAH versus the ERH may vary across different learning contexts. Importantly, the present experiment used a circuit assembly task, which represents a specific type of procedural learning situation involving object manipulation and spatial configuration. It used an object-oriented task centered around getting an object into a desired state – drawing more on the ability to mentally rotate the object rather than on exact motor movements that benefit most from mirror neuron support. Consequently, the findings of the present study likely generalize best to similar object-oriented tasks. While the perspective effect itself has been observed across a range of different learning tasks and contexts, the relative contribution of the underlying mechanisms may vary. For instance, tasks such as knot tying (De Koning et al., 2022) or paper folding (Leng et al., 2024) rely more strongly on precise motor movements, which may give greater weight to the ERH. This potential variation in the mechanisms underlying the perspective effect across different types of tasks represents an interesting avenue for future research.

4.4. *Is knowledge about misalignment beneficial?*

On a practical note, when participants had to replicate the demonstrated configuration from an angle opposite to that of the video, knowing this upfront substantially impaired performance and increased cognitive load during the video. This suggests that transforming the observation on the spot (during task execution) leads to better performance than transforming it during observation. The likely reason is that mentally transforming the information during video observation limits available capacity for processing and disrupts encoding in memory. In contrast, when rotations occur during task execution, memory encoding remains intact, and fewer errors due to missing procedural information occur. Overall, this suggests that if the task angle does not align with the recording angle, it may be beneficial to withhold this information until the task is attempted.

Building on this interpretation, the findings point to a more general principle regarding the timing of cognitively demanding adjustments. Complex operations such as mental rotation appear to compete with encoding processes when executed during initial exposure to instructional material. When learners anticipate the need for transformation, they may proactively restructure the information while it is still being processed, thereby taxing working memory resources and compromising the accuracy of the stored representation. In contrast, when procedural information is first encoded in its presented form, subsequent transformations can operate on a consolidated memory trace rather than on weakly processed input. This sequence – encode first, adapt later – may therefore support both accuracy of the representation and flexible application.

More broadly, this pattern suggests that in learning contexts requiring spatial or structural adaptation, it may sometimes be advantageous not to foreground these demands during initial exposure. Future research might examine whether this temporal separation between encoding and transformation generalizes to other domains, such as technical instruction, repair tasks, or complex procedural learning in which information must later be adapted to specific situational constraints.

4.5. *Mental rotation ability*

Including mental rotation ability (MRT) as a moderator did not reveal consistent patterns that offer deeper insights into the perspective effect. Moderation was found only for task performance: The results showed that high-MRT individuals benefited more from alignment between the recording and expected angle than low-MRT individuals. Both the lack of broader moderating effects of MRT and the particular moderation that *did* emerge are surprising within the EAH framework. High-MRT individuals are generally more skilled at mental transformations (Hegarty & Waller, 2004), and according to the EAH, these transformations are what drive the perspective effect. Based on this, one would expect high-MRT individuals to be less hindered by the need for mental rotation – but the findings for task performance suggest the opposite.

One possible explanation is that high-MRT individuals relied more heavily on mental rotation strategies, whereas low-MRT individuals may have adopted alternative strategies that involved less rotation (e.g., remembering which component is connected to which). This would explain why low-MRT individuals were less affected by misalignment than high-MRT individuals. Exploring these strategies, perhaps through eye-tracking, could be an interesting avenue for future research.

In addition, the required mental transformation in the learning task may have involved a blend of different spatial abilities, such as 2D and 3D mental rotation, perspective-taking and spatial sequential memory. This complexity may obscure the effects of any single spatial ability measure, which could also explain why prior research failed to find a role of spatial ability measures in the perspective effect (Fiorella et al., 2017; Leng et al., 2024).

4.6. Limitations and future directions

One limitation of the present study is the absence of neurological measures like fNIRS to validate the proposed explanations. While the present findings align with neuroscientific evidence (e.g., Leng et al., 2024), it would be valuable to further validate the hypotheses by tying the proposed neural mechanisms (i.e., mirror neuron activation and mental transformations) to the observed effects. In the present study, these mechanisms were inferred from prior literature (e.g., Jackson et al., 2006; Leng et al., 2024), and they were tested by examining whether corresponding behavioral patterns emerged. Consequently, the findings cannot provide direct evidence that the observed patterns reflect these specific neural processes.

Furthermore, the EAH posits that in real-world contexts – where expectations are not deliberately manipulated – individuals inherently assume they will perform tasks from a first-person angle, causing third-person recordings to misalign with these implicit expectations. A limitation of the present study is that it did not directly investigate the presence of these implicit expectations, which could be explored in future research. However, evidence for such implicit expectations might be found in studies showing activation in brain areas related to mental rotation when participants view third-person recordings, despite expectations not being deliberately manipulated (e.g., Leng et al., 2024; Vogeley & Fink, 2003; Watanabe et al., 2011). A default expectation to perform tasks from a first-person angle would explain these findings.

Another limitation of the present study is the gender imbalance in the sample (19 males, 107 females, and two participants identifying with another gender identity). Prior research suggests that the processes underlying the proposed mechanisms – namely mirror neuron activation and spatial transformation – may differ on average between men and women. Specifically, women tend to show stronger neural responses during action observation (as indexed by mu suppression; Cheng et al., 2006, 2008), whereas men generally demonstrate stronger mental rotation abilities (Masters & Sanders, 1993). Although the perspective effect itself has been observed in studies with both balanced samples (De Koning et al., 2022) and unbalanced samples with a predominance of female participants (Fiorella et al., 2017; Leng et al., 2024), the present study specifically investigates competing explanations that rely on mechanisms associated with mirror neuron activation and spatial transformation abilities. Consequently, while the presence of the perspective effect across both balanced and predominantly female samples suggests that gender imbalance may not have been a decisive factor in the present study, the possibility cannot be excluded that the imbalanced sample influenced the relative prominence of the ERH and the EAH in the observed results.

A potential avenue for future research is to further examine how spatial discrepancies between video modeling examples and the associated tasks, or individuals' expectations regarding these discrepancies, influence learning. Intriguing areas for exploration include, for instance, using different devices (e.g., watching a video of soldering components onto a specific motherboard, while expecting to – or having to – solder components to a motherboard with a different layout). Another example is examining how motor movements performed by a right-handed individual are perceived and learned by left-handed viewers and vice versa (e.g., in a video modeling example on sign language).

Furthermore, the current body of research, including the present study, is limited in that all experiments focused on novel and specific procedures (e.g., a particular circuit configuration or knot), which were then tested only once. However, in real-world contexts, newly acquired skills are often applied multiple times, raising the question of whether the perspective effect persists with repeated practice. For instance, nurses bandage lacerations repeatedly but might view a corresponding video modeling example only once. If the perspective effect fades quickly with repeated practice, the advantages of recording from a specific angle may not outweigh the practical burdens (e.g., wearing a GoPro vs. using a static camera). This issue warrants further investigation.

4.7. Practical contributions

Although the present study primarily aimed to delineate the mechanisms underlying the perspective effect, its findings also inform and contextualize existing instructional design guidelines for video modeling examples (Mayer et al., 2020): More critical than the recording angle itself seems to be its alignment with the (expected) task angle. While this task angle is typically first-person, objects can sometimes be rotated – as in the present study – to recreate a third-person task angle. In such cases, learners may actually benefit from third-person recordings.

Educational practitioners can utilize this approach when third-person recordings offer additional advantages. For instance, third-person recordings often capture more contextual information, such as the presenter's gaze or relevant background objects (Debarba et al., 2017). Additionally, a presenter's hands can often block important information, and the amount of blocked information can vary per recording angle. Finally, third-person recordings generally require no specialized equipment, whereas first-person recordings often demand equipment such as an advanced camera mount.

4.8. Conclusion

While the present study supports the idea of a pure perspective effect, it challenges the notion that first-person recordings are universally superior for video modeling examples compared to third-person recordings. Instead, the findings suggest only a modest inherent advantage of first-person recordings, likely due to enhanced mirror neuron activation. More importantly, alignment with learners' expectations regarding the task angle appears to play a critical role: Expecting a misaligned angle triggers cognitively demanding mental rotations, which in turn increase cognitive load and impair learning.

CRediT authorship contribution statement

Tristan H.S. de Jonge: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Bjorn B. de Koning:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Margot van Wermeskerken:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declarations of interest

None.

Acknowledgements

The authors thank Hajo T. Berends for providing the voice-over narration used in the English version of the instructional videos. The authors also thank Timo P. Kegelaar for independently grading a subset of participants' attempts to assess interrater reliability.

Appendix A. Experiment Versions

Version	Trial 1: Speaker			Trial 2: Fan		
	Recording angle	Expected angle	Task angle	Recording angle	Expected angle	Task angle
1	first	first	first	first	first	third
2	first	first	third	first	first	first
3	first	third	first	first	third	third
4	first	third	third	first	third	first
5	third	first	first	third	first	third
6	third	first	third	third	first	first
7	third	third	first	third	third	third
8	third	third	third	third	third	first
Version	Trial 1: Fan			Trial 2: Speaker		
	Recording angle	Expected angle	Task angle	Recording angle	Expected angle	Task angle
9	first	first	first	first	first	third
10	first	first	third	first	first	first
11	first	third	first	first	third	third
12	first	third	third	first	third	first
13	third	first	first	third	first	third
14	third	first	third	third	first	first
15	third	third	first	third	third	third
16	third	third	third	third	third	first

Note. Entries reflect condition angle: "first" = first-person, "third" = third-person.

Appendix B. Models with Mental Rotation Ability

	Task performance	Cognitive load (Video)	Cognitive load (Task)
Intercept	5.70***	5.91***	7.56***
	(0.60)	(0.61)	(0.43)
	0.12	*−0.01*	*−0.03*
Recording angle (third-person)	−0.69*	0.06	0.13
	(0.32)	(0.34)	(0.24)
	−0.23	*0.03*	*0.07*
Match (recording, expected)	1.70***	−1.03*	−0.25
	(0.42)	(0.34)	(0.24)
	0.29	*−0.24*	*−0.06*
Match (recording, task)	−0.27	−0.09	−0.66
	(0.39)	(0.13)	(0.23)
	−0.05	*−0.02*	*−0.17*
Match (recording, expected, task)	4.16**		
	(1.39)		
	0.62		
MRT	−0.02	−0.01	−0.09*
	(0.05)	(0.06)	(0.04)
	−0.04	*−0.03*	*−0.20*
MRT × Match (recording, expected)	0.20*	−0.08	
	(0.08)	(0.36)	
	0.23	*−0.13*	
MRT × Match (recording, task)			−0.08
			(0.05)
			−0.13
MRT × Match (recording, expected, task)	−0.06		
	(0.05)		
	0.28		

Note. $N = 256$ (2 trials per individual). MRT refers to mental rotation ability. Main entries are unstandardized regression coefficients, with entries in parentheses representing standard errors. Entries in italics are standardized regression coefficients. First-person recording and mismatches were used as reference categories. The MRT interaction effects used mean-centered MRT values.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Appendix C. Interdependences

	First-person recording				Third-person recording			
	Performance	Video CL	Task CL	PK	Performance	Video CL	Task CL	PK
Video CL	−0.25**				−0.37***			
Task CL	−0.35***	0.37***			−0.37***	0.46***		
PK	0.02	−0.25**	−0.03		0.13	−0.34***	−0.22*	
MRT	0.19*	−0.14	−0.37***	−0.01	0.04	−0.13	−0.31***	0.10

Note. Pearson correlations. "CL" = cognitive load, "PK" = prior knowledge, "MRT" = mental rotation ability.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

References

Allé, M., Danan, F., Kwok, S., Davies, V., Prudat, C., & Berna, F. (2023). Differential influence of first- vs. third-person visual perspectives on segmentation and memory of complex dynamic events. *Consciousness and Cognition, 111*, Article 103508. https://doi.org/10.1016/j.concog.2023.103508

Bauer, R., Jost, L., Günther, B., & Jansen, P. (2021). Pupillometry as a measure of cognitive load in mental rotation tasks with abstract and embodied figures. *Psychological Research, 86*(5), 1382–1396. https://doi.org/10.1007/s00426-021-01568-5

Bochynska, A., Postma, A., Vulchanova, M., & Laeng, B. (2020). More mental rotation time does not imply more mental effort: Pupillary diameters do not change with angular distance. *Brain and Cognition, 148*, Article 105670. https://doi.org/10.1016/j.bandc.2020.105670

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H., & Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions. *Neuron, 42*(2), 323–334. https://doi.org/10.1016/s0896-6273(04)00181-3

Cheng, Y., Lee, P., Yang, C., Lin, C., Hung, D., & Decety, J. (2008). Gender differences in the mu rhythm of the human mirror-neuron system. *PLoS One, 3*(5), Article e2113. https://doi.org/10.1371/journal.pone.0002113

Cheng, Y., Tzeng, O. J., Decety, J., Imada, T., & Hsieh, J. (2006). Gender differences in the human mirror system: A magnetoencephalography study. *NeuroReport, 17*(11), 1115–1119. https://doi.org/10.1097/01.wnr.0000223393.59328.21

Chong, T. T. J., & Mattingley, J. B. (2008). Automatic and controlled processing within the mirror neuron system. In J. A. Pineda (Ed.), *Mirror Neuron Systems* (pp. 213–233). Humana Press. https://doi.org/10.1007/978-1-59745-479-7_10.

Ciorli, T., & Pia, L. (2023). Spatial perspective and identity in visual awareness of the bodily self-other distinction. *Scientific Reports, 13*(1), Article 14994. https://doi.org/10.1038/s41598-023-42107-z

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

De Koning, B. B., Mok, K., Marcus, N., & Ayres, P. (2022). Investigating the role of hand perspective in learning from procedural animations. *British Journal of Educational Psychology, 93*(S2), 251–269. https://doi.org/10.1111/bjep.12542

Debarba, H. G., Bovet, S., Salomon, R., Blanke, O., Herbelin, B., & Boulic, R. (2017). Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PLoS One, 12*(12), Article e0190109. https://doi.org/10.1371/journal.pone.0190109

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fiorella, L., Van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology, 109*(5), 653–665. https://doi.org/10.1037/edu0000161

Garland, T., & Sanchez, C. A. (2013). Rotational perspective and learning procedural tasks from dynamic media. *Computers & Education, 69*, 31–37. https://doi.org/10.1016/j.compedu.2013.06.014

Gorisse, G., Christmann, O., Amato, E. A., & Richir, S. (2017). First- and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users. *Frontiers in Robotics And AI, 4*. https://doi.org/10.3389/frobt.2017.00033

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*(2), 175–191. https://doi.org/10.1016/j.intell.2003.12.001

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction, 44*, 22–30. https://doi.org/10.1016/j.learninstruc.2016.02.004

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *NeuroImage, 31*(1), 429–439. https://doi.org/10.1016/j.neuroimage.2005.11.026

Jeannerod, M., & Frak, V. (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology, 9*(6), 735–739. https://doi.org/10.1016/s0959-4388(99)00038-0

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*(3), 724–739. https://doi.org/10.1037/edu0000013

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science, 4*(1). https://doi.org/10.1177/2515245920951503

Leng, X., Zhu, W., Mayer, R. E., & Wang, F. (2024). The viewing perspective effect in learning from instructional videos: A replication and neuroimaging extension. *Learning and Instruction, 94*, Article 102004. https://doi.org/10.1016/j.learninstruc.2024.102004

Liang, Z., Ga, R., Bai, H., Zhao, Q., Wang, G., Lai, Q., Chen, S., Yu, Q., & Zhou, Z. (2024). Teaching expectancy improves video-based learning: Evidence from eye-movement synchronization. *British Journal of Educational Technology, 56*(1), 231–249. https://doi.org/10.1111/bjet.13496

Lin, Y., & Yu, Z. (2023). A meta-analysis evaluating the effectiveness of instructional video technologies. *Technology, Knowledge and Learning, 29*, 2081–2115. https://doi.org/10.1007/s10758-023-09669-3

Lohmann, J., Rolke, B., & Butz, M. V. (2017). In touch with mental rotation: Interactions between mental and tactile rotations and motor responses. *Experimental Brain Research, 235*(4), 1063–1079. https://doi.org/10.1007/s00221-016-4861-8

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*(4), 337–341. https://doi.org/10.1007/bf01067434

Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology, Research and Development, 68*(3), 837–852. https://doi.org/10.1007/s11423-020-09749-6

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*(1431), 491–500. https://doi.org/10.1098/rstb.2002.1261

Ouwehand, K., Van Gog, T., & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. *Educational Technology & Society, 18*(4), 78–88.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*, 429–434.

Paas, F., & Sweller, J. (2021). Implications of cognitive load theory for multimedia learning. In R. E. Mayer & Clark (Ed.), *The Cambridge handbook of multimedia learning* (pp. 73–81). Cambridge University Press. https://doi.org/10.1017/9781108894333.009.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39–58. https://doi.org/10.1006/brcg.1995.1032

R Core Team. (2025). R: A language and environment for statistical computing. https://www.r-project.org/.

Rizzolatti, G. (2005). The mirror neuron system and imitation. In S. Hurley, & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science. Volume 1: Mechanisms of imitation and imitation in animals* (pp. 55–76). MIT Press. https://doi.org/10.7551/mitpress/5330.001.0001.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (1999). *Contrasts and effect sizes in behavioral research*. Cambridge University Press eBooks. https://doi.org/10.1017/cbo9780511804403

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience, 4*(5), 546–550. https://doi.org/10.1038/87510

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing, 17*(3), 279–305. https://doi.org/10.1007/s10339-016-0765-6

Torriero, S., Oliveri, M., Koch, G., Caltagirone, C., & Petrosini, L. (2007). The what and how of observational learning. *Journal of Cognitive Neuroscience, 19*(10), 1656–1663. https://doi.org/10.1162/jocn.2007.19.10.1656

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155–174. https://doi.org/10.1007/s10648-010-9134-7

Van Gog, T., Verveer, I., & Verveer, L. (2013). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education, 72*, 323–327. https://doi.org/10.1016/j.compedu.2013.12.004

Van Wermeskerken, M., Ravensbergen, S., & Van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences, 7*(1), 38–42. https://doi.org/10.1016/s1364-6613(02)00003-7

Vogt, S., Taylor, P., & Hopkins, B. (2003). Visuomotor priming by pictures of hand postures: Perspective matters. *Neuropsychologia, 41*(8), 941–951. https://doi.org/10.1016/s0028-3932(02)00319-6

Watanabe, R., Higuchi, T., & Kikuchi, Y. (2013). Imitation behavior is sensitive to visual perspective of the model: an fMRI study. *Experimental Brain Research, 228*(2), 161–171. https://doi.org/10.1007/s00221-013-3548-7.

Watanabe, R., Watanabe, S., Kuruma, H., Murakami, Y., Seno, A., & Matsuda, T. (2011). Neural activation during imitation of movements presented from four different perspectives: A functional magnetic resonance imaging study. *Neuroscience Letters, 503*(2), 100–104. https://doi.org/10.1016/j.neulet.2011.08.016

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/bf03196322