

In the eyes of the beholder: How experts and novices interpret dynamic stimuli

Halszka Jarodzka ^{a,*}, Katharina Scheiter ^b, Peter Gerjets ^a, Tamara van Gog ^c

^a Knowledge Media Research Center, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

^b Applied Cognitive Psychology and Media Psychology, University of Tuebingen, Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

^c Centre for Learning Sciences and Technologies, Open University of The Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands

Abstract

Tasks with a complex, dynamic visual component require not only the acquisition of conceptual/procedural but also of perceptual/attentional skills. This study examined expertise differences in perceiving and interpreting complex, dynamic visual stimuli on a performance and on a process level, including perceptual and conceptual strategies. Performance, eye movement, and verbal report data were obtained from seven experts and 14 novices. Results show that experts compared to novices attend more to relevant aspects of the stimulus, use more heterogeneous task approaches, and use knowledge-based shortcuts. Implications for instructional design for the acquisition of perceptual skills are discussed. © 2009 Elsevier Ltd. All rights reserved.

Keywords: Expertise; Eye tracking; Verbal reports; Instructional design

1. Introduction

Many instructional materials directly or indirectly convey expert knowledge to learners (Feldon, 2007). For instance, a very effective instructional technique is the use of worked examples, in which novices are shown a worked-out expert solution procedure (Sweller, Van Merriënboer, & Paas, 1998). However, worked examples have so far mainly been used to convey conceptual and procedural aspects of expertise. In many domains, expert performance also comprises perceptual/attentional skills, that is, the ability to perceive the relevant out of irrelevant information in complex, highly visual stimuli and to draw inferences based upon the perceived information (e.g., X-rays in medical diagnosis see Lesgold et al., 1988; weather maps in meteorology see Canham & Hegarty, 2010).

An important question is, if these perceptual/attentional aspects of expertise, that is top-down processing of perceptual stimuli, could also be conveyed to novices to facilitate skill acquisition. Since providing information at a conceptual level has been shown to be effective for the acquisition of

conceptual/procedural skills, it might be necessary to grant novices more direct access to the perceptual/attentional processes underlying experts' performance; for example, by guiding the novices' attention to critical perceptual information during the study of worked examples based on evidence regarding experts' perceptual processes, such as eye movements (Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009; for process-oriented worked examples see Van Gog, Paas, & Van Merriënboer, 2008). However, as will be discussed in the next section, not much is known about how experts allocate their attention during task performance, and how their attention allocation differs from novices, especially in domains that involve complex, dynamic visual stimuli. Therefore, the present study examined expertise in the perceptual/attentional and conceptual processes involved in the interpretation of complex dynamic visual stimuli in the domain of fish locomotion.

1.1. Expertise differences in perceiving and interpreting complex visual stimuli

Many studies have addressed the issue of how experts perceive and interpret visual stimuli by using eye tracking

* Corresponding author. Tel.: +49 7071 979 334; fax: +49 7071 979 126.
E-mail address: h.jarodzka@iwm-kmrc.de (H. Jarodzka).

methodology. In particular, these studies provide information on attention allocation through eye movement analyses. Gazes on relevant information have higher densities than on irrelevant information. Haider and Frensch (1999) stated in their information-reduction hypothesis that with increasing expertise people learn to distinguish between relevant and irrelevant information and therefore concentrate on processing mostly relevant information (see also Canham & Hegarty, 2010). Using a letter string task in which the location of relevant information was varied, Haider and Frensch (1999) corroborated this hypothesis with eye movement data. In the domain of art, Antes and Kristjanson (1991) found that experts (i.e., artists) compared to novices (i.e., non-artists) had higher fixation densities on important aspects of the paintings. Charness, Reingold, Pomplun, and Stampe (2001) also found in their studies of expertise effects in chess performance that experts had a greater proportion of fixations on relevant rather than on irrelevant areas.¹

The above studies, however, used static visual stimuli; few eye tracking studies have taken place using *dynamic* visual stimuli (e.g., for air traffic control see Ellis, 1986) to identify differences in visual attention in general, and none has yet investigated attention to visually complex and dynamic stimuli that contain relevant and irrelevant information. One exception to the use of dynamic material is the study by Moreno, Reina, Luis, and Sabido (2002), in which novice and expert gymnastic coaches inspected videos of gymnastic techniques and indicated errors in performance. They found that experts had longer and fewer fixations than novices, and attributed this to the fact that experts attended more to informative (i.e., relevant) areas and ignored uninformative (i.e., irrelevant) ones. However, this assumption was not directly tested in the study. Another exception is the study by Lowe (1999) on the interpretation of weather maps, where novices mentioned more often irrelevant but perceptually salient features after the inspection of the dynamic weather maps, suggesting that they attended more to these features. However, this assumption was not directly tested with eye tracking.

So far, eye tracking has provided interesting insights into how experts differ from novices using *single basic eye tracking indicators* (e.g., number or duration of fixations) when processing tasks with a high visual component. Little research has been done to identify expertise effects in perceptual strategies used, that is, complex *patterns* of eye movements when processing dynamic visual stimuli. There are at least two open questions associated with this issue.

First, one may ask whether experts' perceptual strategies are characterized by an optimization of the strategies that a novice would use or whether experts act within their domains in a qualitatively different way than novices do. In the case of conceptual processing it has been shown that experts' highly integrated knowledge structures enable them to

use shortcuts during task processing (e.g., medical diagnosis based on textual descriptions of cases: Boshuizen & Schmidt, 1992), which results in very different strategies for different expertise levels. In contrast, in case of perceptual processing it has been found that experts process and report visual case information more elaborately than novices (e.g., for diagnosing X-ray pictures see Lesgold, 1984), which rather seems to favor the idea that experts optimize strategies that are used imperfectly by novices.

Second, it is unclear whether experts' perceptual strategies are the same for all experts or differ among experts (cf. Medin et al., 2006). For example, Medin, Lynch, Coley, and Atran (1997) showed that, while some experts were very similar in their approaches to the task (namely, parks maintenance personnel vs. scientific taxonomists in categorizing trees) other experts differed in their approach to the same task (namely, landscapers vs. scientific taxonomists). Moreover, Medin et al. (2006) showed in another study that even in well-structured domains, like the categorization of freshwater fish, expertise did not lead to common conceptualizations. The authors concluded that even if the outcome of a categorization process is similar across experts the underlying processes are not necessarily so.

The above conclusion, however, might be due to the characteristics of biological taxonomies in which one has to deal with a diversity of the features of the various species. These taxonomies are often invented, that is, they are conceptual schemas not inherent to a domain; hence, there may be multiple categories for the same species depending on the focus of the taxonomy (e.g., morphology, genetics, etc.). Furthermore, categorizing species is based on multiple features, which do not need to be considered in a hierarchical order. Experts may differ in the features that they emphasize in order to achieve a categorization depending on their prior experiences, that is, on their learning history. For instance, an expert with lots of outdoor experience will potentially focus on other features (e.g., those easily observable in a natural setting) than an expert who mostly deals with formalin preparations and textbooks (e.g., features requiring a close inspection). Hence, these differences in learning history, which will be reflected in the organization and accessibility of knowledge, may affect how experts attend to features (for endogenous attentional control see Corbetta & Shulman, 2002; Posner, 1980) yielding diverse perceptual strategies for experts. On the other hand, novices' attention may be mainly guided by exogenous features, for instance, by salience (Lowe, 1999). Consequently, their perceptual strategies might be rather homogeneous compared to those of experts.

Therefore, two issues are very important when studying expertise differences: (a) namely the specialization of the experts under investigation; and (b) the nature of the task (Medin et al., 1997, 2006). When investigating perceptual strategies, the task to-be-performed is even more important because eye movements strongly depend on it (Yarbus, 1967).

In sum, many tasks require the use of perceptual skills. However, instructional materials seldom teach these skills directly. To design instructional material that is suited to

¹ In both studies (i.e., Antes & Kristjanson, 1991; Charness et al., 2001) relevant areas were determined a priori by an independent expert in the field of study.

convey perceptual skills, knowledge on expertise differences in the analysis of complex, dynamic visual stimuli is necessary. Although some studies dealing with complex, static visual stimuli exist, research on interpreting complex, dynamic visual stimuli is still rare. The aim of the present study was to investigate expertise differences as a prerequisite for designing effective instructional material in the domain of fish locomotion.

1.2. Describing fish locomotion patterns

In the present study expertise differences were investigated for the description of locomotion patterns of fish swimming forward. This task was chosen for two reasons. First, this task is dynamic and has a highly visual component, thus, perceptual strategies are very likely to play a crucial role in task processing. Second, it is a good example of the topic of biodiversity, which is a core topic in biology and part of the curriculum at school and university, because fishes are the most diverse vertebrates; besides a large diversity of forms, colors, and habitats, they are also very diverse in terms of their locomotion patterns (Videler, 1993).

To describe a locomotion pattern, the following guidelines should be applied (Lindsey, 1978). First, it has to be decided *which* part of the body is used to produce propulsion. This can be either the body itself or the fins. Second, it has to be decided *how* this part moves. This can be either in an undulating (i.e., wavelike) or an oscillating (i.e., paddlelike) way. These decisions form the basis of what will be called a *locomotion description strategy* here, which is taught to university students of biology when they receive training in marine zoology. Application of this strategy, would lead the observer to attend only to those parts of the fish body that might be crucial for the fish's locomotion (i.e., the fins and the body) and ignore the parts that are irrelevant to the locomotion (i.e., eyes or colorful patches).

Both novices and experts have to rely on the locomotion description strategy when classifying locomotion of unfamiliar fish. However, with familiar fish, experts may be able to automatically retrieve knowledge on the specific locomotion pattern associated with this particular fish species from memory. In such a case a *species classification strategy* will be used, and other features of the fish are potentially relevant. Specifically, some species can be easily recognized due to salient static features (e.g., a specific colorful pattern on the fish's body is a characteristic of a particular fish species). Experts can use these features to classify a fish, and upon activating that particular schema, the knowledge on its locomotion pattern is automatically activated. For novices, this type of knowledge-based shortcut is not available, so they need to rely on the locomotion description strategy.

1.3. Research questions – hypotheses

This study investigated how different levels of expertise in biology would be reflected in differences in task performance (i.e., correct description of the locomotion) as well as in

different strategies. These differences were investigated by means of eye tracking and cued retrospective verbal reports (cf. Van Gog, Paas, Van Merriënboer, & Witte, 2005). Using these methods in conjunction was expected, first, to provide insights into the perceptual and conceptual processing of complex visual dynamic stimuli and, second, to provide an example of how eye tracking data on dynamic stimulus material can be analyzed in detail. In addition, findings obtained from this study were intended to inform the instructional design of process-oriented worked examples that teach perceptual skills by guiding students' attention (Van Gog, 2009).

It was hypothesized (Hypothesis 1) that experts would perform more accurately and faster than novices on a locomotion description task (reflected in higher correctness rates and shorter mean viewing times of the stimulus). The respective test, however, mostly served as a manipulation check concerning our assignment of individuals to different levels of expertise.

More important, based on prior findings with static, complex visual stimuli (Antes & Kristjanson, 1991; Charness et al., 2001; Haider & Frensch, 1999) it was hypothesized (Hypothesis 2) that also in this dynamic domain the process data of experts would show that they attend more to relevant information than novices, who would rather attend to perceptually salient, but potentially conceptually irrelevant information (cf. Lowe, 1999). Whether experts attend to features that are relevant for either locomotion description or species classification was explored by means of analyzing the gaze duration on these features. It was predicted that experts would attend more features relevant to the species than novices (Hypothesis 3).

Moreover, whether expertise yields either more diversity or more homogeneity in terms of the perceptual strategies used was investigated by analyzing experts' and novices' gaze sequences. It was assumed that novices' perceptual strategies will be guided by the salience of single features and, thus, possibly leading to a more homogeneous gaze pattern. On the other hand, experts' perceptual strategies were assumed to be controlled in an endogenous way. Hence, experts were expected to be characterized by rather heterogeneous gaze patterns depending on their individual learning history that would lead to different processing strategies (Hypothesis 4).

Finally, it was expected that experts would verbalize less information than novices due to schema automation and thus use fewer words in their description of how they accomplished that task (Ericsson & Simon, 1980). In line with Boshuizen and Schmidt (1992) their verbalizations were expected to contain more encapsulating technical terms (Hypothesis 5).

2. Method

2.1. Participants and design

Participants in the study were 21 individuals ($M = 26.57$ years, $SD = 5.98$; 10 females and 11 males) with two different levels of expertise. Of them seven were experts,

that is, professors, PhDs, or advanced PhD students, with a mean age of 31.43 years ($SD = 8.54$). The novices were 14 biology students of the University of Tuebingen, Germany, with a mean age of 24.14 years ($SD = 1.51$), who had basic knowledge of fish anatomy, terminology and locomotion pattern, but had very little if any experience with classifying different locomotion patterns. We determined via a questionnaire that experts were more interested in fish and had more relevant practical experience with fish locomotion (i.e., they engaged more frequently in snorkeling and diving than novices), because not only years of experience in a domain, but also the nature of this experience is important (Ericsson & Lehmann, 1996).

2.2. Material and apparatus

2.2.1. Stimulus materials

The materials consisted of four digital videos (for screen shots see Fig. 1) in an audio video interleave format (.avi), sized 360 * 480 pxl (corresponding to 3.74 * 4.98 inches). Each video depicted a single fish swimming, whereby each fish deployed a different locomotion pattern (i.e., tetraodonti-, subcarangi-, labri-, and balistiform) that had to be described by the participants. The videos were rather short (8.79 s on average), but looped automatically until participants stopped them. This was done in order to avoid an artificial situation for both groups. On the one hand, novices might not be able to describe the locomotion pattern at all after a too short presentation. On the other hand, if experts were forced to look at a stimulus that they had interpreted already, they might start to look at completely irrelevant features just out of boredom.

2.2.2. Eye tracking equipment

Eye movements were recorded with a Tobii 1750 remote eye tracking system with temporal resolution of 50 Hz, and analyzed with ClearView 2.7.0 software (<http://www.tobii.com>).

The verbal data were recorded by Camtasia 3.0 software using a standard microphone attached to the stimulus PC.

2.2.3. Cued retrospective reporting

Cued retrospective reporting is a verbal reporting procedure in which participants verbalize their thought processes during task performance after completing the task, based on a cue of their performance. The cue used here consisted of the videos with the recordings of participants' own eye movements superimposed onto the video (Van Gog et al., 2005). This so-called "gaze replay" showed fixations with a definition of maximum 50 pxl and at least 200 ms including a gaze trail of 500 ms (for an example screenshot of the gaze replay see Fig. 2). The gaze replay was at full speed and although participants were not allowed to pause it, they could watch it again.

2.3. Procedure

The experiment was run in individual sessions of approximately 20 min. At the beginning, the eye tracking system was adjusted to the individual features of the participant based on a nine-point calibration. Before watching the videos, participants received the following instruction: "While watching the video, please take a look at the way the fish swims. Subsequent to watching the video, you will have to describe the fish's locomotion pattern. You will be allowed to watch the video as often as you like." Then, participants watched the looped video while their eye movements were recorded until they stopped it themselves. After having watched the video, participants were asked to describe the locomotion pattern of the depicted fish verbally. Subsequently, they were asked to provide the cued retrospective reports (Van Gog et al., 2005), for which participants received the following instruction: "In the following you will see a red spot moving over the screen. This is the recording of your eye movements. The lines that

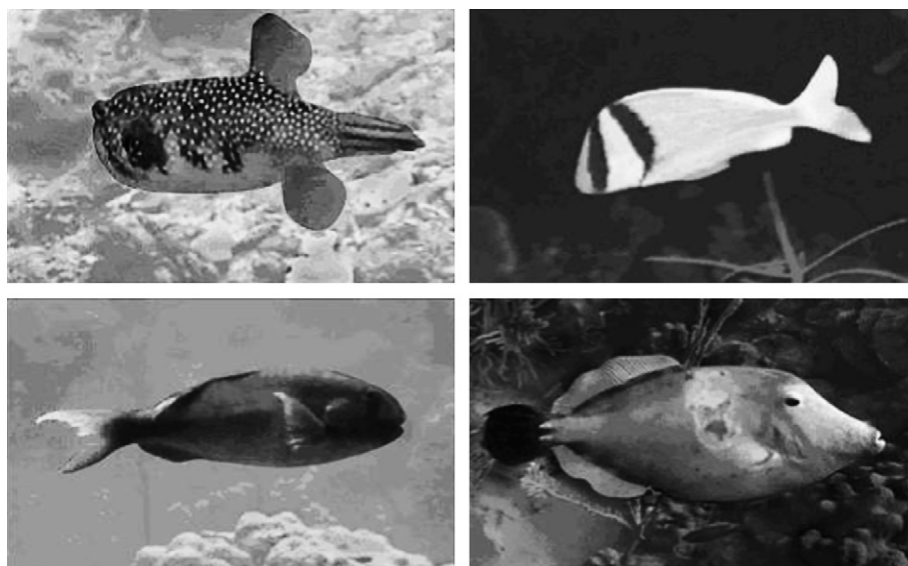


Fig. 1. Screenshots from the four videos used in the study.



Fig. 2. Screenshot of an exemplary gaze replay (gaze plot on 100 ms segment). In the replay, only one dot indicating the fixation was visible at a time and moved across the screen (suggested here by the different dots).

are drawn by the spot represent the path of your eye movement. The size of the spot corresponds to your fixation duration, that is, the bigger the spot is the longer you have looked at this point. Please watch the replay and tell me what you were thinking during your first viewing”. This procedure was repeated for all four videos.

2.4. Data analysis

2.4.1. Correctness of description

Based upon Lindsey's (1978) description of fish locomotion patterns the following coding system for the correctness of description was applied. The first dependent variable was the *correctness of the technical terms* used for describing the locomotion pattern after watching each video. The naming of the correct technical term for describing the locomotion pattern yielded 1 point, whereas the naming of a wrong technical term or none yielded 0 points. Two other dependent variables were obtained from the descriptions of the locomotion pattern. First, it was coded whether participants described correctly *which* part of the fish body had been moving. Second, it was coded whether participants described correctly *how* the part of the fish body had been moving. With regard to the latter two aspects, the technically correct and complete answers yielded 1 point each, a correct but colloquial or only partially complete description yielded 0.5 point, and everything else yielded 0 points. Thus, participants could receive a maximum of 4 points for each of the three dependent variables for all videos, because there were four fish that were presented in four respective videos. The coding system is summarized in Table 1. Two raters conducted the coding of the verbal data independently, and showed complete agreement.

2.4.2. Gaze data

To analyze participants' eye movements, we created “dynamic” areas of interest (AOIs), which are precisely specified areas of an object (e.g., a fin, a patch of colorful stripes, the mouth) and for which it can be determined whether and for which amount of time a participant is looking at this area. Each part of the fish body was an AOI (all fins, the body,

Table 1

Coding system for the locomotion pattern description.

Technical term?	Which part moves?		How does it move?	
	1 point	0.5 point	1 point	0.5 point
Video 1 Tetraodonti	Anal and dorsal fin	e.g., upper and lower fin	Oscillating	e.g., moving like a paddle
Video 2 Subcarangi	Caudal fin	e.g., back part	Undulating	e.g., wavelike
Video 3 Labri	Pectoral fins	e.g., frontal fin	Oscillating	e.g., moving like a paddle
Video 4 Balisti	Anal and dorsal fin	e.g., upper and lower fin	Undulating	e.g., wavelike

and salient features, like eyes, colorful stripes, or taints). Each video was divided into segments of 100 ms each. For each segment AOIs were defined manually. The length of the segments was determined based on the maximum amount of time for which AOIs did not change within each segment (i.e., because the stimuli were dynamic, the positions of AOIs could change). In a first step, the data for each AOI were aggregated per video (i.e., across all 100 ms segments). In a second step, the data for the AOIs were aggregated across all four videos according to whether the body part represented by an AOI was: (a) relevant for the locomotion description strategy only; (b) relevant for the species classification strategy only; (c) relevant for both strategies at the same time; or (d) irrelevant for both classification strategies (see Table 2). The assignment of AOIs to these four categories was determined a priori by a domain expert (cf. Antes & Kristjanson, 1991; Charness et al., 2001). We refer to these *aggregated* AOIs as AAOIs in the remainder of the present article. The AAOIs were not all equal in size; however, this is not problematic as comparisons were made only between groups, instead of between AAOIs within each group.

The first dependent variable was the *mean viewing duration per video*. The second dependent variable was the *distribution of gazes* across the four different AAOIs for the first 4 s. This was done to make total gaze durations comparable, because participants could watch each video as long as they wanted. Every participant had watched each video for at least this amount of time (i.e., minimum viewing time was 4.15 s). Third, *sequence analyses* for the different AAOIs were conducted to compare series of gazes to each other based on the so-called Levenshtein distance. The Levenshtein distance is a measure used in computer science to assess the edit distance between two strings. This edit distance is the minimal number of edit operations needed to transfer one string into another. Edit operations are insertions, deletions, or substitutions of single characters. The edit distance can be converted into a percentage value. In the present study the Levenshtein distance was used to compare the sequences of the gaze locations across participants (cf. Feusner & Lukoff, 2008).

Table 2

Overview of relevant and irrelevant parts of the presented fish according to the locomotion description strategy and the species classification strategy for all four videos.

	Relevant	Irrelevant	Technical term
Locomotion description strategy			
Video 1	Dorsal and anal fin	Pectoral, caudal fin, and eye	Tetraodonti
Video 2	Caudal fin	Dorsal, anal, pectoral fin, and stripes	Subcarangi
Video 3	Pectoral fin	Dorsal, anal, caudal, pectoral fin, and eye	Labri
Video 4	Dorsal and anal fin	Caudal, pectoral fin, eye, and taint	Balisti
Species classification strategy			
Video 1	Dorsal, anal, pectoral fin, and eye	Caudal fin	Arothron hispidus
Video 2	Stripes	Caudal, dorsal, anal, and pectoral fin	Anisotremus virginicus
Video 3	Pectoral fin	Dorsal, anal, caudal, pectoral fin, and eye	Thalassoma lunare (female)
Video 4	Taint	Dorsal, anal, pectoral, caudal fin, and eye	Cantherines macrocerus

To determine the Levenshtein distance, a string of AAOIs that a participant had looked at in a given order (e.g., AAOI 1, AAOI 3, AAOI 1, AAOI 4, ...) was the input data. The number of edit operations (i.e., insertions or deletions of AAOIs) needed to transform the AAOI sequence of this participant into that of another participant describes the similarity between the two sequences of AAOIs. The Levenshtein distance was determined for the gaze sequences of experts as well as for novices to analyze the similarity of the strategies used within groups. In particular, Levenshtein distances were computed for pairs of AAOI strings, whereby each person provided one AAOI string. This procedure resulted in a similarity score for each possible pair of experts and for each possible pair of novices. The calculation of the distribution of gazes and sequence analysis required self-programmed tools, which are described in the [Appendix A](#).

2.4.3. Cued retrospective reporting

The cued retrospective reports obtained during the gaze replay were analyzed with regard to different aspects. First, the *contents* of the participants' initial utterances (i.e., the first term mentioned) of each gaze replay were analyzed to determine whether they referred to either relevant or irrelevant features for both classification strategies (see [Table 2](#)) in order to investigate participants' initial response to the task. The remaining analyses referred to the complete gaze replay. As a second variable, it was analyzed whether the cued retrospective reporting referred to *features* that were either relevant according to one or both classification strategies, or irrelevant according to both strategies (see [Table 2](#)). Third, the *total number of words* used during retrospection was counted.

Fourth, the *number of different technical terms*, which were used during retrospection, was determined.

3. Results

An alpha level of .05 was used for the statistical tests reported.

3.1. Correctness of description

Results of a 2 (expertise) \times 3 (performance) MANOVA, in which the three performance measures were naming locomotion pattern correctly, describing which body part had been moving, and describing how each body part had been moving, showed a main effect of expertise, Pillai's trace = .58, $F(1, 18) = 7.32$, $p = .003$, partial $\eta^2 = .58$, meaning that expertise positively affected the overall correctness of the description of the locomotion pattern. The univariate analyses showed that experts mentioned some technical terms for locomotion pattern ($M = 1.57$, $SD = 1.27$), whereas none of the novices did ($M = .00$, $SD = .00$), $F(1, 18) = 20.82$, $p < .001$, partial $\eta^2 = .54$. All experts were able to identify the body part relevant to the locomotion pattern ($M = 4.00$, $SD = .00$), and although the novices' performance was also quite good in this respect ($M = 3.08$, $SD = 1.22$), it did not reach that of experts, $F(1, 18) = 3.89$, $p = .06$, partial $\eta^2 = .18$. Experts and novices did not differ in the way they described how the relevant parts of the fish body moved (for experts $M = 1.14$, $SD = 1.03$; for novices $M = .92$, $SD = .86$), $F < 1$, ns.

3.2. Gaze data

Novices ($M = 24.69$ s, $SD = 12.15$) had significantly longer mean viewing times for the four videos than experts ($M = 10.93$ s, $SD = 4.06$), $F(1, 19) = 8.32$, $p = .009$, partial $\eta^2 = .31$.

Regarding the distribution of gaze durations for the four different AAOI types for the first 4 s of each video, an ANOVA with expertise as independent variable and the gaze duration for each of the four AAOI types as dependent variable was conducted. The ANOVA revealed that experts had significantly longer gaze durations than novices on the AAOI relevant for the species classification strategy, $F(1, 19) = 7.69$, $p = .01$, partial $\eta^2 = .29$, and marginally longer gaze durations on the AAOI relevant for both strategies, $F(1, 19) = 3.10$, $p = .095$, partial $\eta^2 = .14$. Gaze durations on the AAOI relevant for locomotion pattern description did not differ between groups, $F < 1$, ns. Finally, novices had longer gaze durations on the AAOI irrelevant for each strategy compared to experts, $F(1, 19) = 6.53$, $p = .02$, partial $\eta^2 = .26$. Means and standard deviations are displayed in [Table 3](#).

3.2.1. Sequence analysis

An ANOVA was conducted with the similarity scores as dependent variable and the type of comparison (comparisons among experts vs. comparisons among novices) as the independent variable. A significant main effect of type of

Table 3
Means (and SD) in milliseconds for the four AAOIs as a function of expertise level per video.

AAOI	Experts	Novices
Distribution of gazes for the first 4 s of each video		
Relevant for species classification	375.00 (234.33)	160.36 (124.60)
Relevant for locomotion pattern description	85.71 (93.02)	122.50 (143.56)
Relevant for both strategies	268.57 (47.85)	204.27 (89.72)
Irrelevant for both strategies	2054.30 (280.53)	2336.80 (216.85)

comparison was found on Levenshtein distance, $F(1, 107) = 9.28$, $p = .003$, partial $\eta^2 = .08$. The group of experts ($M = 67.41\%$; $SD = 5.62\%$) was less similar than the novices group ($M = 72.09\%$; $SD = 6.49\%$), that is, the gaze behavior of experts was more heterogeneous than that of novices.

3.3. Cued retrospective reporting

First, the initial utterances during watching the gaze replay were analyzed by conducting ANOVAs with expertise as independent variable. Experts mentioned contents relevant for classifying the fish species more often ($M = 1.57$, $SD = 1.27$) than novices ($M = .21$, $SD = .43$), $F(1, 19) = 13.53$, $p = .002$, partial $\eta^2 = .42$. However, the experts did not differ from novices in the case of mentioning content relevant for classifying the locomotion pattern, $F < 1$, ns, or contents relevant for both classification strategies, $F(1, 19) = 3.07$, $p = .10$, or of irrelevant content, $F < 1$, ns. Means and standard deviations are displayed in Table 4.

Second, ANOVAs, with expertise as independent variable, were conducted in order to investigate utterances of the participants that occurred during the whole task performance. Experts mentioned features that were relevant to the locomotion description strategy significantly more often than novices, $F(1, 19) = 23.84$, $p < .001$, partial $\eta^2 = .56$, as well as features that were relevant according to the species

classification strategy, $F(1, 19) = 9.11$, $p = .007$, partial $\eta^2 = .32$. However, there were no significant differences for features relevant to both strategies, $F(1, 19) = 1.49$, $p = .24$, and irrelevant statements, $F < 1$, ns. Means and standard deviations are displayed in Table 4.

Third, an ANOVA for the overall number of words used during retrospection, with expertise as independent variable, revealed no differences between experts ($M = 242.14$, $SD = 110.54$) and novices ($M = 247.36$, $SD = 151.44$), $F < 1$, ns. Finally, a similar ANOVA for the complete gaze replay showed that experts tended to use more technical terms during retrospection ($M = 16.29$, $SD = 4.75$) than novices ($M = 11.14$, $SD = 5.78$), $F(1, 19) = 4.12$, $p = .06$, partial $\eta^2 = .18$.

4. Discussion

The present study aimed at identifying expertise effects in perceiving and interpreting complex, dynamic visual stimuli, both at an outcome and a process level. It was hypothesized that experts would perform more accurately and faster than novices (Hypothesis 1) and that the process data of experts would show that they attended more to relevant information than novices, who would attend more to irrelevant information (Hypothesis 2). Furthermore, Hypothesis 3 addressed the issue of which relevant features the experts would use (i.e., for locomotion description strategy or species classification strategy, or both). Hypothesis 4 addressed the issue of whether expertise would yield either more diversity or more homogeneity in terms of the perceptual strategies used. Finally, Hypothesis 5 predicted that experts would verbalize less information than novices and thus use fewer words, because they use more encapsulating technical terms during retrospection.

In line with Hypothesis 1, performance differences in favor of experts were verified. Experts were able to perform the task faster, as indicated by shorter mean viewing times of the stimulus, and more accurately, as indicated by their better description of locomotion patterns and their higher use of correct technical terms. As expected, novices were not able to do so. Not only did they not use technical terms, but they also were not able to identify the parts of the fish body relevant for the locomotion pattern displayed. While both experts and novices refrained from appropriately describing how the crucial body parts moved, experts named the correct technical term for the locomotion pattern, which by definition encapsulates this information.

The main focus of this study, however, was to investigate process differences between experts and novices both at a conceptual level, by means of cued retrospective reporting, and at a perceptual level, by means of eye tracking. These processes were analyzed with regard to whether experts would attend more to relevant rather than irrelevant features compared to novices during the task as indicated by Hypothesis 2. The comparison of initial utterances of the cued retrospective reports between experts and novices showed that experts considered more of the relevant information in the

Table 4
Means (and SD) in percent for the four content types as a function of expertise per video in the two time slots.

Content type	Experts	Novices
Initial utterance		
Relevant for species classification	39.00 (31.75)	5.25 (10.75)
Relevant for locomotion pattern description	21.50 (22.50)	28.50 (23.75)
Relevant for both strategies	14.25 (19.75)	28.50 (16.50)
Irrelevant for both strategies	7.25 (12.25)	12.50 (16.25)
Utterances for entire task performance		
Relevant for species classification	57.25 (23.75)	26.75 (20.75)
Relevant for locomotion pattern description	78.50 (17.25)	50.00 (9.75)
Relevant for both strategies	50.00 (0.00)	42.75 (15.25)
Irrelevant for both strategies	60.75 (43.00)	75.00 (26.00)

Values do not add up to 100%, because off-topic utterances were not taken into account.

beginning of their task processing. Furthermore, analyzing the distribution of gaze durations on AAOIs revealed that experts also attended more to relevant areas compared to novices. Finally, the cued retrospective reports for the entire duration of task accomplishment showed that the experts' attention remained focused on relevant areas. Thus, the findings of the present study show, in line with results from studies on static stimuli (Antes & Kristjanson, 1991; Charness et al., 2001; Haider & Frensch, 1999), that experts attend more relevant features of a complex dynamic stimulus than novices. Therefore Hypothesis 2 was confirmed.

Moreover, the results showed that experts did not primarily focus on features that were crucial for identifying a locomotion pattern. Instead they mainly concentrated on features that allowed identifying the fish species; these features were not related to a locomotion pattern at all. In line with Boshuizen and Schmidt (1992) this finding indicates that experts use knowledge-based shortcuts, like using the features associated with a fish species, activating the appropriate schema, which also contains the knowledge on the locomotion pattern. Evidence that experts used this strategy can be found at both the conceptual and the perceptual level. For instance, experts often began their cued retrospective reports with mentioning the fish species and had longer gaze durations on AAOI relevant for species classification compared to novices. These findings confirmed Hypothesis 3.

Furthermore, the results from cued retrospective reporting revealed that although experts used indeed more technical terms, which encapsulate their knowledge (cf. Boshuizen & Schmidt, 1992), they did not use less words than novices, contrary to the assumptions of Ericsson and Simon (1980) and Hypothesis 5. This finding may be due to the fact that novices also verbalize little because they lack knowledge. However, it might also be due to the use of cues for retrospective reports. Cued retrospective reports seem to be a suitable method to enable experts to inspect their cognitive processes during task performance (Van Gog et al., 2005). Hence, cueing reports by means of gaze replays may be suited to overcome the potential drawbacks of retrospective compared to concurrent data on strategy use that have been noted by other researchers (Brinkman, 1993).

Finally, the similarity of sequences of gaze allocations was investigated at a perceptual level. The results showed that experts had a more diverse gaze pattern compared to novices. This finding is in line with Hypothesis 4 and with the results of Medin et al. (2006) on experts classifying fish species. There are several possible explanations for this finding: It might either be that experts act upon their individual case-based knowledge rather than upon a more generic knowledge base that many experts share. Or it might be that experts use diverse strategies, while novices use no strategies at all and thus their gazes converge around a more neutral pattern that might be, for instance, influenced by salience. Indeed, it has to be noted that the group of experts was rather diverse (i.e., advanced PhD students as well as professors), which might have led to a diversity in strategies resulting in a high variability of the perceptual patterns.

4.1. Limitations and implications of the study

The present study has at least two potential limitations. The first is the small number of participants, which is, however, not uncommon in expertise research as experts are scarce. There is a relatively small number of specialists in marine zoology. Still, because of the small sample size, care has to be taken in interpreting the results from this study. The second potential limitation concerns the stimulus material itself. The videos were chosen with the help of a domain expert in such a way that all features needed in order to classify the particular locomotion pattern were clearly visible. However, because the fish moved, the relevant areas of a fish were not always entirely visible as they were sometimes partly concealed for a short time due to the fact that the fish had changed its relative position to the camera. These facts might lead to generalization problems.

Nonetheless, the present study suggests some interesting implications for teaching novices perceptual and conceptual strategies of experts. Novices need to acquire factual knowledge that enables them to express their observations correctly. However, to gain this factual knowledge, they need to know what to observe. Therefore, since experts pay more attention to relevant features of the locomotion pattern at a conceptual level as well as at a perceptual level, novices may benefit from instruction that contains attentional guidance in order to recognize relevant body parts involved in the locomotion patterns. It has been shown that thought processes (conceptual strategies) of experts can be taught directly (Van Gog, Paas, et al., 2008). Thus, a form of cueing (see Boucheix & Lowe, 2010; De Koning, Tabbers, Rikers, & Paas, 2010) by guiding novices' attention based on experts' eye movements (cf. Van Gog, 2009) might also be effective, especially for highly visual tasks dealing with complex dynamic stimuli.

The results from the present study provide first hints on how such attentional guidance could be designed in order to be useful for novices. First, expert–novice differences in cued retrospective reporting show that experts do not have difficulties in verbalizing their reasoning with this reporting technique (a drawback that concurrent reporting might have). Thus, it is possible to acquire experts' eye movements as well as cued retrospective verbalizations for instructional purposes, that is, the experts can act as a model for a successful strategy for novices, in particular when experts would be instructed to explain all technical terms used.

Second, the findings indicate that each expert might have her or his own understanding of what is relevant. For instance, most experts focused on features that were relevant for the species classification strategy rather than on features that were relevant for the locomotion description strategy. Thus, findings from the present study lead to the assumption that it might be more meaningful to use the perceptual processes from one carefully chosen expert for instructional purposes instead of averaging across several experts. They also suggest that a direct use of experts' process data for instructional purposes might not always be useful for novices. Rather, to be useful for novices, experts' should model the locomotion description

strategy as it is taught. Thus, additional instructions for experts to “behave in a more didactic manner” are probably needed.

In sum, the present study has provided some insight into expertise differences concerning the perception and interpretation of complex, dynamic visual stimuli, also in terms of the underlying processes of task performance, and seems to hold promising implications for instructional design in this area.

Acknowledgements

This work is part of a larger research project on “Resource-adaptive design of visualizations for supporting the comprehension of complex dynamics in the Natural Sciences” funded by the Leibniz-Gemeinschaft.

During the realization of this work Tamara Van Gog was supported by a Rubicon grant from the Netherlands Organization for Scientific Research (NWO; Grant nr. 446-07-001).

Appendix A. Additional tools programmed for the analysis of the eye tracking data

Defining “dynamic” AAOIs

The computer tool that enables the transformation of AOIs to “dynamic” AAOIs was programmed in Java (<http://java.sun.com>). The input files are ClearView 2.7.0 output. TXT files. The tool first calculates the dwell time for each AOI across all 100 ms segments per video. It then aggregates the AOIs into AAOIs (relevant for locomotion, species, both, or irrelevant) across all four videos. In addition, the tool generates a sequence of the AAOIs for each video, which serves as input for the sequence analysis tool. All output of the tool is given in .txt files.

Sequence analysis

The computer tool that was used for the sequence analysis was programmed in Ruby (<http://www.ruby-lang.org>). The input data files are in TXT file types, which were converted in CSV file types. The output data files are in TXT file types. The program calculates the edit distance and the similarity percentage between two strings according to Levenshtein. The program can be assessed at the URL code.google.com/p/eye-tracker-tools.

References

Antes, J. R., & Kristjanson, A. F. (1991). Discriminating artists from nonartists by their eye-fixation patterns. *Perceptual and Motor Skills*, 73, 893–894.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates, and novices. *Cognitive Science*, 16, 153–184.

Boucheix, J.-M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, 20(2), 123–135.

Brinkman, J. A. (1993). Verbal protocol accuracy in fault diagnosis. *Ergonomics*, 36, 1381–1397.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20(2), 155–166.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: evidence from eye movements. *Memory and Cognition*, 29, 1146–1152.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Review Neuroscience*, 3, 201–215.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: seeing is understanding? *Learning and Instruction*, 20(2), 111–122.

Ellis, S. R. (1986). Statistical dependency in visual scanning. *Human Factors*, 28, 421–438.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Reviews in Psychology*, 47, 273–305.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251.

Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review*, 19, 91–110.

Feusner, M., & Lukoff, B. (2008). Testing for statistically significant differences between groups of scan patterns. In S. N. Spencer (Ed.), *Proceedings of the 2008 symposium on eye tracking research & applications* (pp. 43–46). New York: ACM.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: more evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 172–190.

Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson, & S. M. Kosslyn (Eds.), *Tutorials in learning and memory* (pp. 31–60). San Francisco, CA: Freeman.

Lesgold, A., Robinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

Lindsey, C. C. (1978). Form, function, and locomotory habits in fish. *Fish Physiology*, 7, 1–88.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education*, 14, 225–244.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology*, 32, 49–96.

Medin, D. L., Ross, N., Atran, S., et al. (2006). Folkbiology of freshwater fish. *Cognition*, 99, 237–273.

Moreno, F. J., Reina, R., Luis, V., & Sabido, R. (2002). Visual search strategies in experienced and inexperienced gymnastic coaches. *Perceptual and Motor Skills*, 95, 901–902.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296.

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model’s eye movements. *Computers in Human Behavior*, 25, 785–791.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, 18, 211–222.

Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11, 237–244.

Videler, J. J. (1993). *Fish swimming*. New York: Chapman & Hall.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.