The influence of gestures and visuospatial ability during learning about movements with dynamic visualizations – An fNIRS study

Birgit Brucker [a,*], Björn de Koning [b], David Rosenbaum [c], Ann-Christine Ehlis [c], Peter Gerjets [a,d]

[a] *Leibniz-Institut für Wissensmedien, Schleichstr. 6, 72076, Tübingen, Germany*
[b] *Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062PA, Rotterdam, the Netherlands*
[c] *Department of Psychiatry & Psychotherapy, University Hospital Tübingen, Calwerstr. 14, 72076, Tübingen, Germany*
[d] *University of Tübingen, Schleichstr. 4, 72076, Tübingen, Germany*

ABSTRACT

There is increasing evidence that learning manual tasks from dynamic visualizations (e.g., origami folding) is facilitated when human hands are shown or gestures can be observed in the visualizations. This study examined whether observing and making gestures improves learning about non-human biological movements (i.e., fish locomotion) and whether gestures that correspond to the to-be-learned movement are superior to non-corresponding gestures. Moreover, learners' visuospatial ability was assessed as a possible moderator. Regarding underlying neurophysiological processes, functional near-infrared spectroscopy (fNIRS) was used to investigate whether gestures activate the human mirror-neuron system (hMNS) and whether this activation mediates the facilitation of learning. During learning, participants viewed animations that were supplemented with either a self-gesturing instruction (yes/no) and/or a gesture video (corresponding/non-corresponding/no gesture) resulting in six conditions (2x3-between-subjects design). Results showed that higher-visuospatial-ability learners benefitted from learning with non-corresponding gestures, whereas those gestures were detrimental for lower-visuospatial-ability learners. Furthermore, activation of the inferior frontal cortex (part of hMNS) tended to predict better learning outcomes. Making gestures did not influence learning, but participants observing corresponding gestures showed higher inferior frontal cortex activation if they self-gestured than when they did not self-gesture. Implications of the results for the design of instructional materials are discussed.

1. Introduction

Due to the rapid technological development of dynamic visualizations such as videos and animations in the recent decades (De Koning, Hoogerheide, & Boucheix, 2018), it has become easier and also more attractive to use these visualizations as instructional tools. They have become particularly popular to visualize processes and phenomena that are dynamic in nature, such as lightning formation, the cardiovascular system, or animal movements. Naturally, dynamic visualizations are well-suited to display dynamic phenomena given that they explicitly depict visuospatial information over time (e.g., Ploetzner, Berney, & Bétrancourt, 2021). Nevertheless, research thus far indicates that dynamic visualizations are not in all cases superior to learning from static visualizations (e.g., Castro-Alonso, Ayres, & Paas, 2016; Mayer, Hegarty, Mayer, & Campbell, 2005). Dynamic visualizations seem to be particularly effective for learning about dynamic phenomena when

biological movement is involved (see Höffler & Leutner, 2007) and for supporting understanding in learners with lower visuospatial ability (see Höffler, 2010). Examples of biological movements studied in prior research with dynamic visualizations relate to human biological movement, such as learning to tie knots (with the hands; see for example Marcus, Cleary, Wong, & Ayres, 2013) or learning to play the piano (with additional mimicking gestures, see for example, Mierowsky, Marcus, & Ayres, 2020). Learning from dynamic visualizations depicting human biological movements depends on observational learning, which is not only important in many social situations (e.g., Bandura, 1986), but also during learning how-to tasks, such as motor skills, by watching human models in real life as well as in visualizations (cf. Mierowsky et al., 2020). However, biological movements do not only comprise human movements, but also non-human biological movements, such as when learning to classify fish movement patterns (see for example Brucker, Ehlis, Häußinger, Fallgatter, & Gerjets, 2015), for which

---

dynamic visualizations also showed beneficial learning outcomes (e.g., Imhof, Scheiter, Edelmann, & Gerjets, 2012; Imhof, Scheiter, & Gerjets, 2011).

At present, there is a small, yet emerging, number of studies investigating the instructional potential of dynamic visualizations addressing biological movement and most of them focus on human biological movements in terms of hand-manipulative tasks (e.g., De Koning, Marcus, Brucker, & Ayres, 2019). Secondly, the existing studies mainly focused on investigating what people can learn from visualizations showing (human) biological movement; in other words, on how effective these visualizations are for learning. Except for the study by Brucker et al. (2015), there is no study that has investigated to what extent learning about (non-human) biological movements from dynamic visualizations can be enhanced by providing instructional support in terms of observing additional human gestures. Thirdly, it is yet fairly uninvestigated which role learners' visuospatial ability plays during learning about biological movements with dynamic visualizations and additional gestures. These three aspects provided the basis for the present study wherein we investigated the value of observing and making gestures when learners with different levels of visuospatial ability learn to classify fish movement patterns (i.e., non-human biological movement) from dynamic visualizations.

1.1. Learning from dynamic visualizations and gestures

So far, it is relatively well-established that both observing and making gestures is beneficial for acquiring knowledge about different scientific topics and spatial problem solving (e.g., Chu & Kita, 2011; Cook & Goldin-Meadow, 2006; Korbach, Ginns, Brünken, & Park, 2020; Valenzeno, Alibali, & Klatzky, 2003). These beneficial effects occur, because gesturing has the potential to promote learners' engagement (similar to other additional interventions during learning like drawing, Kastner et al., 2021; Stieff, 2017) and to guide learners' attention (De Koning & Jarodzka, 2017). In learning about movements from dynamic visualizations, there is also increasing evidence that observing hands or gestures improves learning outcomes. One line of research addresses learning procedural information from human movement tasks, such as knot tying or origami folding (e.g., De Koning et al., 2019; Marcus et al., 2013). In these hand manipulation learning tasks, the hands are inherently present as they are used to conduct the movements. However, with newer visualization techniques it is possible to not only show those movements with but also without the hands allowing for comparisons of dynamic visualizations with and without hands.

Results on the presence of hands in hand manipulation learning tasks are mixed so far (cf. Castro-Alonso, Ayres, & Paas, 2015). It might be the case that for these human movement tasks the visibility of the hands is not needed for learning as it is possible to observe and imitate such motor skill tasks effortlessly as they can be considered as biologically primary knowledge concepts (e.g., Geary, 2007, see also Ayres, Marcus, Chan, & Qian, 2009; Van Gog, Paas, Marcus, Ayres, & Sweller, 2009). Due to evolutionary processes humans have developed the ability to learn biologically primary knowledge concepts effortlessly because learners can rely on brain regions involved in the observation, understanding, and imitation of other persons' actions, that is the so-called human mirror-neuron system (hMNS; Fogassi & Ferrari, 2011; Rizzolatti & Craighero, 2004). This hMNS operates relatively effortlessly so it does not require cognitive load from learners that would be associated with deliberately processing the information in working memory (Ayres et al., 2009; Van Gog et al., 2009). It follows that, if observing and imitating human movement tasks that represents biologically primary knowledge and thus can be accomplished effortlessly, the visibility of the human body parts might not be necessary as the brain 'fills in' the blanks (i.e., missing hands) for such biologically primary tasks, so these tasks can be performed with and without observing the hands in the visualization.

However, this may vary for biological movement tasks, such as for example animal movements (e.g., Brucker et al., 2015) or non-biological movement tasks, such as lightning formation (e.g., De Koning & Tabbers, 2013), in which gestures or hands are depicted in addition to the learning visualizations that inherently did not contain hands or gestures. In these latter tasks (biological and non-biological movements) – that according to Geary (2007) would be classified as biologically secondary knowledge concepts (see also Ayres et al., 2009; Van Gog et al., 2009) – the addition of human movements by adding hands or gestures to the visualization (which can be considered the addition of biologically primary knowledge and thereby involving the hMNS) improves learning outcomes (Brucker et al., 2015; De Koning & Tabbers, 2013). It is assumed that these positive effects of observing hands or gestures are due to the activation of the hMNS (e.g., De Koning & Tabbers, 2011). This assumption might extend the hypothesis that the stimulation and involvement of the hMNS might be beneficial for learning about complex continuous aspects with dynamic visualizations in general (Ayres et al., 2009; De Koning & Tabbers, 2011; Van Gog et al., 2009). Dynamic visualizations usually impose the observer with higher working memory load than static visualizations due to their transient nature (e.g., Hegarty, 2004; Lowe, 1999). However, if the content of the dynamic visualizations can be mapped by the observers onto their own (human) body by anthropomorphizing them via the additionally depicted hands or gestures, such biologically secondary knowledge concepts might be processed more easily, because the hMNS is activated (and thus the biologically primary knowledge is addressed – by incorporating the hands or gestures). Thereby, the same neural networks might be activated that provide advantages in terms of additional cognitive resources for handling biologically primary knowledge concepts thus unburdening working memory. Thus, learners might have more working memory capacity available to process the contents (e.g., elaboration, reflection) and therefore benefit from dynamic in comparison to static visualizations (Van Gog et al., 2009). Up to now, it is still unclear, whether dynamic visualizations activate the hMNS per se, or which aspects of dynamic visualizations are accountable for such an activation. Yet, one promising approach to foster hMNS activation during learning with dynamic visualizations might be the depiction of hands and gestures under the aforementioned assumption that this hMNS activation is beneficial for learning.

A study by Brucker et al. (2015) provides first evidence for the hMNS assumption. In this study, lower- and higher-visuospatial-ability learners had to learn to discriminate fish movement patterns. The movements were displayed via dynamic visualizations whilst the participants observed additional gestures that did correspond (e.g., fingers of a human hand that are moved in the same undulating manner as fin spines of a dorsal fin) or did not correspond (e.g., moving the hands alternating up and down) to the depicted fish movements. Lower-visuospatial-ability learners who observed gestures that corresponded to the to-be-learned fish movements showed better learning outcomes and higher cortical activation in the inferior frontal cortex (which is part of the hMNS) compared to lower-visuospatial-ability learners who observed non-corresponding gestures. Higher-visuospatial-ability learners achieved high learning outcomes with both – corresponding as well as non-corresponding – gestures. Unexpectedly, learners who had neither higher visuospatial ability nor corresponding gestures at their disposal (i.e., lower-visuospatial-ability learners who observed the non-corresponding gestures) could also achieve high learning outcomes if they activated their inferior parietal cortex (which is also part of the hMNS). Thus, both an activation of the inferior frontal cortex and an activation of the inferior parietal cortex helped learners to discriminate different fish movements. The activation of parts of the hMNS might have provided the learners with additional resources to deal with the to-be-learned dynamic information (in this case the fish movements). These findings provide the first indication that the hMNS is not only involved in representing human (biological) movements but also non-human biological movements if the observer is able to map these movements onto the human body by

anthropomorphizing them (cf. De Koning & Tabbers, 2011; Engel, Burke, Fiehler, Bien, & Rösler, 2008). De Koning and Tabbers (2011) argue that this might be even possible for non-biological movements (i. e., mechanical or technical processes, e.g., lightning formation). To sum up, enabling learners to draw on their hMNS by showing them gestures (i.e., biologically primary knowledge) seems to be an effective instructional strategy to improve learning about biological movements (i.e., biologically secondary knowledge) from dynamic visualizations.

Compared to observing gestures made by someone else, learner-generated gestures have a more direct and stronger influence on the degree to which the hMNS is activated (e.g., Montgomery, Isenberg, & Haxby, 2007). Based on this notion, a way to further enhance learning from dynamic visualizations may be to ask learners to make gestures related to the movements depicted in a dynamic visualization themselves (cf. De Koning & Tabbers, 2011). This approach of self-performed gestures provides additional advantages in terms of the manner (e.g., amplitude, speed) in which the gestures are made and the possibility to draw on one's personal experiences (e.g., with fish movements) in order to perform the gestures. According to Memmert, Hagemann, Althoetmar, Geppert, and Seiler (2009), for example, exaggerating movements can support and even speed up understanding of movements or learning about motor skills. Additionally, because the learners embody the learning content in their own sensory and motor systems based on the physical movements they conduct (i.e., gestures), the information is coded in a distinct, visuospatial representational format that enriches the way the information is represented, thereby creating a higher-quality mental representation (Paas & Sweller, 2012). Higher-quality mental representations are associated with better learning (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), yielding faster and more accurate performance on learning outcome measures. Accordingly, Mierowsky et al. (2020) showed that producing mimicking gestures during learning how to play the piano with instructional dynamic visualizations fostered learning. In the domain of biological movements, Scheiter, Brucker, and Ainsworth (2020) recently demonstrated that enacting the to-be-learned movements had a positive effect on learning how to classify the movements as long as the movement pattern was neither too easy (the act of making gestures is not needed) nor too difficult (e.g., demanding, complex, or distracting) to be recognized (see also De Koning & Tabbers, 2013; Skulmowski, Bunge, Kaspar, & Pipa, 2014). Moreover, making gestures must not interfere with another instructional strategy that is simultaneously performed (such as for example making and also observing gestures; see Imhof, Scheiter, Edelmann, & Gerjets, 2013, for another example on competing instructional strategies during learning about fish movement patterns). Taken together, by focusing on self-performed gestures whilst learning about biological movements from dynamic visualizations, we move into a promising but yet rather unexplored field of research.

1.2. Learners' visuospatial ability, gestures, and learning

As the processing of continuous changes in dynamic visualizations as well as observing and making gestures requires visuospatial ability (cf. Hegarty, 1992), it is highly probable that learners' visuospatial ability will influence how much a learner benefits from dynamic visualizations as well as additional gestures (cf. Hegarty & Waller, 2005). Accordingly, previous research on learners' visuospatial ability (e.g., Höffler, 2010) shows that learners with higher visuospatial ability outperform learners with lower visuospatial ability during learning with visualizations.

Moreover, previous research reveals that visuospatial ability may moderate the effectiveness of learning with different instructions and visualization formats (e.g., dynamic versus static visualizations or corresponding versus non-corresponding gestures). There are two alternative interaction hypotheses: the ability-as-compensator versus the ability-as-enhancer hypothesis (e.g., Höffler, 2010).

With regard to the ability-as-compensator hypothesis, learners who are equipped with higher visuospatial ability may not require well-

designed visualizations and instructions as they manage to achieve high learning outcomes even if they only see suboptimal instructions or visualizations. However, learners who have lower visuospatial ability at their disposal may suffer from suboptimally designed visualizations and instructions (cf. ability-as-compensator hypothesis; Höffler, 2010; for research on dynamic visuospatial ability as a particular spatial skill see also Sanchez & Wiley, 2017). In accordance, certain visualization formats – for example dynamic visualizations – compensate the missing visuospatial ability and learners with lower visuospatial ability thus achieve the same learning outcomes with these visualizations as learners with higher visuospatial ability (e.g., Lee, 2007). For example, relating this to the Brucker et al. (2015) study, higher-visuospatial-ability learners likely possess the skills and resources to see when gestures are in conflict with the depicted content (non-corresponding gestures) and come up with their own strategy to elaborate on the relevant movements, whereas lower-visuospatial-ability learners do not possess these skills and therefore are less able to deal with situations where gestures are in conflict with the dynamic visualizations resulting in lower learning outcomes. However, well-designed visualizations (such as these supplemented with corresponding gestures) might compensate the missing visuospatial ability of lower-visuospatial-ability learners.

With regard to the ability-as-enhancer hypothesis, learners with a higher visuospatial ability might even be able to take advantage from specific visualization or instructional formats that are less optimally designed, such as dynamic visualizations with non-corresponding gestures as investigated in the Brucker et al. (2015) study. Assuming an enhancing effect of visuospatial ability in this case, would have led to a different result pattern. In such a case, learners with higher visuospatial ability would even have benefitted from observing the non-corresponding gestures that are in conflict with the depicted content. That is, the higher visuospatial ability might facilitate learning from such specific visualizations or instructional formats.

In sum, considering learners' visuospatial ability is relevant when studying the value of gestures in learning about movements from dynamic visualizations.

1.3. Present study

This study aimed to investigate to what extent learning about non-human biological movements from dynamic visualizations can be enhanced by adding information in the form of gestures. We implemented gesture-information during learning of fish movements from dynamic visualizations in two ways. Firstly, by asking learners to *observe* gestures displayed in a video and secondly, by asking participants to *make* gestures themselves during learning from the dynamic visualizations. For the observation of gestures participants studied the dynamic visualization whilst (1) observing corresponding gestures, (2) observing non-corresponding gestures, or (3) not observing additional gestures. When asked to make gestures, learners were (1) studying the dynamic visualizations whilst making gestures or (2) studying the visualizations without making gestures. Moreover, we addressed learners' visuospatial ability as a potential moderating variable. Furthermore, functional near-infrared spectroscopy (fNIRS), which is a non-intrusive neurophysiological method to gather information about cortical activation of humans (e.g., Ehlis, Schneider, Dressler, & Fallgatter, 2014), was used to investigate whether the hMNS is activated during observing and/or making gestures and learning about biological movements from dynamic visualizations (cf. Sun et al., 2018). Thus, the present study contributes to the field of research by combining (and hopefully disentangling) the two aspects of observing and making gestures for learners with different levels of visuospatial ability and by directly measuring cortical activation to uncover the respective processes underlying learning.

Based on research showing cognitive benefits of making gestures in instructional visualizations (e.g., Scheiter et al., 2020), we hypothesize that making gestures while studying the dynamic visualizations yields

higher learning outcomes and higher cortical activation (in regions associated with planning and performing the movements as well as the hMNS) than studying without making gestures (Hypothesis 1, see Table 1 for an overview of our hypotheses). Moreover, we hypothesized that studying the dynamic visualization with additionally observing gestures yields higher learning outcomes and higher cortical activation (in regions associated with the hMNS) than studying without observing gestures (due to the assumed additional hMNS activation, both types of gestures should lead to higher values than the no gesture condition); whereby – regarding the type of gesture – we expected corresponding gestures to result in higher learning outcomes and higher cortical activation (in regions associated with the hMNS) than non-corresponding gestures (Hypothesis 2). In accordance with Brucker et al. (2015) and as a more nuanced alternative, we hypothesize this pattern to vary as a function of gesture correspondence and level of learner's visuospatial ability: we expect that lower-visuospatial-ability learners show higher learning outcomes and higher cortical activation only with corresponding gestures (Hypothesis 3a), whereas higher-visuospatial-ability learners show improved learning outcomes and higher cortical activation for corresponding and non-corresponding gestures (Hypothesis 3b). Besides, we expected – in line with prior research on learners' visuospatial ability (e.g., Höffler, 2010) – higher-visuospatial-ability learners to outperform lower-visuospatial-ability learners (Hypothesis 4).

As we expected both observing and making gestures to be effective, one might argue that the combination of both might result in the highest learning outcomes and highest cortical activation. However, observing gestures might be particularly helpful if the learners were not allowed to gesture themselves, because observed gestures might interfere with self-performed gestures. This interference could for example be due to the fact that self-performed gestures do not fully overlap with the gestures observed in the video or the fact that learners make attempts to match the movements observed in the video when making gestures themselves (cf. De Koning & Tabbers, 2013). Moreover, it might be the case that if learners were already instructed to make gestures, they already activate their hMNS and thus observing gestures might not add any additional benefit. Thus, we expected the result patterns of the factors observing gestures and making gestures to be stronger in the conditions when the other factor was not present (Hypothesis 5). Finally, we hypothesized that higher hMNS activation is associated with higher learning outcomes. This is expected to be particularly true for lower-visuospatial-ability learners: if they show higher cortical activation in cortical regions associated with the hMNS they should also show higher learning outcomes (Hypothesis 6).

2. Methods

2.1. Participants and design

One hundred and twenty university students ($M = 24.33$ years, $SD = 3.98$; 86 females; 111 right-handed) from a German university were recruited via the online system ORSEE (http://www.orsee.org/). They were compensated for their participation with 10 Euro, and due to the chosen learning domain of fish movement patterns Biology majors were excluded from participation. The study protocol was approved by the local ethics committee and all participants gave consent and participated voluntarily.

Based on dynamic visualizations the participants had to learn how to discriminate different fish according to their swimming movements. During the learning phase the participants saw each fish movement pattern twice: Firstly, they saw an animation of the specific fish movement pattern. Secondly, they saw the animation of the specific fish movement pattern again. This time, depending on the experimental condition, the second presentation of the animation could have been complemented with up to two additional sources: participants could in addition have seen a video of a person performing gestures with his hands and arms (observing gestures) and/or they could in addition have received a written instruction to self-gesture (making gestures). So, there was a 2-by-3 between-subjects design with the two independent factors *observing gestures* and *making gestures*. The first factor, *observing gestures*, was varied in three variants: Participants either saw gestures that did correspond or that did not correspond (i.e., were unrelated) to the fish movement patterns or they saw no gesture at all. The second factor, *making gestures*, was varied in two variants: Participants either did or did not get the instruction "Please make your own gestures, that help you to better understand the movement." Combining these factors resulted in the six conditions displayed in Fig. 1.

For the observing gestures conditions, we used the same gestures that were already used in Brucker et al. (2015). For the corresponding gestures, an expert regarding fish movements displayed with his hands and arms representations of the respective movements as clearly as possible, whereas for the non-corresponding gestures the (same) expert performed gestures with his hands and arms that were unrelated to the fish movement patterns (i.e., waving, circulating the forearms around each other, drumming, and pointing; see Fig. 2).

2.2. Materials

Participants were asked to study dynamic visualizations to learn to classify four different fish movement patterns. These fish movement patterns differed in terms of the parts of the body that generate propulsion (i.e., several fins or the body itself) and also in the manner of how these body parts move in the three-dimensional space (i.e., different paddle-like or wave-like movements). The four different movement patterns were: 1. *balistiform*: undulation of the dorsal and anal fins; 2. *labriform*: oscillation of the pectoral fins; 3. *subcarangiform*: undulation of the body; and 4. *tetraodontiform*: oscillation of the dorsal and anal fins (and undulation of the pectoral fins). Identifying these movement patterns is very challenging given that fish may deploy other movements in addition (e.g., to navigate), which can easily be mistaken for movements used for propulsion in another movement pattern. The fish animations and gesture videos used in this study were taken from Brucker et al. (2015). The movement cycles of the movement patterns were presented in loops in the animations (30 s in total per movement pattern). The fish movement animations were depicted with 25 frames per second. They were presented at the upper left-hand side of the screen and their size was 480×360 pixels. The gestures were presented in the respective conditions in loops in the videos (30 s in total per movement pattern). The gesture videos were also depicted with 25 frames per second and in the same size as the animations (480×360 pixels) on the upper right-hand side of the screen. The presentation of all visualizations was

Table 1
Overview over the hypotheses of the present study.

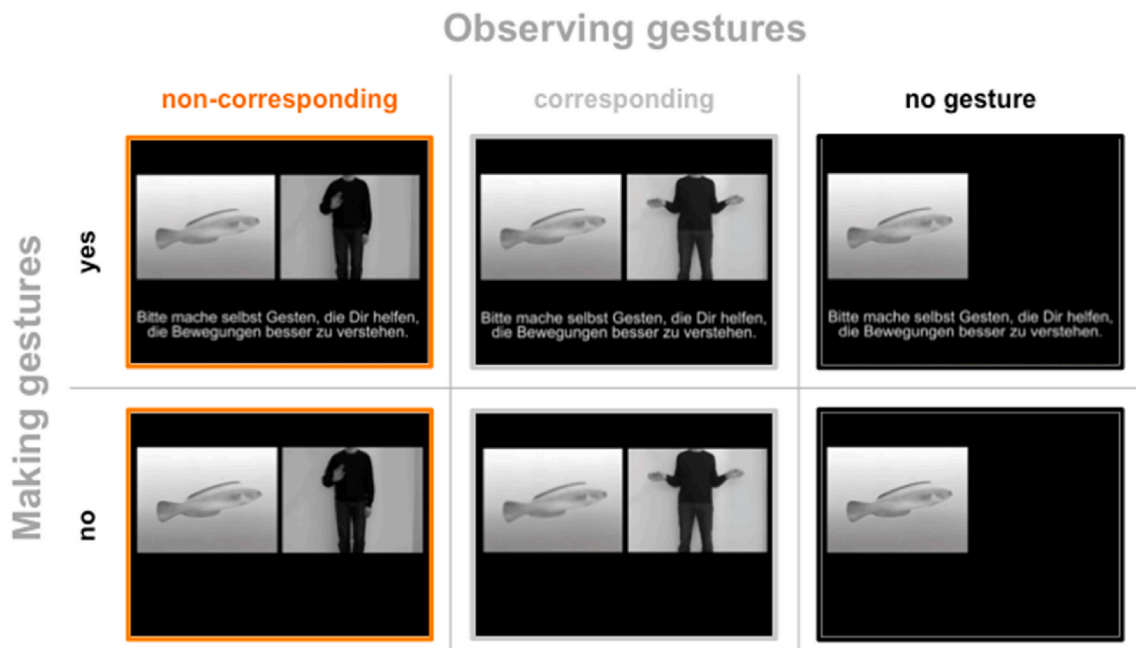|	Effects on learning outcomes and cortical activation
Hypothesis 1: main effect making gestures	yes > no
Hypothesis 2: main effect observing gestures	corresponding > non-corresponding > no gesture
Hypothesis 3: interaction observing gestures x learners' visuospatial ability	2a: low: corresponding > non-corresponding, no gesture 2b: high: corresponding, non-corresponding > no gesture
Hypothesis 4: main effect learners' visuospatial ability	higher > lower
Hypothesis 5: interaction observing gestures x making gestures	no gestures < observing gestures >/ = observing gestures and making gestures no gestures < making gestures >/ = observing gestures and making gestures
Hypothesis 6: moderating role of cortical activation on learning outcomes	Higher cortical activation (in areas associated with the hMNS) is associated with higher learning outcomes (particularly for lower-spatial-ability learners).

Observing gestures

Fig. 1. Screenshots of the six conditions in the 2-by-3-design of the study.

system-controlled. The self-gesturing instructions were presented centered at the bottom of the screen (see also Fig. 1). The gesture videos and the self-gesturing instructions were presented simultaneously with the fish visualization only during the second viewing of the movement pattern.

2.3. Measures

2.3.1. Demographics and familiarity with the domain

A questionnaire to assess participants' demographics and their *familiarity with the domain* was administered to gather demographical data (i.e., gender, age, major), to ensure that all participants were novices in the domain of this study (fish movement patterns), and to ensure that randomization was successful across the six conditions. The part of the questionnaire addressing *familiarity with the domain* was already used in former studies on fish movements (Brucker et al., 2015; Imhof et al., 2011, 2012). It was developed together with a domain expert, and was used to assess whether participants had encountered the domain of fish movement patterns prior to participating in the experiment by addressing details of participants' Biology school education; participants' familiarity with the marine world in terms of their experience with diving, snorkeling, swimming, rowing, and/or owning an aquarium; participants' interest in related topics, such as fish, biology, zoology, physics, aircraft construction, and/or shipbuilding; and participants' use of related media, including documentaries and/or books on fish or sea life and/or visits to an aquarium. The questionnaire consisted of 19 questions. For each answer indicating any familiarity with the domain, participants received one point. Higher numbers of diving and/or snorkeling experiences could each give up to two additional points. This resulted in a total amount of 0–23 points (= 19 + 2 + 2 points; for a detailed description see Brucker et al., 2015).

2.3.2. Learning outcomes

We administered a movement pattern classification test comprising 45 dynamic multiple-choice items to assess learning outcomes. These items consisted of underwater videos of real fish performing one of the four to-be-learned movement patterns or a distractor movement pattern. Thus, all items had five possible answer options (one for each of the four to-be-learned movement pattern and the additional answer "new

movement"). To correctly identify the movement pattern that was depicted in each item, learners had to identify the body parts relevant for propulsion and their way of moving. Each item was awarded one point for the correct answer (0–45 points). The classification test score of each participant was transformed into % correct.

2.3.3. Learners' visuospatial ability

To assess learners' visuospatial ability, we used a short version of the paper folding test (PFT, Ekstrom, French, Harman, & Dermen, 1976). Following Blazhenkova and Kozhevnikov (2009, p. 640) this test measures the ability to form representations of "object location, movement, spatial relationships, and transformations". Castro-Alonso and Atit (2019) point out that it is important to use appropriate visuospatial ability tests to investigate the effects on different visualization formats. They classify the PFT as a mental folding test in contrast to mental rotations tests (such as the mental rotation test [MRT] from Vandenberg & Kuse, 1978) and state that mental folding is particularly of interest for changing – in contrast to rigid – objects (such as bending fins of fish). Thus, the visuospatial ability task used in the present study (i.e., the PFT) is well suited for learning about fish movements, during which participants must identify different movements of the fins and the body of the fish in terms of bendings and turn downs as well as to locate distinct features (e.g., a certain fin ray) in different positions. The short version of the PFT consists of ten multiple-choice items. For each item, participants see five options from which they must choose the correct answer. The stimuli are depictions of papers that are folded stepwise and then were punched in the folded state. The answer options depict unfolded papers with punches being either in the correct or incorrect positions. Each correct answer was awarded one point (0 to max. 10 points) and participants had a maximum of 3 min to work on this task.

2.3.4. Cortical activation (fNIRS measurement and analyses)

During viewing the animations of the fish movements for the second time in the learning phase (i.e., the phase in which the experimental manipulation took place), cortical activation was assessed via fNIRS measurements with an ETG-4000 (Hitachi Medical Co.). To cover left and right lateral surfaces of the head, we used a 2 × 22 channel array as probe set. The probe set was placed over fronto-temporo-parietal regions and was centered by positioning the middle lower channel at the T3-T4
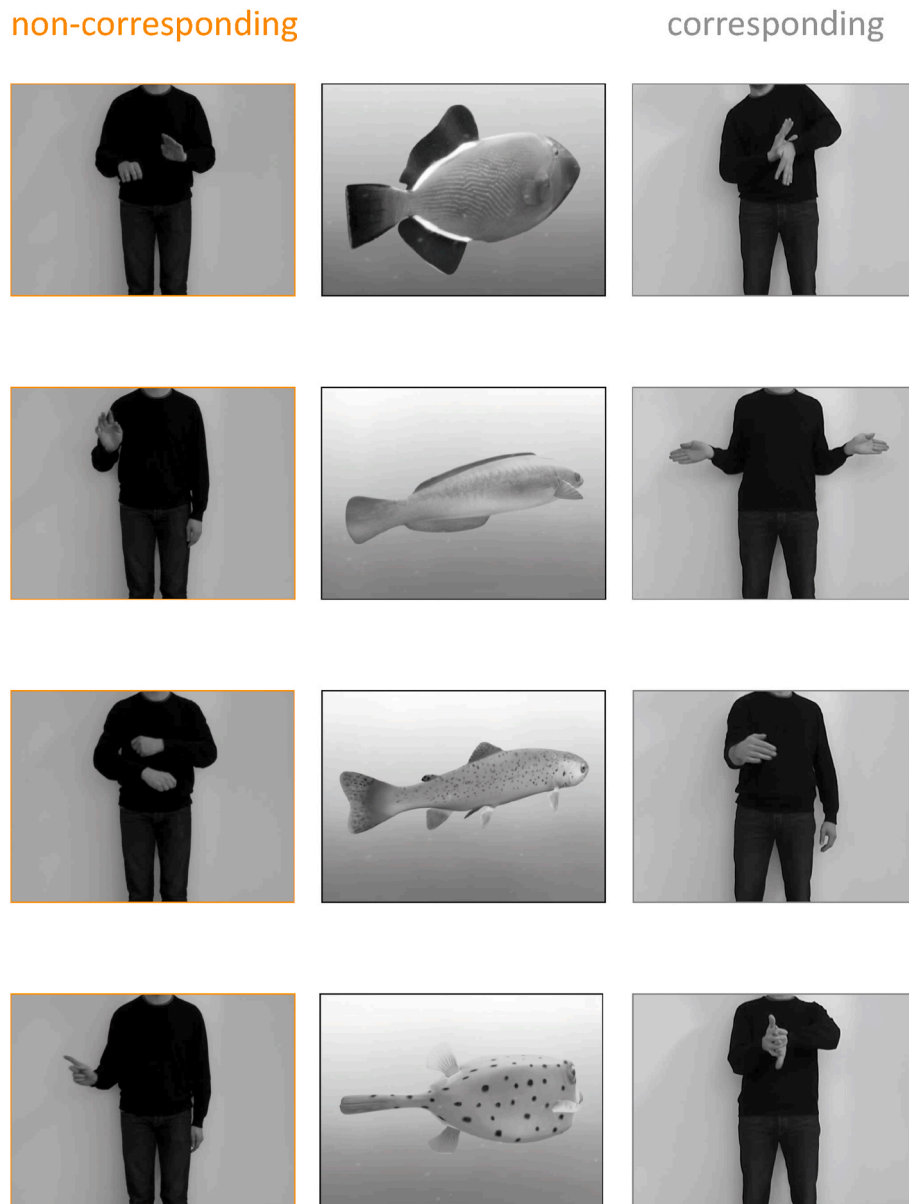
Fig. 2. Screenshots of non-corresponding (left) and corresponding (right) gestures for each of the four fish movement patterns (top down: *balisti-, labri-, subcarangi-, tetraodontiform*).

position and orienting the probe set upwards in a straight line towards the C3–C4 position (not exactly terminating on these positions because of the fixed interoptode distances) according to the standard locations of the international 10–20 system for electrode placement (Jasper, 1958). Consequently, and in line with previous publications (e.g., Okamoto et al., 2004), the measurement array was thus covering the motor cortex as well as adjacent inferior frontal and temporo-parietal brain areas. The fNIRS system measures the change in the product of hemoglobin (Hb) concentration and effective optical path length in human brain tissue. The unit of Hb change is molar concentration (mM = mmol/l) multiplied by optical path length (mm). Local increases of oxygenated Hb and decreases of deoxygenated Hb are indicators of cortical activity (Obrig & Villringer, 2003). For data analyses using customized MATLAB (Math-Works Inc, Natick, USA) scripts, hemodynamic raw data were initially lowpass filtered at 0.1 Hz and highpass filtered at 0.005 Hz. After that, the original data series of oxygenated and deoxygenated Hb were combined to one "true oxy signal" by using the algorithm proposed by Cui, Bray, Bryant, Glover, and Reiss (2010) for reduction of motor artefacts. After a visual inspection of the resulting data and manual

interpolation of single noisy channels, we performed an independent component analysis (ICA) to identify and remove clenching artefacts related to temporal muscle activation (cf. Schecklmann et al., 2017). Then, a second bandpass filtering was performed (at 0.01 and 0.1 Hz, respectively) before a global signal reduction was done with a spatial Gaussian kernel filter (see Zhang et al., 2016). Following these pre-processing steps, the 30 s task segments were averaged over the four task repetitions (i.e., the four different to-be-learned fish movements) for the second viewing of the fish movements (plus additional gesture video and/or self-gesturing instruction depending on the experimental condition). Finally, for statistical analyses, mean values of these 30 s activation segments were calculated and exported for each NIRS channel and participant.

To analyze the cortical activation, we defined four cortical regions of interest (ROIs) on the left hemisphere (see Fig. 3): one for the left motor cortex (MC; involved in planning, controlling, and executing movements, cf. Sanes & Donoghue, 2000), one for the left dorsolateral prefrontal cortex (DLPFC; involved in executive functions and motor planning, organization, and regulation, cf. Miller & Cohen, 2001;
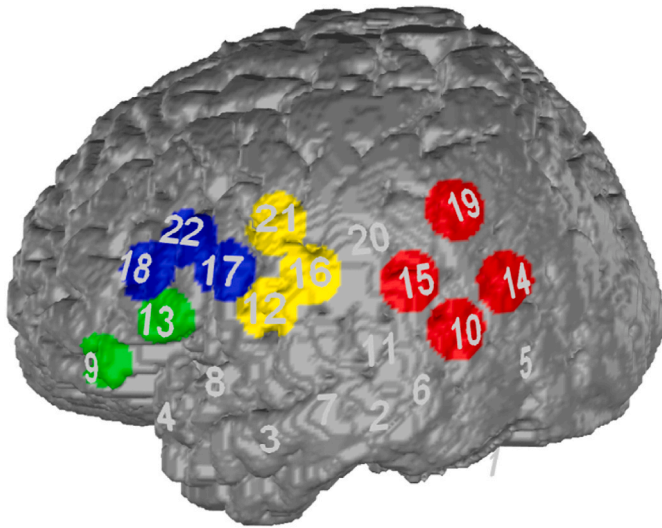
Fig. 3. Spatial arrangement of the left probe set with the four ROIs (MC = 12, 16, 21 in yellow, DLPFC = 17, 18, 22 in blue, IFC = 9, 13 in green, and IPC = 10, 14, 15, 19 in red).

Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004), and two for the hMNS among the respective channels. The two ROIs for the hMNS were the left inferior frontal cortex (IFC; involved in language processing and associated with the hMNS; cf. Rizzolatti & Craighero, 2004) and the left inferior parietal cortex (IPC; involved in the perception of emotions in facial stimuli and also associated with the hMNS; cf. Rizzolatti & Craighero, 2004).

2.4. Procedure

Participants were tested individually. Firstly, they had to read a printed overview with information about the procedure of the study and fill in the consent form. Then, they had to answer the questionnaire on demographics and participants' familiarity with the domain, and the PFT. After answering these questionnaires, the experimenter placed and adjusted the fNIRS probe set on the scalp of the participants. Subsequently, the computer-based learning materials were presented in the learning phase. For each of the four to-be-learned movement patterns, learners were always presented with two presentations of each of the fish animations: presentation 1 showed the fish animation, and presentation 2 showed the same fish animation plus an additional gesture video and/or self-gesturing instruction depending on the experimental condition. Participants saw the animation of the first fish movement for 30 s. Then a pause of 30 s (black screen) followed before they saw the animation of the first fish movement with its additions (depending on the experimental condition; fNIRS measurement took place) for 30 s again. Then again, a pause of 30 s (black screen) followed before the presentation of the next fish movement started. Following the learning phase – that lasted 8 min – learners performed a filler task for about 8 min in which they answered some questions on object positions of depicted objects. Finally, learners completed the movement classification test which took 15 min. In the testing phase, each item was visible for 7 s and immediately afterwards participants had 3 s time to choose the correct answer by pressing a corresponding key on the keyboard. Participants were instructed to put both their forefingers and both their middle fingers on predefined keys as well as one of their thumbs on the space bar to answer the test items. The predefined keys were labeled on the screen with static screenshots from the learning animations of the four movement patterns and the spacebar was labeled with a grey bar indicating movements that were not part of the learning phase (i.e., distractor items). In total, one experimental session lasted approximately 50 min.

2.5. Statistical analysis

The data were analyzed with Matlab (version 9.2.0, The MathWorks, Inc. Released 2017) and SPSS® (version 25, IBM Corp. Released 2017). We used a *p*-value of .05 as an indicator of significance and a *p*-value of .10 as an indicator of a tendency in all analyses.

3. Results

3.1. Learner prerequisites

To test for the equivalence of the six experimental groups, we compared them with regard to their learner prerequisites: familiarity with the domain, visuospatial ability, age, and gender. We conducted ANOVAs (univariate analysis of variance) with *condition* as between-subjects factor on the scores for familiarity with the domain, learners' visuospatial ability, and age. A chi-squared test was conducted for examining the distribution of gender over all conditions. Means and standard deviations are reported in Table 2.

Regarding familiarity with the domain there was no effect for condition ($F < 1$, ns). In general, the means indicated that learners' familiarity with the domain was very low in all experimental conditions, revealing that the selection of low domain knowledge participants was successful. Regarding visuospatial ability, there was also no effect for condition ($F < 1$, ns). Regarding age, there was also no effect for condition, $F(5, 114) = 1.597$, $MSE = 15.433$, $p = .166$, $\eta^2_p = .065$, ns). Furthermore, there were no significant associations between the six experimental conditions and participants' gender, $\chi^2(5) = 2.791$, $p = .732$, ns. Taken together, the six experimental conditions are comparable with regard to learners' prerequisites in terms of familiarity with the domain, visuospatial ability, age, and gender.

3.2. Learning outcomes

To analyze learning outcomes, we conducted an ANCOVA (univariate analysis of covariance) on the classification performance with the between-subjects factors *observing gestures* and *making gestures,* and the continuous variable *learners' visuospatial ability* as a third factor. To test the moderating role of learners' visuospatial ability, z-standardized values of the PFT were used and all possible two-way and three-way interactions between the three factors were inserted as interaction terms in the ANCOVA model (for adjusted means and standard errors, see Table 3).

Results showed no significant main effect for making gestures ($F < 1$, ns) and observing gestures, $F(2,108) = 1.281$, $MSE = 121.764$, $p = .282$, $\eta^2_p = .023$. However, there was a significant main effect for learners' visuospatial ability, $F(1,108) = 11.969$, $MSE = 121.764$, $p = .001$, $\eta^2_p = .100$. This effect has to be interpreted in terms of the significant interaction between observing gestures and learners' visuospatial ability, $F(2,108) = 7.417$, $MSE = 121.764$, $p = .001$, $\eta^2_p = .121$; see Fig. 4). This interaction showed that for participants with higher visuospatial ability (defined as one standard deviation above the sample mean) the non-corresponding gesture led to a higher classification performance than the corresponding gesture ($p = .005$) and tended to outperform also learning without gesture ($p = .083$), whereas there was no difference between the corresponding gesture and learning without gesture ($p > .999$). For participants with lower visuospatial ability (defined as one standard deviation below the sample mean) non-corresponding gestures led to a lower classification performance than when learning without gesture ($p = .015$). There was no significant difference between the corresponding gesture condition and the no gesture condition ($p = .552$) or the corresponding gesture condition and the non-corresponding gesture condition ($p = .415$) for lower-visuospatial-ability learners. Thus, the non-corresponding gestures are beneficial for higher-visuospatial-ability learners, but detrimental for lower-visuospatial-ability learners. In other words, learners who have higher visuospatial

Table 2
Means and standard deviations (in parentheses) for values of familiarity with the domain, of visuospatial ability, of classification performance, and number of females/males as a function of the between-subjects factors "observing gestures" and "making gestures".

	observing corresponding gestures		observing non-corresponding gestures		observing no gesture	
	making gestures yes (n = 20)	making gestures no (n = 20)	making gestures yes (n = 20)	making gestures no (n = 20)	making gestures yes (n = 20)	making gestures no (n = 20)
Domain familiarity (0–23)	4.700 (2.638)	4.263 (3.364)[a]	5.350 (2.907)	4.500 (2.666)	4.250 (2.751)	4.650 (2.739)
Visuospatial ability (0–10)	7.800 (1.673)	6.850 (2.346)	7.250 (2.099)	7.700 (1.593)	7.050 (1.959)	7.250 (1.803)
Age (years)	24.100 (2.954)	25.550 (4.442)	23.150 (3.543)	23.500 (4.274)	25.850 (3.787)	23.850 (4.356)
Females	13	15	13	16	13	16
Males	7	5	7	4	7	4

[a] Missing values for one participant.

Table 3
Means and standard errors (in parentheses) for values of classification performance (in % correct) as a function of the between-subjects factors "observing gestures", "making gestures" and "visuospatial ability".

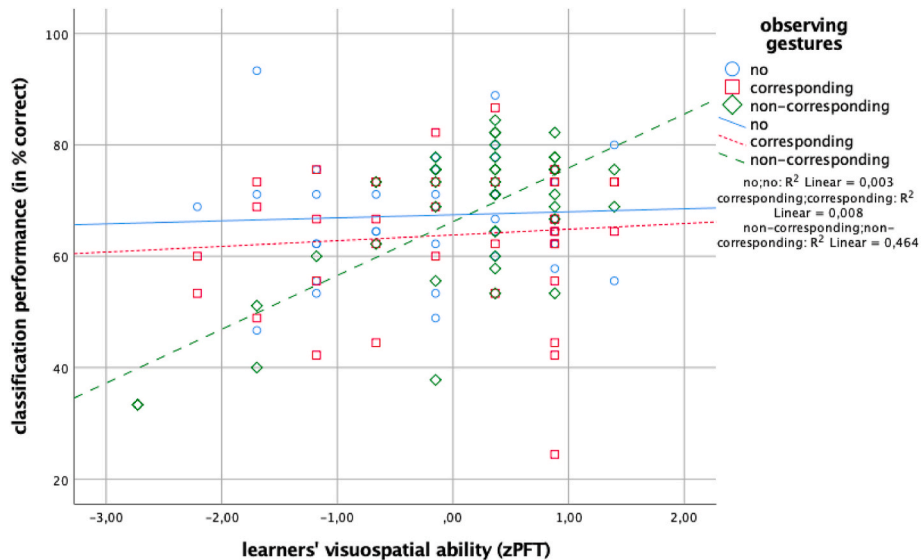|	observing corresponding gestures				observing non-corresponding gestures				observing no gesture			
	making gestures yes (n = 19)		making gestures no (n = 19)		making gestures yes (n = 20)		making gestures no (n = 19)		making gestures yes (n = 18)		making gestures no (n = 19)	
	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)
Classification performance (in % correct)	66.862 (3.287)	62.182 (4.392)	62.528 (3.566)	62.573 (2.923)	74.159 (3.438)	56.241 (3.328)	77.487 (3.470)	57.108 (4.413)	66.822 (3.749)	63.818 (3.262)	68.756 (3.723)	70.516 (3.585)

VSA = visuospatial ability

Fig. 4. Interaction between observing gestures and learners' visuospatial ability on classification performance.

ability at their disposal achieve better learning results with the non-corresponding gestures, whereas learners who have lower visuospatial ability as a (missing) prerequisite obtain poorer learning results with these non-corresponding gestures. There were no other significant two-way or three-way-interactions regarding learning outcomes (all $Fs < 1$).

3.3. Cortical activation

The cortical activation in the four ROIs on MC, DLPFC, IFC, and IPC was analyzed with four ANCOVAs (one for each ROI) with the between-subjects factors *observing gestures* and *making gestures*, and the continuous variable *learners' visuospatial ability* as a third factor (z-standardized; all possible interaction terms included in the ANCOVA model; for adjusted means and standard errors, see Table 4).

We had to exclude six participants from the analyses on cortical activation because of poor data quality resulting in a total number of 114 participants in these analyses. One participant had to be excluded from each of the four conditions "observing corresponding gestures + making

Table 4

Means and standard deviations (in parentheses) for cortical activation in all ROIs as a function of the between-subjects factors "observing gestures", "making gestures", and "visuospatial ability".

	observing corresponding gestures				observing non-corresponding gestures				observing no gesture			
	making gestures yes (n = 19)		making gestures no (n = 19)		making gestures yes (n = 20)		making gestures no (n = 19)		making gestures yes (n = 18)		making gestures no (n = 19)	
	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)	higher VSA (M + 1 SD)	lower VSA (M - 1 SD)
Motor Cortex (MC)	0.023 (0.007)	0.004 (0.009)	- 0.007 (0.007)	0.000 (0.006)	0.002 (0.007)	0.003 (0.007)	- 0.010 (0.007)	- 0.008 (0.009)	0.011 (0.008)	0.005 (0.007)	0.013 (0.008)	0.008 (0.007)
Dorsolateral Prefrontal Cortex (DLPFC)	0.015 (0.007)	0.002 (0.010)	- 0.015 (0.008)	0.004 (0.006)	−0.005 (0.007)	0.007 (0.007)	- 0.006 (0.008)	- 0.002 (0.009)	0.015 (0.008)	0.000 (0.007)	0.004 (0.008)	0.002 (0.008)
Inferior Frontal Cortex (IFC)	0.043 (0.009)	- 0.001 (0.013)	- 0.008 (0.010)	- 0.011 (0.008)	- 0.014 (0.010)	0.002 (0.009)	- 0.003 (0.010)	- 0.004 (0.012)	0.013 (0.011)	0.005 (0.009)	0.009 (0.011)	0.007 (0.010)
Inferior Parietal Cortex (IPC)	- 0.006 (0.007)	- 0.001 (0.010)	−0.014 (0.008)	0.003 (0.007)	0.002 (0.008)	0.013 (0.007)	- 0.019 (0.008)	- 0.004 (0.010)	0.001 (0.009)	−0.002 (0.008)	0.009 (0.009)	−0.002 (0.008)

VSA = visuospatial ability

gestures", "observing corresponding gestures + not making gestures", "not observing gestures + not making gestures", "observing non-corresponding gestures + not making gestures" and two participants had to be excluded from the condition "not observing gestures + making gestures" resulting in 19 participants in the conditions "observing corresponding gestures + making gestures", "observing corresponding gestures + not making gestures", "not observing gestures + not making gestures", "observing non-corresponding gestures + not making gestures", 18 participants in the condition "not observing gestures + making gestures" and unchanged 20 participants in the condition "observing non-corresponding gestures + making gestures".

For observing gestures there was a significant main effect on MC activation, $F(2,102) = 3.328$, $MSE = 0.00048$, $p = .040$, $\eta^2_p = .061$. Pairwise Bonferroni-adjusted comparisons revealed that participants observing non-corresponding gestures showed significantly lower MC activation than participants observing no gesture ($p = .040$). There were no significant differences in MC activation between observing corresponding gestures and observing non-corresponding gestures ($p = .275$) or not observing gestures ($p > .999$).

Even though making gestures did not influence results on learning outcomes, analyses on cortical activation showed that for making gestures there was a significant main effect on MC activation, $F(1,102) = 4.138$, $MSE = 0.001$, $p = .045$, $\eta^2_p = .039$, and a tendency on DLPFC activation, $F(1,102) = 3.106$, $MSE = 0.001$, $p = .081$, $\eta^2_p = .030$. The activation for both MC and DLPFC was higher for participants who self-gestured than for participants who did not self-gesture.

Moreover, analyses on cortical activation showed a tendency for an interaction between observing gestures and making gestures for IFC activation, $F(2,102) = 3.004$, $MSE = 0.001$, $p = .054$, $\eta^2_p = .056$. Pairwise Bonferroni-adjusted comparisons revealed that participants observing corresponding gestures showed higher IFC activation if they self-gestured than when they did not self-gesture ($p = .005$), whereas there was no such difference for participants who observed non-corresponding gestures ($p = .830$) and participants who did not observe gestures ($p = .923$). There were no other significant main effects or interactions in the four analyses (MC, DLPFC, IFC, IPC) on cortical activation (for statistical values see Table 5).

3.3.1. Effects of cortical activation on learning

To address the question whether higher cortical activation directly effects learning outcomes (i.e., classification test score), we conducted four separate ANCOVAs with the between-subjects factors *observing gestures* and *making gestures,* and *learners' visuospatial ability* and *cortical activation* as continuous variables in terms of *IFC activation, IPC*

activation, MC activation, and DLPFC activation respectively (all continuous variables were z-standardized; all possible interaction terms were included in the ANCOVA model; for adjusted means and standard errors, see Tables 6–9 respectively for the different cortical activations).

In this section only the main effects and interactions with regard to the respective cortical activation are reported, because the effects of making gestures, observing gestures, learners' visuospatial ability, and their potential interactions have already been reported in the section on learning outcomes (see 3.2. Learning outcomes).

There was a tendency that higher IFC activation predicted higher learning outcomes, $F(1,90) = 3.372$, $MSE = 126.639$, $p = .070$, $\eta^2_p = .036$. There were no other significant interactions with IFC activation (all $Fs < 1$, ns).

For MC activation as well as IPC activation there were no significant main effects or interactions (all $Fs < 1$, ns).

The analysis on DLPFC activation showed a tendency for a main effect of DLPFC activation on learning outcomes, $F(1, 90) = 3.430$, $MSE = 116.697$, $p = .067$, $\eta^2_p = .037$; a tendency for an interaction between observing gestures and DLPFC activation, $F(2,90) = 2.387$, $MSE = 116.679$, $p = .098$, $\eta^2_p = .050$; a significant interaction between making gestures and DLPFC activation, $F(1,90) = 4.859$, $MSE = 116.679$, $p = .030$, $\eta^2_p = .051$; a tendency for an interaction between visuospatial ability and DLPFC activation, $F(1,90) = 3.898$, $MSE = 116.679$, $p = .051$, $\eta^2_p = .042$; a significant three-way interaction between observing gestures x making gestures x DLPFC activation, $F(2,90) = 3.308$, $MSE = 116.679$, $p = .041$, $\eta^2_p = .068$; and a significant three-way interaction between making gestures x visuospatial ability x DLPFC activation, $F(1,90) = 6.054$, $MSE = 116.697$, $p = .016$, $\eta^2_p = .063$. For the sake of completeness: The three-way interaction observing gestures x visuospatial ability x DLPFC activation was not significant ($F(2,90) = 1.872$, $MSE = 116.697$, $p = .160$, $\eta^2_p = .040$, ns). All of the aforementioned significant or marginal significant effects have to be interpreted in the light of a tendency for a significant four-way-interaction between observing gestures x making gestures x visuospatial ability x DLPFC activation, $F(2,90) = 2.545$, $MSE = 116.697$, $p = .084$, $\eta^2_p = .054$ (see Fig. 5).

The four-way-interaction indicated no significant differences between making or not making gestures for participants with higher visuospatial ability and higher DLPFC activation (no gestures: $p = .937$, corresponding gestures: $p = .925$, non-corresponding gestures: $p = .301$).

For participants with higher visuospatial ability and lower DLPFC activation, making gestures tended to be beneficial in the corresponding conditions (compared to not making gestures, $p = .085$), whereas

Table 5

Statistical values for the four ANCOVAs on cortical activation (MC, DLPFC, IFC, IPC) with the between-subjects factors "observing gestures", "making gestures", and "visuospatial ability".

	ROIs	F-value	MSE-value	p-value	η^2_p-value
Main effect	**MC**	**F(2,102) = 3.328**	**MSE = .000**	**p = .040**	**η^2_p = .061**
observing gestures	DLPFC	F < 1, ns			
	IFC	F(2,102) = 1.933	MSE = .001	p = .150, ns	η^2_p = .037
	IPC	F < 1, ns			
Main effect	**MC**	**F(1,102) = 4.138**	**MSE = .000**	**p = .045**	**η^2_p = .039**
making gestures	**DLPFC**	**F(1,102) = 3.106**	**MSE = .001**	**p = .081**	**η^2_p = .030**
	IFC	F(1,102) = 2.700	MSE = .001	p = .103, ns	η^2_p = .026
	IPC	F(1,102) = 1.413	MSE = .001	p = .237, ns	η^2_p = .014
Main effect	MC	F < 1, ns			
visuospatial ability	DLPFC	F < 1, ns			
	IFC	F(1,102) = 1.357	MSE = .001	p = .247, ns	η^2_p = .013
	IPC	F(1,102) = 1.427	MSE = .001	p = .235, ns	η^2_p = .014
Interaction observing gestures x	MC	F(2,102) = 1.871	MSE = .000	p = .159, ns	η^2_p = .035
making gestures	DLPFC	F < 1, ns			
	IFC	**F(2,102) = 3.004**	**MSE = .001**	**p = .054**	**η^2_p = .056**
	IPC	F(2,102) = 2.247	MSE = .001	p = .111, ns	η^2_p = .042
Interaction	MC	F < 1, ns			
observing gestures x	DLPFC	F(2,102) = 1.117	MSE = .001	p = .331, ns	η^2_p = .021
visuospatial ability	IFC	F(2,102) = 2.191	MSE = .001	p = .117, ns	η^2_p = .041
	IPC	F(2,102) = 1.767	MSE = .001	p = .176, ns	η^2_p = .033
Interaction	MC	F(1,102) = 1.251	MSE = .000	p = .266, ns	η^2_p = .012
making gestures x	DLPFC	F(1,102) = 1.853	MSE = .001	p = .176, ns	η^2_p = .018
visuospatial ability	IFC	F < 1, ns			
	IPC	F < 1, ns			
Interaction	MC	F < 1, ns			
observing gestures x	DLPFC	F(2,102) = 1.544	MSE = .001	p = .218, ns	η^2_p = .029
making gestures x	IFC	F(2,102) = 1.955	MSE = .001	p = .147, ns	η^2_p = .037
visuospatial ability	IPC	F < 1, ns			

ROIs = regions of interest
MC = motor cortex
DLPFC = dorsolateral prefrontal cortex
IFC = inferior frontal cortex
IPC = inferior parietal cortex

making gestures tended to be detrimental in the non-corresponding conditions (compared to not making gestures, $p = .053$). There was no difference for these participants between making and not making gestures in the conditions in which they did not observe gestures ($p = .381$).

For participants with lower visuospatial ability and higher DLPFC activation, making gestures was detrimental in the non-corresponding conditions (compared to not making gestures, $p = .025$). There were no significant differences for these participants between making and not making gestures in the corresponding conditions ($p = .192$) and in the no gesture conditions ($p = .629$).

For participants with lower visuospatial ability and lower DLPFC activation, making gestures was beneficial in the non-corresponding conditions ($p = .016$). There were no differences for these participants between making and not making gestures in the corresponding conditions ($p = .273$) and in the no gesture conditions ($p = .232$).

Thus, in essence the four-way interaction indicates that for different learners (higher versus lower visuospatial ability) self-gesturing could either be beneficial or detrimental depending on the combinations of which gesture (corresponding or non-corresponding) participants observed and whether they showed higher or lower DLPFC activation. It should be noted, that self-gesturing showed neither positive nor negative effects in the conditions in which participants did not observe gestures. In sum, self-gesturing is beneficial for higher-visuospatial-ability learners, who showed lower DLPFC activation and observe corresponding gestures as well as for lower-visuospatial-ability learners, who showed lower DLPFC activation, but observe non-corresponding gestures. In contrast, self-gesturing is detrimental for higher-visuospatial-ability learners, who showed lower DLPFC activation and observe non-corresponding gestures as well as for lower-visuospatial-ability learners, who showed higher DLPFC activation and observe non-corresponding gestures.

4. Discussion

This study investigated whether observing and making gestures improves learning about non-human biological movements from dynamic visualizations and to what extent this is related with the cortical activation in areas associated with the hMNS.

4.1. Effects of observing corresponding and non-corresponding gestures

Regarding learning outcomes, our results indicate that there were neither significant differences between observing corresponding gestures and observing no gestures for higher-as well as for lower-visuospatial-ability learners. However, the observation of gestures has differential effects for higher- and lower-visuospatial-ability learners when it comes to observing non-corresponding gestures. This is particularly in line with the more nuanced assumption that the effectiveness of observing gestures depends on learners' visuospatial ability (Hypothesis 3). For higher-visuospatial-ability learners, non-corresponding gestures improved learning (even beyond corresponding gestures, which is in line with Hypothesis 3b), whereas for lower-visuospatial-ability learners the observation of non-corresponding gestures had detrimental effects on learning (which is in line with Hypothesis 3a). These findings are largely in line with the pattern of results reported by Brucker et al. (2015).

One possible explanation for these results is that particularly when higher-visuospatial-ability learners are challenged by a desirable difficulty (cf. Schüler, 2017; Yue, Castel, & Bjork, 2013), in this case by creating a conflict between the visualized fish movements and the (mismatching) non-corresponding gestures, they might be stimulated to intensify their effort in order to reduce the conflict detected (cf. Scharinger, Schüler, & Gerjets, 2020), for instance by coming up with a strategy to more elaborately process the relevant movements. This in

Table 6

Means and standard errors (in parentheses) for values of classification performance (in % correct) as a function of the between-subjects factors "observing gestures", "making gestures", "visuospatial ability", and "IFC activation".

	observing corresponding gestures								observing non-corresponding gestures								observing no gesture							
	making gestures yes (n = 19)				making gestures no (n = 19)				making gestures yes (n = 20)				making gestures no (n = 19)				making gestures yes (n = 18)				making gestures no (n = 19)			
	higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)	
	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)	higher IFC (M + 1 SD)	lower IFC (M - 1 SD)
Classification performance (in % correct)	67.005 (3.379)	68.952 (5.830)	60.013 (8.731)	60.991 (8.528)	71.405 (6.376)	56.918 (4.333)	66.294 (6.610)	59.518 (3.649)	76.668 (5.857)	73.105 (4.091)	56.257 (5.056)	56.802 (5.868)	79.258 (9.433)	75.478 (5.872)	73.339 (21.919)	48.195 (12.223)	68.892 (5.802)	61.328 (8.095)	69.726 (7.578)	56.371 (7.500)	67.948 (7.114)	67.478 (8.059)	73.105 (7.287)	67.104 (7.530)

VSA = visuospatial ability
IFC = inferior frontal cortex activation

turn might increase the chance of properly understanding the depicted movements.

At first sight, one might also assume an alternative explanation for the overall pattern of results, which is based on the fact that the videos of the gestures and the visualizations of the to-be-learned fish movements were presented simultaneously to the participants, potentially evoking a split attention effect in the observation phase (e.g., Ayres & Sweller, 2014). For instance, this split might be used to explain the missing main effect for observing gestures, and particularly the missing beneficial effects of corresponding gestures that were expected in Hypothesis 2 due to an assumed activation of the hMNS. However, the split-attention effect with regard to the two visualizations might have imposed additional cognitive load onto the learners due to the need to view and process two visualizations simultaneously and thus to split the attention between them. Accordingly, in the corresponding gestures condition, in which learners might have realized that the gesture videos were in principle helpful but also a bit redundant to the videos of the fish movements themselves, they might have stopped viewing them and concentrated on the visualization of the fish movements instead to avoid split-attention effects and even more negative redundancy effects (see Kalyuga & Sweller, 2014). This assumption of an ignorance-reaction with regard to gestures could on the one hand explain a lack of beneficial effects for the corresponding gestures but it would on the other hand require additional empirical evidence to be substantially supported, for instance by means of collecting eye-tracking data (e.g., Van Gog & Scheiter, 2010) and comparing the amount of attention devoted to both types of visualizations.

When it comes to the effects of non-corresponding gestures for higher- and for lower-visuospatial-ability learners one might be tempted to hypothesize that learners with higher visuospatial ability might have better realized than learners with lower visuospatial ability that the non-corresponding gestures were irrelevant and that they might therefore have immediately stopped viewing them and concentrated on the visualization of the fish movements instead. Lower-visuospatial-ability learners on the contrary might have found it more difficult to realize the irrelevance of the non-corresponding gestures and might have invested time to understand the relation between these gestures and the fish movements, thereby experiencing detrimental effects of the non-corresponding gestures on learning outcomes.

However, the fact that for higher-visuospatial-ability learners non-corresponding gestures at least tended to outperform the condition with no gestures at all, in which no split attention effect could have occurred, speaks rather in favor of a desirable difficulty effect than in favor of an avoided split attention effect. If higher-visuospatial-ability learners would have tried to ignore the non-corresponding information in order to avoid splitting their attention, because they realized the mismatch, we would have expected identical or even worse learning outcomes in the non-corresponding gesture condition than in the no gesture condition, for instance due to the necessity ot inhibit distraction by the gesture videos. Our results, however, show on the contrary that at least at the descriptive level (and with a statistical tendency) non-corresponding gestures yielded better learning outcomes than in the no gesture condition. Particularly, if one has a closer look at participants with higher visuospatial ability in the two conditions (a) fish movement visualization without any further gesturing instruction (only fish movement visualization, no gesture video, no self-gesturing) versus (b) fish movement visualization with non-corresponding gesture, but without self-gesturing, the split attention explanation would clearly claim that the condition with only one information source (i.e., the fish movement visualization alone) should outperform the condition in which the two competing information sources were available at the same time (i.e., the fish movement visualization plus the non-corresponding gesture). However, the results indicate the reversed pattern, that the combination of fish movement visualization plus non-corresponding gestures led to better results for these higher-visuospatial-ability learners (77% correct, cf. Table 3) than the fish

Table 7

Means and standard errors (in parentheses) for values of classification performance (in % correct) as a function of the between-subjects factors "observing gestures", "making gestures", "visuospatial ability", and "IPC activation".

	observing corresponding gestures								observing non-corresponding gestures								observing no gesture							
	making gestures yes (n = 19)				making gestures no (n = 19)				making gestures yes (n = 20)				making gestures no (n = 19)				making gestures yes (n = 18)				making gestures no (n = 19)			
	higher VSA (M - 1 SD)		lower VSA (M + 1 SD)		higher VSA (M - 1 SD)		lower VSA (M + 1 SD)		higher VSA (M - 1 SD)		lower VSA (M + 1 SD)		higher VSA (M - 1 SD)		lower VSA (M + 1 SD)		higher VSA (M - 1 SD)		lower VSA (M + 1 SD)		higher VSA (M - 1 SD)		lower VSA (M + 1 SD)	
	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)	higher IPC (M - 1 SD)	lower IPC (M + 1 SD)
Classification performance (in % correct)	63.887 (6.358)	67.630 (4.294)	65.010 (8.115)	59.336 (5.824)	62.250 (6.971)	61.687 (4.454)	63.931 (4.114)	56.093 (5.998)	74.526 (4.270)	70.819 (6.815)	59.253 (5.767)	57.002 (3.815)	77.723 (7.690)	77.713 (4.886)	59.910 (11.272)	54.386 (11.530)	67.545 (7.735)	64.040 (9.554)	60.381 (4.676)	66.141 (4.595)	68.982 (4.949)	65.093 (6.183)	72.486 (8.200)	68.076 (6.772)

VSA = visuospatial ability

IPC = inferior parietal cortex activation

visualization alone (69% correct, cf. Table 3). Thus, we still presume, that the non-corresponding gestures might provide advantages for higher-visuospatial-ability learners in terms of some kind of desirable difficulty, because these learners seem to somehow make value out of the conflicting information simultaneously provided to them in this study.

In contrast, lower-visuospatial-ability learners presumably are insufficiently equipped for coping with such a situation of two conflicting information sources – independently of whether these information sources might in principle provide desirable difficulties or whether they are just potentially interfering. The lower-visuospatial abiltiy learners might not only lack the necessary resources to quickly identify the mismatch and then ignore the gestures, but also the strategies required to cope with a detected mismatch in terms of a more effortful and elaborated processing of the movements. Accordingly, these lower-visuospatial-ability learners suffer from reduced learning outcomes when confronted with non-corresponding gestures.

Our findings on observing gestures regarding the connection between learning outcomes and cortical activation did not show direct evidence for one of the two alternative explanations for the overall pattern of results: we did not find higher DLPFC activation in the conditions with gestures, which might have indicated split attention processes (cf. Osaka, Komori, Morishita, & Osaka, 2007), nor did we find higher activation in the areas associated with the hMNS (IFC as well as IPC) in the conditions with gestures (and particularly not in the non-corresponding gestures conditions and for higher-visuospatial-ability learners), which might have indicated an embodied processing (e.g., Van Gog et al., 2009) of the (non-corresponding) gestures what at least would have given a hint that (higher-visuospatial-ability) learners processed the gestures based on the hMNS. Nevertheless, our results seem to suggest that the conflict created by non-corresponding gestures does not occur at the motor level. Participants who saw non-corresponding gestures activated their MC less than participants who saw no additional gestures. Observing only the fish visualization without any accompanying gestures seems to be processed partly in the areas associated with actual body movements, which might be related to processes of motor imagery (e.g., Savaki & Raos, 2019), whereas non-corresponding gestures seem to be processed in a non-motoric way, that not evokes or rather suppresses such motor imagery strategies, potentially based on the identification of the mismatch between the gestures and the to-be-learned movements. It is possible that the non-corresponding gestures rather had a signalizing effect, for example by alerting learners to pay close attention thereby inviting higher-visuospatial-ability learners to engage in a deeper processing, thereby finding a non-motoric way of representing the information. Lower-visuospatial-ability learners, on the contrary, did not react to non-corresponding gestures in this way or did not know how to do that. Future research is needed to investigate these issues in more detail, particularly by replicating this study with additionally process measures based on eye-tracking (e.g., Van Gog & Scheiter, 2010) as well as an online-assessment of cognitive load (e.g., Cierniak, Scheiter, & Gerjets, 2009). This might provide deeper insights into the question whether the non-corresponding gestures worked for higher-visuospatial-ability learners as positive desirable difficulties or whether different learners mainly differ in their ability to avoid negative split attention effects in the face of conflicting visualizations that are simultaneously presented. Despite these open theoretical issues, our findings suggest in addition to earlier work showing that it is relevant to consider learners' visuospatial ability for deciding whether to provide them with static or dynamic instructional visualizations (Höffler, 2010), that the level of visuospatial ability might also be important for choosing an appropriate way to support learners' processing when learning from (dynamic) visualizations enriched by gestures.

In future research, it might also be possible to provide more causal evidence for the beneficial instructional effects of providing higher-visuospatial-ability learners with non-corresponding gestures and

Table 8
Means and standard errors (in parentheses) for values of classification performance (in % correct) as a function of the between-subjects factors "observing gestures", "making gestures", "visuospatial ability", and "MC activation".

	observing corresponding gestures								observing non-corresponding gestures								observing no gesture							
	making gestures yes (n = 19)				making gestures no (n = 19)				making gestures yes (n = 20)				making gestures no (n = 19)				making gestures yes (n = 18)				making gestures no (n = 19)			
	higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)	
	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)	higher MC (M + 1 SD)	lower MC (M - 1 SD)
Classification performance (in % correct)	66.375 (3.414)	71.355 (6.382)	55.967 (6.185)	66.044 (7.117)	61.459 (7.579)	61.340 (4.359)	64.855 (4.994)	58.910 (4.381)	76.397 (4.537)	69.471 (5.287)	54.570 (4.694)	60.492 (5.742)	77.515 (5.782)	77.534 (16.188)	70.675 (36.732)	52.973 (12.218)	69.602 (6.117)	62.146 (7.897)	58.668 (6.516)	66.515 (6.057)	67.108 (4.849)	68.664 (6.419)	72.266 (4.864)	67.824 (5.433)

VSA = visuospatial ability
MC = motor cortex activation

lower-visuospatial-ability learners with corresponding gestures by designing adaptive experimental conditions that provide exactly these combinations (or the opposite) based on a visuospatial ability pretest: The matching pattern would define an advantageous experimental group and the opposite pattern a disadvantageous group (i.e., providing higher-visuospatial-ability learners with corresponding gestures and lower-visuospatial-ability learners with non-corresponding gestures). Based on our current results, we would expect better overall learning outcomes in the advantageous group as compared to the disadvantageous group due to the improved instructional match between learner prerequisites and learning materials. But even without this further evidence we can already derive as practical implications from our results that it would be useful to assess learners' visuospatial ability and to use non-corresponding gestures only for learners with higher visuospatial ability, whereas for learners with lower visuospatial ability these non-corresponding gestures should be avoided.

4.2. Effects of making gestures

In contrast to our Hypothesis 1, self-performed gestures did not improve learning outcomes in our study. This finding is in line with several recent studies that attempted to augment learning about other topics than non-human movement (e.g., lightning formation, grammar rules) by instructing learners to make gestures while studying an animation and which also did not find an improvement of learning performance due to self-performed gestures (e.g., De Koning & Tabbers, 2013; Post, van Gog, Paas, & Zwaan, 2013). It should be noted that in the present study, learners were instructed to gesture in a way that they thought to be useful in assisting them in understanding the fish movements, whereas in other studies participants were much more restricted in how to gesture or were simply told how to move their hands (e.g., De Koning & Tabbers, 2013). Collectively, the conclusion from this and other studies is that independent from the instructional approach (i.e., instruct specific ways to perform gestures or let learners decide how to perform gestures), making gestures does not seem to particularly benefit learning from dynamic visualizations (nor does it hinder learning; for some specific exceptions under which making gestures might be helpful when considering DLPFC activation in addition see section 4.3.).

However, consistent with our expectations (cf. Hypothesis 1) self-performed gestures activated the areas involved in planning and performing movements – namely the MC and the DLPFC. Also in line with Hypothesis 1, making gestures did activate the hMNS: Participants observing corresponding gestures showed higher IFC activation if they self-gestured than when they did not self-gesture. This might be an indicator that during observing corresponding gestures the IFC is particularly important – at least when the participants were instructed to self-gesture. This result pattern can be brought in line with earlier findings by Brucker et al. (2015), who also found evidence that in the observing-corresponding-gestures conditions particularly lower-visuospatial-ability learners activated their IFC, which probably helped them to achieve the same learning outcomes as higher-visuospatial-ability learners. IFC activation thus is not only important during observing corresponding gestures, but it might be even more important when the observation of corresponding gestures is combined with self-gesturing. However, future research is needed to explore these processes in more detail.

4.3. Influence of cortical activation on learning outcomes

In line with Hypothesis 6, a higher hMNS activation, in terms of IFC activation, was associated with better learning outcomes. In other words, participants who activated their IFC (which is part of the hMNS) more strongly tended to show better learning outcomes than participants who did activate this area less strongly. Thus, our results showed again that the activation of the hMNS might be helpful for learning about biological movements with dynamic visualizations. This pattern of results was obtained independently of learners' visuospatial ability

Table 9

Means and standard errors (in parentheses) for values of classification performance (in % correct) as a function of the between-subjects factors "observing gestures", "making gestures", "visuospatial ability", and "DLPFC activation".

	observing corresponding gestures								observing non-corresponding gestures								observing no gesture							
	making gestures yes (n = 19)				making gestures no (n = 19)				making gestures yes (n = 20)				making gestures no (n = 19)				making gestures yes (n = 18)				making gestures no (n = 19)			
	higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)		higher VSA (M + 1 SD)		lower VSA (M - 1 SD)	
	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)	higher DLPFC (M + 1 SD)	lower DLPFC (M - 1 SD)
	SD	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*	*SD*
Classification performance (in % correct)	65.090 (3.575)	72.729 (5.692)	54.010 (6.871)	67.244 (7.051)	64.408 (6.254)	60.979 (3.637)	64.503 (4.072)	57.531 (5.265)	72.634 (4.296)	69.879 (5.311)	55.717 (3.655)	65.311 (7.172)	108.531 (22.940)	85.415 (5.883)	61.036 (10.299)	23.721 (15.273)	67.635 (5.007)	70.341 (5.075)	59.436 (8.253)	57.553 (4.331)	67.033 (5.714)	68.251 (5.673)	74.356 (6.544)	66.500 (6.045)

VSA = visuospatial ability
DLPFC = dorsolateral prefrontal cortex activation

and independently of the different gesturing instructions provided to participants (observing and/or making different types of gestures). To sum up, all participants (with higher and lower visuospatial ability) in all gesturing instructions (observing corresponding, non-corresponding, or no gesture, making or not making gestures) tended to benefit from activating their hMNS in terms of their IFC.

Interestingly, our results yielded a very differentiated pattern of how DLPFC activation in combination with the three factors *observing gestures, making gestures,* and *learners' visuospatial ability* affects learning outcomes. For higher-visuospatial-ability learners, who showed higher DLPFC activation, the gesturing factors (observing different types of gestures and/or making gestures) did not influence learning outcomes. However, for higher-visuospatial-ability learners, who showed lower DLPFC activation, self-gesturing seemed to be beneficial during observing corresponding gestures, whereas it seemed to be detrimental during observing non-corresponding gestures. The combination of the non-corresponding gestures with self-gesturing might stimulate competing strategies that cannot be solved even by higher-visuospatial ability learners without the activation of the DLPFC. It must be noted that the result pattern that the combination of corresponding gestures and self-gesturing (for participants with higher visuospatial ability and lower DLPFC activation) leads to better learning outcomes, speaks partly against our Hypothesis 5 stating that a combination of both observing gestures and making gestures might be detrimental due to competing approaches. At least for this specific subgroup (higher visuospatial ability, lower DLPFC activation) this particular combination (corresponding gestures, making own gestures) is helpful for learning.

For lower-visuospatial-ability learners, who showed higher DLPFC activation, making gestures was detrimental during observing non-corresponding gestures, whereas for lower-visuospatial-ability learners, who showed lower DLPFC activation, making gestures was beneficial during observing these non-corresponding gestures. Taken together, making gestures is particularly helpful for higher-visuospatial-ability learners, who showed lower DLPFC activation and observe corresponding gestures as well as for lower-visuospatial-ability learners, who showed lower DLPFC activation and observe non-corresponding gestures. Particularly the latter group has no other helping factors at their disposal: They neither possess sufficient visuospatial ability, nor do they have access to corresponding gestures, nor recruit higher levels of DLPFC activation. Thus, in this special case, where no other facilitating factors are available, making gestures developed its potential. Interestingly, for lower-visuospatial-ability learners with lower DLPFC activation, self-gesturing is (as aforementioned) helpful, but if these learners observe the non-corresponding gestures and show higher DLPFC activation, they cannot make use of self-gesturing: in contrast, it even has detrimental effects. The DLPFC is involved in executive functions and motor planning, organization, and regulation (cf. Miller & Cohen, 2001; Ridderinkhof et al., 2004) and our results indicate that it's activation might replace the positive effects of making gestures for this subgroup (lower-visuospatial ability learners observing non-corresponding gestures) or at least indicate that there occur different approaches or strategies (by simultaneously self-gesture and showing higher DLPFC activation) that might not be compatible with each other. However, as these analyses took place on an exploratory level, future research is needed to explicitly address the role of DLPFC activation during learning with dynamic visualizaions and additonal gestures in more detail.

4.4. Limitations and future directions

In this study, IFC activation tended to predict better learning outcomes (Hypothesis 6). However, compared to the Brucker et al. (2015) study, we did not find the result pattern that IPC activation compensates for missing support based on learners' visuospatial ability or corresponding gestures. This might be explained by the fact that in the present study participants who neither have higher visuospatial ability nor corresponding gestures at their disposal (i.e., the group of
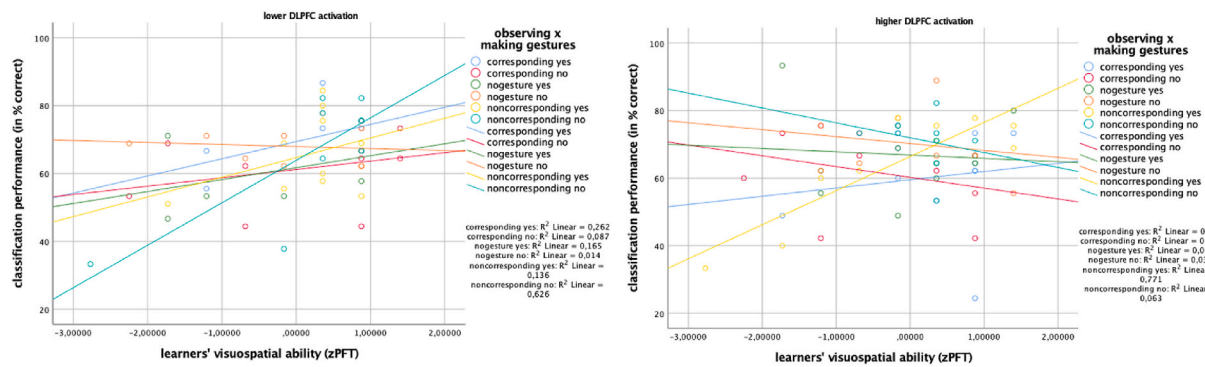
Fig. 5. Marginal significant four-way-interaction between observing gestures, making gestures, learners' visuospatial ability, and DLPFC activation on learning outcomes.

lower-visuospatial-ability learners who observed non-corresponding gestures) still could focus on the fish visualization. This was possible because in this study the gestures were presented at the same time as the fish movements, whereas in the Brucker et al. (2015) study the gestures and the fish visualizations were presented sequentially. To disentangle these effects, further research should investigate direct comparisons of sequential and simultaneous presentations of additional gestures. Moreover, one caveat of the study is that the combination of the fish movement visualizations plus the video of (corresponding or non-corresponding) gestures plus the instruction to self-gesture might have led to high working memory load for the participants in this experimental group due to the requirement to integrate these three sources. Thus, future research should also address this limitation of the present study by replicating this study with cognitive load measurement (cf. Sweller, Ayres, & Kalyuga, 2011) as well as think-aloud protocols so that it is possible to discover the strategies learners use when observing and making gestures in learning from dynamic visualizations. Furthermore, as Wakefield, Congdon, Novack, Goldin-Meadow, and James (2019) point out it is important to further identify potential neural correlates of (gesture-supported) learning – in our case with dynamic visualizations – and to further unravel the relations between activation in different parts of the brain and learning outcomes.

5. Conclusion

The present study provides a starting point from which future research endeavors within this emerging field of research can be explored with the goal to incorporate (observing and making) gestures in a way that learning about non-human movements from dynamic visualizations is enhanced. In conclusion, this study shows that observing additional gestures might be helpful for learning about biological movements, but that learners react differently to non-corresponding gestures depending on their level of visuospatial ability. Thus, different types of gestures might be best deployed by using an adaptive approach: Higher-visuospatial-ability learners should be challenged with non-corresponding gestures, whereas lower-visuospatial-ability learners might even be impaired by these gestures and need to be supported by means of other instructional aids.

Credit author statement

Birgit Brucker: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing- Reviewing and Editing, Visualization, Supervision, Project administration. **Björn de Koning:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing- Reviewing and Editing, Visualization. **David Rosenbaum:** Methodology, Software, Formal analysis, Data curation, Visualization. **Ann-Christine Ehlis:** Conceptualization, Methodology, Software,

Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization, Supervision, Project administration. **Peter Gerjets:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – original draft, Writing- Reviewing and Editing, Supervision, Funding acquisition.

References

Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior, 25*, 348–353. https://doi.org/10.1016/j.chb.2008.12.013

Ayres, P., & Sweller, J. (2014). The split-attention-principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 206–226). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.011.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology, 23*, 638–663. https://doi.org/10.1002/acp.1473

Brucker, B., Ehlis, A.-C., Häußinger, F. B., Fallgatter, A. J., & Gerjets, P. (2015). Watching corresponding gestures facilitates learning with animations by activating human mirror-neurons: An fNIRS study. *Learning and Instruction, 36*, 27–37. https://doi.org/10.1016/j.learninstruc.2014.11.003

Castro-Alonso, J. C., & Atit, K. (2019). Different abilities controlled by visuospatial processing. In J. C. Castro-Alonso (Ed.), *Visuospatial processing for education in health and natural sciences.* https://doi.org/10.1007/978-3-030-20969-8_2

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015). Animations showing Lego manipulative tasks: Three potential moderators of effectiveness. *Computers & Education, 85*, 1–13. https://doi.org/10.1016/j.compedu.2014.12.022

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2016). Comparing apples and oranges? A critical look at research on learning from statics versus animations. *Computers & Education, 102*, 234–243. https://doi.org/10.1016/j.compedu.2016.09.004

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*, 102–116. https://doi.org/10.1037/a0021790

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*, 315–324. https://doi.org/10.1016/j.chb.2008.12.020

Cook, S. M., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development, 7*, 211–232. https://doi.org/10.1207/s15327647jcd0702_4

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2010). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage, 54*, 2808–2821. https://doi.org/10.1016/j.neuroimage.2010.10.069

De Koning, B. B., Hoogerheide, V., & Boucheix, J.-M. (2018). Developments and trends in learning with instructional video. *Computers in Human Behavior, 89*, 395–398. https://doi.org/10.1016/j.chb.2018.08.055

De Koning, B. B., & Jarodzka, H. (2017). Attention guidance strategies for supporting learning from dynamic visualizations. In R. Lowe, & R. Ploetzner (Eds.), *Learning from dynamic visualization* (pp. 255–278). Cham: Springer. https://doi.org/10.1007/978-3-319-56204-9_11.

De Koning, B. B., Marcus, N., Brucker, B., & Ayres, P. (2019). Does observing hand actions in animations and static graphics differentially affect learning of hand-manipulative tasks? *Computers & Education, 141*, 103636. https://doi.org/10.1016/j.compedu.2019.103636

De Koning, B. B., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: An embodied perspective. *Educational Psychology Review, 23*, 501–521. https://doi.org/10.1007/s10648-011-9173-8

De Koning, B. B., & Tabbers, H. K. (2013). Gestures in instructional animations: A helping hand to understanding non-human movements? *Applied Cognitive Psychology, 27*, 683–689. https://doi.org/10.1002/acp.2937

Ehlis, A.-C., Schneider, S., Dresler, T., & Fallgatter, A. J. (2014). Application of functional near-infrared spectroscopy in psychiatry. *NeuroImage, 85*, 478–488. https://doi.org/10.1016/j.neuroimage.2013.03.067

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton: Educational Testing Service.

Engel, A., Burke, M., Fiehler, K., Bien, S., & Rösler, F. (2008). What activates the human mirror neuron system during observation of artificial movements: Bottom-up visual features or top-down intentions? *Neuropsychologia, 46*, 2033–2042. https://doi.org/10.1016/j.neuropsychologia.2008.01.025

Fogassi, L., & Ferrari, P. F. (2011). *Mirror systems* (Vol. 2, pp. 22–38). Wiley Interdisciplinary Reviews: Cognitive Science. https://doi.org/10.1002/wcs.89

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516–522. https://doi.org/10.1111/1467-9280.00395

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084–1102. https://doi.org/10.1037/0278-7393.18.5.1084

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction, 14*, 343–351. https://doi.org/10.1016/j.learninstruc.2004.06.007

Hegarty, M., & Waller, D. (2005). Individual differences in spatial ability. In P. Shah, & A. Miyake (Eds.), *Handbook of visuospatial thinking*. Cambridge University Press.

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—a meta-analytic review. *Educational Psychological Review, 22*, 245–269. https://doi.org/10.1007/s10648-010-9126-7

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722–738. https://doi.org/10.1016/j.learninstruc.2007.09.013

Imhof, B., Scheiter, K., Edelmann, J., & Gerjets, P. (2012). How temporal and spatial aspects of presenting visualizations affect learning about locomotion patterns. *Learning and Instruction, 22*, 193–205. https://doi.org/10.1016/j.learninstruc.2011.10.006

Imhof, B., Scheiter, K., Edelmann, J., & Gerjets, P. (2013). Learning about locomotion patterns: Effective use of multiple pictures and motion-indicating arrows. *Computers & Education, 65*, 45–55. https://doi.org/10.1016/j.compedu.2013.01.017

Imhof, B., Scheiter, K., & Gerjets, P. (2011). Learning about locomotion patterns from visualizations: Effects of presentation format and realism. *Computers & Education, 57*, 1961–1970. https://doi.org/10.1016/j.compedu.2013.01.017

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 370–375.

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247–262). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.013.

Kastner, L., Umbach, N., Jusyte, A., Cervera-Torres, S., Ruiz Fernández, S., Nommensen, S., et al. (2021). Designing visual-arts education programs for transfer effects: Development and experimental evaluation of (digital) drawing courses in the art museum designed to promote adolescents' socio-emotional skills. *Frontiers in Psychology, 11*, Article 4031. https://doi.org/10.3389/fpsyg.2020.603984

Korbach, A., Ginns, P., Brünken, R., & Park, B. (2020). Should learners use their hands for learning? Results from an eye-tracking study. *Journal of Computer Assisted Learning, 36*(1), 102–113. https://doi.org/10.1111/jcal.12396

Lee, H. (2007). Instructional design of web-based simulations for learners with different levels of spatiality. *Instructional Science, 35*, 467–479. https://doi.org/10.1007/s11251-006-9010-5

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*, 225–244. https://doi.org/10.1007/BF03172967

Marcus, N., Cleary, B., Wong, A., & Ayres, P. (2013). Should hand actions be observed when learning hand motor skills from instructional animations? *Computers in Human Behavior, 29*, 2172–2178. https://doi.org/10.1016/j.chb.2013.04.035

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*, 256–265. https://doi.org/10.1037/1076-898X.11.4.256

Memmert, D., Hagemann, H., Althoetmar, R., Geppert, S., & Seiler, D. (2009). Conditions of practice in perceptual skill learning. *Research Quarterly for Exercise & Sport, 80*, 32–43. https://doi.org/10.1080/02701367.2009.10599527

Mierowsky, R., Marcus, N., & Ayres, P. (2020). Using mimicking gestures to improve observational learning from instructional videos. *Educational Psychology, 40*, 550–569. https://doi.org/10.1080/01443410.2019.1650896

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Montgomery, K. J., Isenberg, N., & Haxby, J. V. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Social Cognitive and Affective Neuroscience, 2*, 114–122. https://doi.org/10.1093/scan/nsm004

Obrig, H., & Villringer, A. (2003). Beyond the visible – imaging the human brain with light. *Journal of Cerebral Blood Flow and Metabolism, 23*, 1–18. https://doi.org/10.1097/01.WCB.0000043472.45775.29

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., et al. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10-20 system oriented for transcranial functional brain mapping. *NeuroImage, 21*, 99–111. https://doi.org/10.1016/j.neuroimage.2003.08.026

Osaka, M., Komori, M., Morishita, M., & Osaka, N. (2007). Neural bases of focusing attention in working memory: An fMRI study based on group differences. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 130–139. https://doi.org/10.3758/CABN.7.2.130

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*, 27–45. https://doi.org/10.1007/s10648-011-9179-2

Ploetzner, R., Berney, S., & Bétrancourt, M. (2021). When learning from animations is more successful than learning from static pictures: Learning the specifics of change. *Instructional Science, 49*, 497–514. https://doi.org/10.1007/s11251-021-09541-w

Post, L. S., van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior, 29*, 1450–1455. https://doi.org/10.1016/j.chb.2013.01.005

Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56*, 129–140. https://doi.org/10.1016/j.bandc.2004.09.016

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

Sanchez, C., & Wiley, J. (2017). Dynamic visuospatial ability and learning from dynamic visualizations. In R. Lowe, & R. Ploetzner (Eds.), *Learning from dynamic visualization* (pp. 155–176). Cham: Springer. https://doi.org/10.1007/978-3-319-56204-9_7.

Sanes, J. N., & Donoghue, J. P. (2000). Plasticity and primary motor cortex. *Annual Review of Neuroscience, 23*, 393–415. https://doi.org/10.1146/annurev.neuro.23.1.393

Savaki, H. E., & Raos, V. (2019). Action perception and motor imagery: Mental practice of action. *Progress in Neurobiology, 175*, 107–125. https://doi.org/10.1016/j.pneurobio.2019.01.007

Scharinger, C., Schüler, A., & Gerjets, P. (2020). Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations. *International Journal of Psychophysiology, 158*, 201–214. https://doi.org/10.1016/j.ijpsycho.2020.09.014

Schecklmann, M., Mann, A., Langguth, B., Ehlis, A.-C., Fallgatter, A. J., & Haeussinger, F. B. (2017). The temporal muscle of the head can cause artifacts in optical imaging studies with functional near-infrared spectroscopy. *Frontiers in Human Neuroscience, 11*(456), 1–13. https://doi.org/10.3389/fnhum.2017.00456

Scheiter, K., Brucker, B., & Ainsworth, S. (2020). Now move like that fish": Can enactment help learners come to understand dynamic motion presented in photographs and videos? *Computers & Education, 155*, Article 103934. https://doi.org/10.1016/j.compedu.2020.103934

Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learning and Instruction, 49*, 218–231. https://doi.org/10.1016/j.learninstruc.2017.03.001

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience, 8*, 426. https://doi.org/10.3389/fnbeh.2014.00426

Stieff, M. (2017). Drawing for promoting learning and engagement with dynamic visualizations. In R. Lowe, & R. Ploetzner (Eds.), *Learning from dynamic visualization* (pp. 333–356). Cham: Springer. https://doi.org/10.1007/978-3-319-56204-9_14.

Sun, P.-P., Tan, F.-L., Zhang, Z., Jiang, Y.-H., Zhao, Y., & Zhu, C.-Z. (2018). Feasibility of functional near-infrared spectroscopy (fNIRS) to investigate the mirror neuron system: An experimental study in a real-life situation. *Frontiers in Human Neuroscience, 12*, 86. https://doi.org/10.3389/fnhum.2018.00086

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory*. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-8126-4_6.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28*, 187–204. https://doi.org/10.1016/S0361-476X(02)00007-3

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror-neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review, 21*, 21–30. https://doi.org/10.1007/s10648-008-9094-3

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99. https://doi.org/10.1016/j.learninstruc.2009.02.009

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three dimensional spatial visualization. *Perceptual & Motor Skills, 47*, 599e604. https://doi.org/10.2466/pms.1978.47.2.599

Wakefield, E. M., Congdon, E. L., Novack, M. A., Goldin-Meadow, S., & James, K. H. (2019). Learning math by hand: The neural effects of gesture-based instruction in 8-year-old children. *Attention, Perception, & Psychophysics, 81*, 2343–2353. https://doi.org/10.3758/s13414-019-01755-y

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments

and memory. *Memory & Cognition, 41*, 229–241. https://doi.org/10.3758/s13421-012-0255-8

Zhang, B., Li, Z., Cao, X., Ye, Q., Chen, C., Shen, L., et al. (2016). Output constraint transfer for kernelized correlation filter in tracking. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47*(4), 693–703. https://doi.org/10.1109/TSMC.2016.2629509