Journal of Expertise
2026. Vol. 9(1)

ISSN 2573-2773

# Visual Expertise in History: Historians' and Novices' Visual Processing and Interpretation of Historical Images

Marjaana Puurtinen[1], Johanna K. Kaakinen[2], and Halszka Jarodzka[3]
[1]Department of Teacher Education, University of Turku, Finland
[2]Department of Psychology and Speech-Language Pathology, University of Turku, Finland
[3]Research and Innovation Center for Open Distance Education, Open University of the Netherlands, the Netherlands

Correspondence: Marjaana Puurtinen, marjaana.puurtinen@utu.fi

## Abstract
Drawing on expertise research and scene perception studies, this study investigates the characteristics of visual expertise in the domain of history by examining historians' and novices' visual processing and task performance in an image integration task. Task complexity was manipulated by altering the ease of identifying the topic depicted in a series of historical images. Historians ($n$ = 40) and novices ($n$ = 40) were shown a series of black-and-white images. Their eye movements were recorded, and each participant identified a topic for each series. Their written responses were coded for appropriateness and whether they described the topic as part of a historical process. The results showed that compared to novices, historians exhibited more consistent gaze behavior for each series of images, indicating a systematic approach to interpreting them. Furthermore, novices' fixation durations increased, and the number of fixations tended to decrease from the first to the final image, suggesting that their gaze became more focused only for the later images. Expertise also affected task performance: In the case of a more challenging set of images, historians provided more appropriate topic answers and interpreted the topics as part of a historical process, indicating that they were able to situate even the more complex series within a meaningful bigger picture. Thus, even in the ill-structured domain of history, domain expertise induces differences in the visual processing and interpretation of domain-related images compared to lack of such expertise.

## Keywords
expertise, eye tracking, history, interpretation, scene perception

## Introduction

Although history is often considered a text-based discipline, it is, to a large extent, also visual. In professional historical studies, studying the past through pictures has an enduring and increasingly visible role (e.g., Bleichmar & Schwartz, 2019). Novices, in their turn, may binge-watch historical TV dramas, play digital games inspired by past worlds, and share witty memes that refer to historical events and characters. Despite the everyday presence of historical visual representations, as well as their significance for historical research, empirical evidence on the effect of history expertise on visual processing and interpretation of these representations is extremely limited. Therefore, to enhance our understanding of the characteristics of expertise in this relatively unexplored domain, we investigated historians' and novices' visual processing and interpretation of a series of historical images, basing our work on expertise theories and research on scene perception, with eye tracking employed as our key methodology.

## Expert–Novice Differences in the History Domain

Experts are individuals with a large amount of prior knowledge and experience in a specific domain, who repeatedly excel in performing tasks representative of their area of expertise compared to less experienced individuals (Ericsson & Smith, 1991; Ericsson & Lehmann, 1996). In the domain of history, however, research on expertise has had a rather limited empirical basis (for a review of early work, see Voss & Wiley, 2006; for more recent reviews, see, Voet & De Wever, 2017; Kainulainen, et al., 2025). The field is also somewhat scattered, and empirical studies about historians' domain skills and reasoning have been published in the fields of cognitive or educational psychology (e.g., Wineburg, 1991; 1998; Leinhardt & Young, 1996; Rouet et al., 1997; Baron, 2012; Greene & Yu, 2014; Shreiner, 2014; McGrew, 2022; Kainulainen, et al., 2025), history education (e.g., Wineburg, 1994; Leinhardt et al., 1994, Nye et al., 2009, Monte-Sano, 2017; Monte-Sano & Thomson, 2022; Kainulainen, et al., 2022a), teaching, learning or teacher education (e.g., Gottlieb & Wineburg, 2012; Lévesque et al., 2014; Schneider & Zakai, 2016; Goulding, 2021; Kainulainen et al., 2022b) and even information sciences and literacy research (e.g., Cole, 1998; 2000; Shanahan, et al., 2011), to name a few examples.

Nonetheless, all of the above studies build on the assumption that the key task of an expert historian is to make source-based interpretations about the past (Rouet et al., 1997). In this regard, historians differ from domain novices in two key ways. First, drawing on their education and years of practice, historians tend to view their domain as that involving interpretation, conceptualizing history as human production as opposed to the factual past (Greene & Yu, 2014; see also Kainulainen et al., 2019). In other words, expert historians view historical sources as objects that need to be interpreted and understood, focusing on relationships over historical facts (Greene & Yu, 2014). In contrast, novices typically consider historical knowledge to be established facts and treat sources as bits of information that have to be pieced together (Wineburg, 1991; Greene & Yu, 2014; Lévesque et al., 2014). Second, historians treat historical questions as open-ended, using multiple interpretations to ask questions about what might have actually taken place, while novices prefer choosing one *correct* or *better* alternative among multiple historical interpretations or sources (Wineburg, 1991; Rouet et al., 1997; Goulding, 2021). Similar differences between historians and novices have also been observed in cases where the participating historians were not specialists of the specific historical topic being discussed but applied their general domain knowledge to solve the given tasks (Wineburg, 1991; 1998; Shreiner, 2014). Thus, even though expertise is typically considered highly domain-, subdomain-, or even task-specific (Ericsson & Kintsch, 1995), historians have outperformed novices even in source analysis tasks outside their specific area of expertise.

In their seminal study, Wineburg (1991) invited eight historians to think aloud while reading historical materials. They concluded that historians' source analysis includes the use of the heuristics of corroboration (i.e., comparing documents with others), sourcing (i.e., considering the origins of the source), and contextualization (i.e., situating the document in a spatial and temporal context). These heuristics have since been complemented and detailed with additional cognitive tools and strategies relevant for historical analysis. Historians may, for instance, turn to rereading or close reading, critique the material, or consider text structure or their personal interest toward the topic (Shanahan, et al., 2011). They may also reflect on their virtues and affects, or apply broad approaches or specific methodologies that all impact their interpretation (Kainulainen, et al., 2025). Prior studies have, however, typically focused on historians' reading of textual sources and evidence about expert-like interpreting of historical images is scarce. It still appears that while novices tend to rely on descriptive analysis of an image, historians are able to evaluate images, too, through the means of corroboration, sourcing and contextualization (e.g., Wineburg, 1991; Lévesque et al., 2014). These heuristics may, however, need to be complemented or modified according to the type of inspected visual material (Goulding, 2021). Historians may also express a more critical attitude toward a visual source, compared to non-historians (Gottlieb & Wineburg, 2012; Lévesque et al., 2014).

In summary, historians possess skills that enable them to engage with historical materials in manifold ways (Leinhardt & Young, 1996; Kainulainen, et al., 2025). However, most of the empirical evidence thus far has been gained through qualitative studies and self-report protocols, such as think-alouds or interviews, leaving visual processing of the image uncaptured (for the few exceptions, see the following section). Thus, to expand our understanding of visual expertise in history, we study both the visual inspection and interpretations of historical images of large groups of historians and novices and use eye tracking as our key method. Our purpose is to explore whether the effects of history expertise emerge at the visual processing or the interpretation level and what forms these effects take. Due to the scarcity of closely related empirical work, we formulate our hypotheses based on expertise research on visual-task behavior conducted in other domains, as well as on general views pertaining to human scene perception.

## Seeing History in Images: Integrating Expertise and Scene Perception Research

The domain knowledge of experts is generally highly organized, providing them fast and efficient access to task-relevant information (Ericsson & Kintsch, 1995). Their organized memory structures, including schemata-type knowledge, allow them to rapidly encode and retrieve relevant information from long-term memory, as well as categorize new information both quickly and efficiently (VanLehn, 1989; about the history domain, see Leinhardt & Young, 1996). This also holds for visual tasks, in which experts' efficient use of their cognitive resources leads to purposeful and successful visual processing and task performance across a range of domains, compared to domain novices (e.g., Gegenfurtner, et al., 2024). For instance, eye-tracking studies have demonstrated how domain experts visually study domain-relevant materials more efficiently than novices (Gegenfurtner et al., 2011; Sheridan & Reingold, 2017). It has also been established that experts can interpret, classify, diagnose, evaluate, or judge visualizations related to their own domain in ways that novices are unable to (Gegenfurtner, et al., 2011).

Only a few studies have examined history experts' visual processing. The existing evidence suggests that historical domain knowledge may support the simple encoding and recognition of history-related images (Humphrey & Underwood, 2009; Underwood et al., 2009), and that descriptive differences can be identified between a novice, intermediate, and expert historian's gaze patterns when studying a historical image (Lévesque et al., 2014). Considering the scarcity of relevant evidence, we needed to build the hypotheses for our experiment by drawing on research conducted in other areas of expertise, as well as on insights from scene perception research.

We begin by addressing how experts visually process materials pertaining to their domain. In general, compared to intermediates and novices, experts demonstrate a larger number of short duration fixations on task-relevant areas (for reviews, see Gegenfurtner et al., 2011; Jarodzka et al., 2017; see also Gegenfurtner, et al., 2020). Experts can also use their parafoveal or peripheral vision, i.e., holistic perception, to process large portions of an image, benefiting even from a rapid initial glimpse (Kundel et al., 2007; Sheridan & Reingold, 2017; see also Gegenfurtner et al., 2011). This expert characteristic facilitates an efficient visual strategy—if the visual task so requires, experts can first apply their holistic perception before turning to a search-to-find strategy. In contrast, less-experienced viewers begin with the search-to-find approach, resulting in slower and less accurate task performance (Kundel et al., 2007).

However, we argue that, when examining how experts and novices respond to images, the interplay of top-down and bottom-up factors in gaze control must also be accounted for. This factor has been extensively examined in scene perception research—a field of cognitive psychology focusing on how one perceives and makes sense of a complex scene, or “a pictorial view of the environment” (Foulsham, 2015, p. 257). Although this research is often overlooked in studies about expert–novice differences in complex and real-life visual tasks, we see it highly informative and therefore apply it in crafting our hypotheses on characteristics of visual expertise in history.

In practice, both the semantic relevance and visual saliency of the elements within an image play significant roles in how its visual information is inspected (e.g., Borji & Itti, 2014; Loftus & Mackworth, 1978; Rayner, 2009). While visual saliency determines the early capture of attention, it is then surpassed by object-based purpose to the inspection (Underwood et al., 2009; Drew et al., 2013), and top-down control takes over. When investigating top-down gaze control, Henderson et al. (2007) identified three factors that control one’s gaze when inspecting images. The first factor is knowledge of the inspected image, which is gained during the initial encounter or across multiple encounters with the same image. Thus, both the working memory representation of a specific image and its image-relevant long-term memory representation guide visual attention during image inspection (Woodman & Luck, 2015). The second factor is the task being performed by the viewer. For instance, gazes are distributed wider and target more of the objects within an image during a memorization task than during a visual search (Castelhano et al., 2009). Since viewers tend to focus on those areas of images that provide them task-relevant visual information (e.g., DeAngelus & Pelz, 2009; Tatler et al., 2010; Borji & Itti, 2014), the increased cognitive demand of a visual task may lead to restricted visual study of an image, at least in terms of the number of fixations (Walter & Bex, 2021). The third top-down factor that affects gaze control—one that most clearly ties scene perception studies to expertise research—is the viewer’s scene schema knowledge. This factor also influences how gaze is distributed across an image. Large amounts of prior knowledge are typically stored in one’s long-term memory in the form of schemata (Rumelhart, 1980). When inspecting an image, an activated schema quickly guides visual processing. While the first fixations made on an image is often guided by visual saliency, already the second fixation is more likely to target an element based on interest (Kaakinen et al., 2011). An expert usually has a wide range of appropriate schemata available for the efficient interpretation of image content, context, and meaning. Schema activation should, in turn, guide visual processing and aid in finding task-relevant image features (e.g., Kundel & Nodine, 1983).

Schema knowledge is also important in the domain of history. For instance, when tasked with encoding and recognizing black-and-white photographs, history and engineering specialists identified images related to their own domain more accurately and, based on just 3-second viewings of images, made fewer overall fixations when looking at stimuli from their domain (Humphrey & Underwood, 2009; Underwood et al., 2009). In addition, during both encoding and recognition, saliency had less of an effect on the domain specialists’ scanpaths when they viewed domain-specific stimuli (Humphrey & Underwood, 2009). Thus, domain knowledge results in the rapid activation of domain-related schemata, overriding the effects of visual saliency on gaze control.

As mentioned above, domain experts differ from novices not only in the visual processing of domain-specific material, but they also outperform novices in related tasks (Gegenfurtner et al., 2011; Sheridan & Reingold, 2017). In view of this, one can expect that the ways in which experts and novices respond to variations in *task complexity* should differ as well—in most cases, experts should be able to apply their task and schema knowledge to optimize their visual processing according to the given task requirements (see Henderson et al., 2007). For instance, in radiology, experts exhibited more sensitivity to visual materials and task requirements, altered their visual strategies accordingly, and showed less variability in gaze behavior compared to novices (van der Gijp et al., 2017).

In general, image items that are semantically consistent with an assumed image category are usually identified faster and more precisely, whereas inconsistent image items are looked at for a longer time (Võ, 2021). Thus, when one image activates a schema, encountering a new image with items that are inconsistent with the schema should lengthen fixation durations. In addition, experts are generally more efficient in their processing of domain-specific visual materials; thus, in the case of inconsistencies with new image items and their active schema, they should be able to activate new and better-fitting schemata with relative ease, owing to their extended and efficient expert memory (as in van der Gijp et al., 2017). Novices, however, should be more perplexed by challenges in identifying semantic consistency.

However, it must be noted that the nature of the benefits arising from expertise in visual tasks is not always straightforward. On some occasions, experts can outperform novices even when there are no notable differences in the visual processing patterns of the groups (for an example in medicine, see Kok et al., 2012). On other occasions, experts and novices may show similar task performance even when their visual processing differed due to task complexity (for an example in music reading, see Penttinen et al., 2015). Finally, experts' schemata-driven processing of visual stimuli may also result in the so-called proof-readers error, where errors in stimuli are misread as "correct" instead (for an example in music reading, see Wolf, 1976). Thus, it appears that task characteristics may modify if and how expertise benefits gaze control and/or task performance during visual tasks—we need to note, however, that evidence from the domain of history is scarce.

To summarize, all three top-down factors of gaze control—image knowledge, task knowledge, and relevant schema knowledge—are closely connected, not just to each other, but also to task performance. This is perfectly captured in Võ's (2021) broadened conception of schemata, described as encompassing viewer's domain-specific schemata about the types of images, task requirements, what the images represent, and how they are created and used. In all of these respects, domain experts should be in an advantageous position compared to novices due to their efficient use of expert memory. In this study, we build on this general idea to explore if and how history expertise is manifested in a task involving the processing and interpretation of historical images.

## Research Questions

In our experiment, we employed a quasi-experimental task, in which historians and novices were invited to view three historical images, presented one by one, and group them according to what they considered an appropriate topic (see Puurtinen et al., 2015). The following research questions (RQs) are addressed in this study:

### RQ1. How do domain knowledge and task complexity affect the visual processing of a series of historical images?

We hypothesize that owing to their ability to quickly activate appropriate schemata through the efficient use of expert memory, as well as effectively employ holistic perception, historians generally exhibit shorter fixation durations and inspect a smaller portion of domain-specific images compared to novices. We also expect to identify the effects of expertise on fixation frequency most prominently in the first scene, assuming the experts are able to quickly activate an appropriate schema. After identification of the initial scene and activation of related schemata, we expect experts to demonstrate more strategic gaze behavior as the series proceeds. We also assume that an increase in task complexity, which implies an increase in cognitive demand, may result in fewer and longer fixations, and that experts should be able to overcome potential challenges more efficiently than novices.

### RQ2. How do domain knowledge and task complexity affect task performance in interpreting a series of historical images?

We hypothesize that due to their ability to quickly activate appropriate schemata through the efficient use of expert memory, historians are more capable of producing appropriate domain-relevant interpretations of a series of historical images compared to novices, despite only a rapid view of a series of images and irrespective of task complexity.

## Methods

### Participants

A total of 92 adult Finnish-speaking volunteers participated in this study (*MEDIAN* = 38 years of age, *MIN.* = 19 years of age, *MAX.* = 75 years of age; 58 female, 28 male, and three other or not specified; information from three participants was missing). After excluding three participants with missing background and task-response data, one participant who recognized having seen some task stimuli before, and eight with poor eye-tracking data, the data obtained from 80 participants were ultimately included in the final analyses.

To recruit participants with domain knowledge, we contacted universities and cultural history museums at five municipalities. Subsequently, we (a) delivered presentations about the study at staff meetings, seminars, or lectures targeting master's degree students, and/or (b) sent email announcements through staff and student mailing lists. In addition, an online announcement was posted on a national humanities web forum. University and museum staff could participate in our study during working hours, and history students could get course credit for participation. For recruiting history novices, open-to-all data collections were organized at the Medieval Turku Event and at Luostarinmäki Museum, both located in the city of Turku, Finland. The opportunity to participate in the data collection was advertised on Medieval Turku and Luostarinmäki websites and on social media. In all announcements, the study was advertised as pertaining to viewing and interpretation of historical images.

Participants were divided into two groups based on their self-reported level of education and whether this education involved a history-related discipline. Of the 80 participants, *historians* (*n* = 40) had a university degree (bachelor, master's, or PhD) in a history-related discipline, whereas *novices* (*n* = 40) reported not having a degree in history. These group sizes were considered adequate with respect to the effort invested in recruitment, the data collection taking place after the COVID-19 pandemic, and the general availability of representatives, especially of the domain-knowledge group. No financial compensation was offered to study participants. All participants gave their informed consent, and the study adhered to the guidelines of the Finnish National Board on Research Integrity. The research plan was approved by the Ethics Committee for Human Sciences at the University of Turku (decision 42/2021).

### Visual Task and Stimuli

We adapted the image-integration task and stimuli developed by Puurtinen et al. (2015). A participant was briefly shown a series of three historical images (see Figure 1) presented one by one and invited to answer the question: "What do you think the series is about?" This open-ended task was equally approachable to both historians and novices and left room for different interpretations. As for professional historians, the task might resemble a situation when they, for instance, scan through digital archives in search of images for research or teaching purposes.

We employed two image series created and piloted in the original study, titled "Female Workers" (FW)[1] and "Doctors and Infants" (DI), and we created a new series titled "School Children" (SC; see Figure 1), as well as a practice series. Contents of images in each series represented changes across time in terms of, e.g., environments, tasks, or outfits (from oldest/least modern to newest/most modern), and the presentation order within one series was therefore kept constant. Each image was shown to the participants for 4 s. The 4 s duration was selected based on Puurtinen et al. (2015), whose history-student participants responded after 2 s and 8 s viewings (see also Kundel & Nodine, 1983; Humphrey & Underwood, 2009; Underwood et al., 2009).

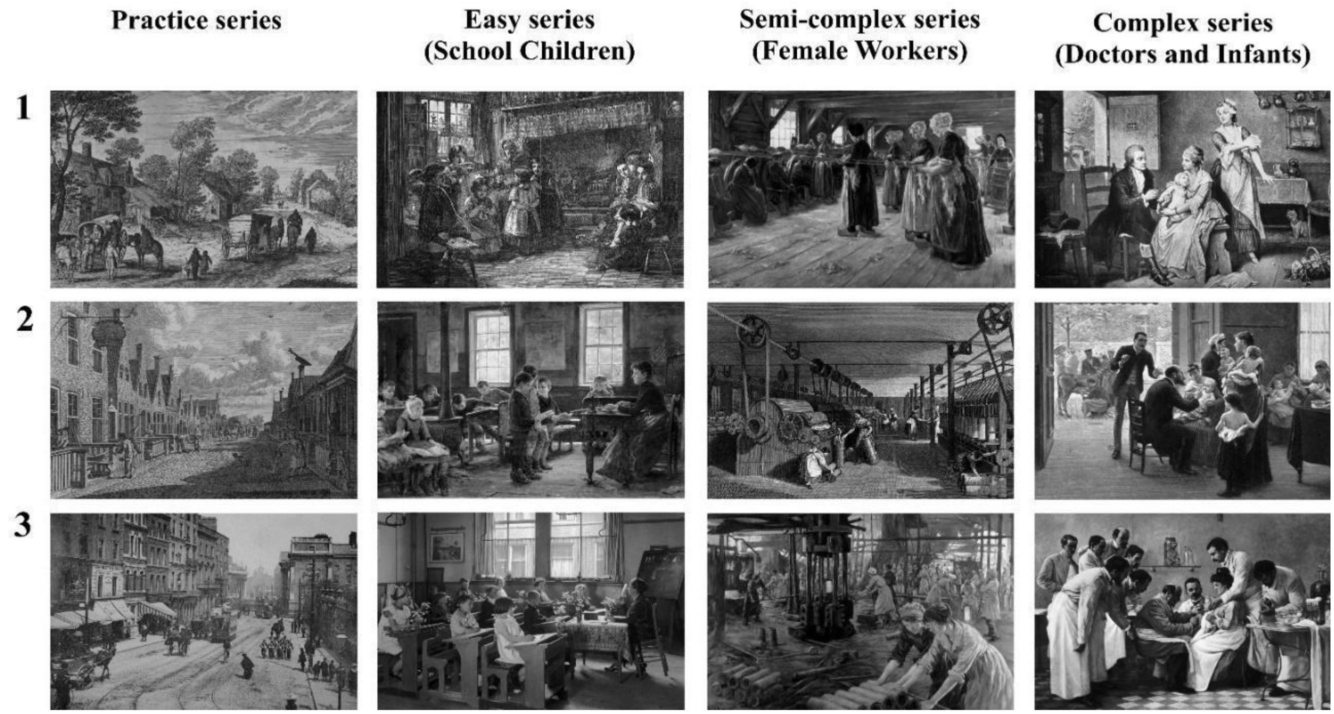


**Figure 1.** Image Series Used in the Experiment

**Note.** All images are modified versions of the originals. See Appendix A for image details and copyright information.

The series were arranged to reflect different levels of task complexity in terms of integrating their manifest content (SC < FW < DI). The easy SC series was created for the present experiment to include a series with an easily identifiable topic. In the series, all scenes present one adult with a group of children in a relatively formal teaching environment. The pictured spaces, outfits of people and original media (the final image using a more modern medium, a photograph) further signal a temporal progression from the first to the last image (see Appendix A for image details).

In the FW series, created for Puurtinen et al. (2015), the two initial images represent women doing textile work, first in a manufacture and then in an industrial milieu. The early stages of the textile industry are a standard topic in Finnish history education, when learning about the development of industrialization in Europe. However, the third image depicting heavy machinery challenges the assumption that development of the textile industry would be an appropriate topic for the whole series. Furthermore, the original media, engravings and paintings, do not give hints about any temporal progression from one image to the next. In the original study (Puurtinen, et al., 2015) history students coped with the new information in the third image by, for instance, generalizing the topic to industrialization, and most were able to integrate the images under an appropriate historical topic. Unfortunately, in the present study, an analysis of the participants' responses, supported by additional post-task interview data in the case of some participants (reported elsewhere; see Procedure), revealed that the third image in the FW series did not result in a sufficiently clear discrepancy in the participants' responses, as a result of which the series was not considered for further analyses. Details of how the FW series was handled are available from the authors.

The DI series was initially created by Puurtinen et al. (2015) to represent a more rarely discussed historical topic, the development of medical care of infants. In the series, the first two scenes present a male doctor vaccinating a baby, who is held by a woman. The doctor is without the outfit and hospital milieu of today's medical procedures. The third scene shows the characters in a hospital setting. In this series, too, the original media do not give hints about any temporal progression. Indeed, in the original study, history students responded with a diverse range of topics to explain this series, struggled in integrating them under a coherent topic and even failed to identify the medical procedure depicted in all the scenes (Puurtinen et al., 2015). Thus, of the three, this series was considered as the most complex one to integrate.

All images were retrieved from online web archives for historical images, and their use, modification, and publishing were carried out with permission. The series were compiled by searching for images that were matched in terms of their manifest content (see above), and, to some extent, in terms of composition, the people depicted in them (gender, role, size of the image, and tasks being performed), and the space in which the people were located (indoor or outdoor; large or closed space). They also deliberately excluded any recognizable historical characters. The original scenes were converted into a grayscale format to minimize the effect of differences in coloring on visual processing. The final coloring scheme for each series was manually adjusted using MATLAB. In addition, some scenes were slightly cut to ensure an equal size for all images (1163 x 768 pixels), and two scenes were rotated horizontally so that they better matched the composition of the other two scenes in their series. During the experiment, each scene was placed at the center of the computer screen against a midtone gray background (hex color #B1B1B1).

## Apparatus

A Tobii X3-120 Eye Tracker, with a sampling frequency of 120 Hz, was employed to track eye movements, while Tobii Pro Lab software was used for collecting and preprocessing the eye-tracking data (data export with version 1.20744884, x64). The eye tracker was attached to a 15.6-inch screen laptop, in which the software was installed. The screen resolution was 1920 x 1080 pixels. A 5-point calibration was performed before recording the data for each image series, and all measures were recorded as averages of both eyes. The room lighting during the measurement process was kept as constant as possible by conducting the measurements in rooms with no windows or those equipped with curtains. Nonetheless, different recording locations resulted in somewhat different lighting conditions between measurements. The REDcap online survey tool was employed to collect the participants' written responses to the visual task and the background survey. Notably, the survey tool was presented to the participants on a separate laptop.

## Procedure

Data collection took place at universities, museum facilities, and in an indoor space at the Medieval Turku Event. Upon entering the data collection site, the participant first signed the informed consent form and was then introduced to the eye tracker. Next, the participant rehearsed the calibration procedure and was familiarized with the visual task using the practice series (see Figure 2). All relevant instructions were provided to the participant in writing on a computer screen, the reading of which was self-paced. In the instructions, the participant was informed that (a) they would view three images, (b) they would be asked to write down their response to the question "What do you think the series is about?" after viewing all three images, and (c) each image would appear for 4 seconds, and they should look at an "X" on the screen before an image appears. After viewing the three scenes, the participant was (d) reminded about the task and instructed to give their response in 1–3 sentences on a separate laptop. After the practice stage, the same procedure was repeated using the three series (SC, FW, DI), with their presentation order randomized between participants.

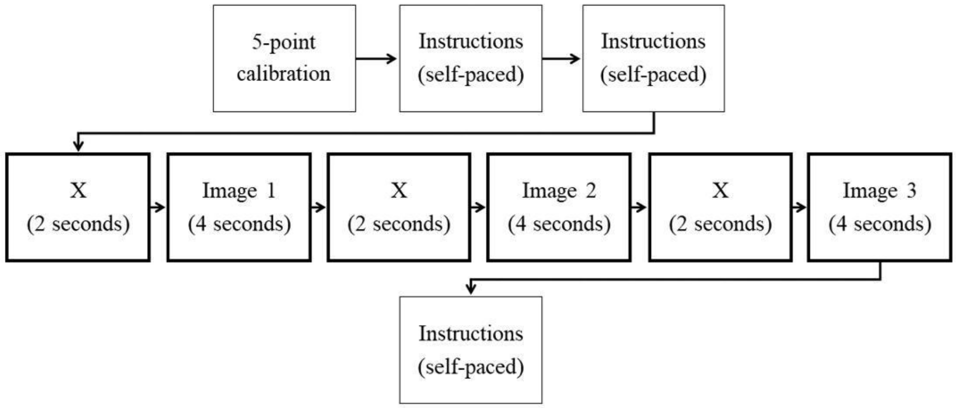


**Figure 2.** Presentation of Instructions, Fixation Crosses, and Images in the Experiment
**Note.** For the practice series, an additional instruction slide, which was presented after calibration, informed the participants about the practice stage.

After the completion of the eye-tracking procedure, the participants recruited from history departments and museums were interviewed about their interpretations and the use of visual materials in their work while simultaneously showing them their own eye-tracking recordings. The qualitative data obtained through these interviews address a different set of research questions reported elsewhere. All participants completed a background survey, which marked the end of the session for the interviewed participants. Those participants recruited at the Medieval Turku Event and at Luostarinmäki Museum engaged in an informal discussion about their experience and the eye-tracking method with the experimenter at the end of their session while simultaneously watching their eye-tracking recordings. Measurement sessions for the interviewed participants lasted approximately 1.5 hours, while those for the others lasted approximately 30 minutes.

## Data Handling and Analysis

### Eye-Movement Data

To classify fixations, we applied a Tobii I-VT fixation filter with a 30 degree/s threshold. We used the "merge adjacent fixations" function, with the maximum time and angle between fixations set to 75 ms and 0.5 degrees, respectively, and only included fixations with a minimum duration of 60 ms. Noise reduction was considered the moving median, with a window size of 3 samples, while the window length for the velocity calculator was maintained at 20 ms.

Data quality was assessed by the first author—a researcher with extensive experience working with eye-tracking data from Tobii eye trackers. They reviewed the gaze videos of all recordings on Tobii Pro Lab by applying the Tobii I-VT fixation filter, maintaining the settings described above, while also consulting the notes from the recording diary. Moreover, video replays of the raw eye movement data were repeatedly inspected as needed. Subsequently, eight participants with large amounts of missing data (e.g., loss of gaze tracking or fixations landing outside the stimulus image) were excluded from further eye movement analyses (as explained in the Participants section).

The average duration of whole fixations (hereby, average [mean] fixation duration) and the number of whole fixations for each participant and scene were exported using the Metrics Export function in Tobii Pro Lab. To analyze fixation dispersion, an area of interest (AOI) grid with 6 x 4 squares (see Figure 3) was created using the Tobii Pro Lab software, following which the AOI visit duration per participant was exported using the Metrics Export function. All visit durations larger than 0 ms were coded as hits (1), while durations of 0 ms were coded as misses (0).

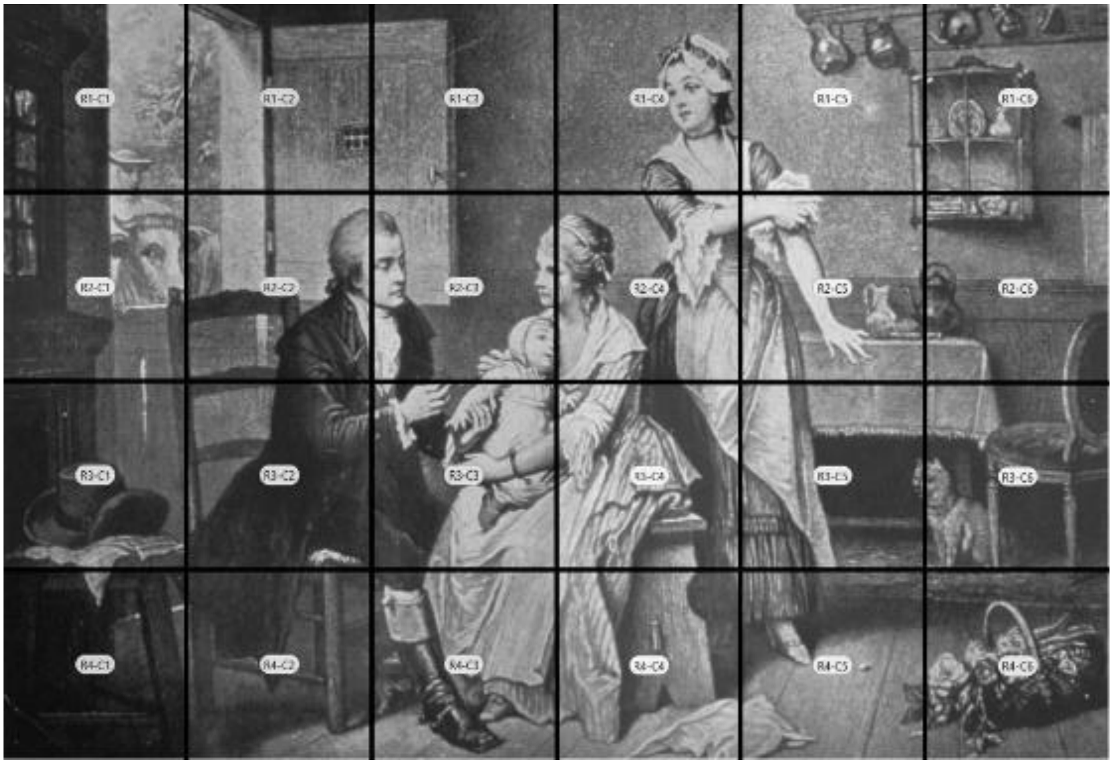

**Figure 3.** The 6 x 4 Grid Used for Analyzing Fixation Dispersion
**Note.** Modified for area of interest visualization by Tobii Pro Lab. Original image:
Vaccinating the Baby. Artist, Edouard Hamman. Published in New York, Wm. Wood & Co., 1890. CC 0.

To answer RQ1, eye movement measurements were analyzed using (generalized) linear mixed models by implementing the lme4 package (Bates et al., 2015) for R statistical software (version 4.1.3; R Core Team, 2022) in RStudio (version 2023.09.0). Notably, separate models were fitted for the three dependent measures—number of fixations, average (mean) fixation duration, and number of visited AOIs. Furthermore, history degree (1 = yes, 0 = no), scene number in a series (1, 2, or 3), and task complexity (complex, easy) served as fixed effects, while the participant was considered a random effect (i.e., dependent measure ~ History degree x Scene number x Task complexity + (1 | Participant)).

## Task Performance

A data-driven thematic analysis scheme was devised to characterize each written response based on two perspectives: (1) content of the response (i.e., "What do you think the series is about?") and (2) whether the response presented a historical process. This scheme was developed through multiple iterative steps until two independent coders reached an acceptable consensus.

First, we identified the recurrent themes in the responses and grouped them into categories, hereby referred to as *topics* (see Appendix B for detailed descriptions). Notably, the complex image series resulted in a wider variety of topics (see Table 1, next page), which were further grouped into two categories: topics that accurately captured all scenes, and topics that did not accurately capture the activities or roles of the people in all scenes in a series.

Second, each response was additionally coded for whether the topic described a historical process. In particular, responses that described a temporal process, change, or development were coded as presenting a historical process. Expressions referring to a historical process were, for example, "the history of [theme]," "the development of [theme]," "first – then," "from – to," or "[theme] in different era" (see Appendix C for examples of how the responses were coded).

After establishing the coding guidelines, the reliability of the coding scheme was tested by two coders, who independently coded 19 responses. For the SC series, the similarity between the two coders in identifying the topic was 94.7% (18/19), and that for identifying both the topic *and* lack/presence of a historical process in the answer (yes/no) was 84.2% (16/19), while the similarities in the case of the DI series were 94.7% (18/19) and 84.2% (16/19), respectively. Consequently, the reliability of the coding scheme was considered adequate, and the coding of the first author was used for further analyses.

**Table 1.** Topics for the Two Series Based on the Data-Driven Content Analysis of All Responses

| | Topic captures all scenes | |
|---|---|---|
| **Series** | **Yes** | **No** |
| **Easy (School Children)** | Children's everyday life<br>Going to school<br>Teaching and learning<br>Education and educational systems | [None] |
| **Complex (Doctors and Infants)** | A child at the doctor's<br>Medicine | People's lives<br>Family life<br>Childcare<br>People's lives, including a doctor's visit[1]<br>Family life, including a doctor's visit[1]<br>Childcare, including a doctor's visit[1] |

**Note.** [1]The response mentions a doctor, a hospital, examination, or a visit to the doctor as an additional topic, but not in a way indicating that the medical topic integrates all scenes of the series.

To answer RQ2, we analyzed the data using generalized linear models, employing the glm function for R statistical software (version 4.1.3; R Core Team, 2022) in RStudio (version 2023.09.0). Separate models were fitted for the two dependent measures—topic quality (1 = topic does not cover all images, 0 = topic covers all images) and topic described as a historical process (1 = yes, 0 = no). Due to differences in the number of topics identified for the image series, we fitted separate models for the easy (School Children) and difficult (Doctors and Infants) series, with a degree in history (1 = yes, 0 = no) considered a fixed effect (i.e., dependent measure ~ HistoryDegree).

# Results

## Effects of Domain Knowledge and Task Complexity on Visual Processing

The global visual processing measures for historians and novices are presented in Table 2.

With regard to the number of fixations, we observed significant interaction between history degree and image number (Pr[>|z|] < 0.05; see Table 3 and Figure 4). For novices, the number of fixations tended to reduce from the first to the third image in a series, whereas historians' number of fixations remained more consistent throughout their viewing of the three images.

As for the average fixation duration, we observed significant interactions between image number and history degree, as well as between image number and task complexity (Pr[>|z|] < 0.05; see Table 4 and Figure 5). The average fixation duration of the novices increased from the first to the third image of a series, whereas that of the historians showed only a trend in this direction. Notably, this observation is consistent with the observed effects of the number of fixations. Second, irrespective of history degree, the average fixation duration for the first to the second and then to the third image displayed a greater increase for the complex series than for the easy series.

A significant effect of image number (Pr[>|z|] < 0.01) on the number of visited AOIs (per image) was observed. However, no such effect was observed for history degree and task complexity (see Table 5). This implies that in a series of three images, the first image is scanned most globally.

**Table 2.** Means and Standard Deviations (in Parentheses) of the Global Visual Processing Measures for Novices ($n$ = 40) and Historians ($n$ = 40), Presented Separately for Images in the Easy and Difficult Series

| | Number of fixations | | Average fixation duration (ms) | | Number of visited AOIs (max. 24) | |
|---|---|---|---|---|---|---|
| **Image** | **Novices** | **Historians** | **Novices** | **Historians** | **Novices** | **Historians** |
| | | | **Easy (School Children)** | | | |
| **1** | 12.03 (2.49) | 11.75 (2.58) | 231.88 (61.70) | 237.85 (57.79) | 7.00 (1.32) | 6.48 (1.40) |
| **2** | 10.08 (2.54) | 9.78 (2.88) | 250.08 (66.85) | 255.20 (78.71) | 6.55 (1.22) | 6.78 (1.25) |
| **3** | 9.90 (2.51) | 10.80 (2.32) | 255.23 (71.29) | 252.88 (90.22) | 6.45 (1.34) | 6.75 (1.46) |
| | | | **Complex (Doctors and Infants)** | | | |
| **1** | 12.43 (2.45) | 11.38 (2.46) | 226.00 (56.47) | 228.53 (61.27) | 6.43 (1.89) | 6.90 (2.20) |
| **2** | 11.78 (2.53) | 11.48 (2.81) | 245.83 (77.12) | 237.65 (61.90) | 6.20 (1.60) | 6.68 (1.58) |
| **3** | 9.88 (2.96) | 11.03 (2.55) | 301.30 (146.44) | 253.08 (63.02) | 5.03 (1.49) | 5.75 (1.81) |

**Table 3.** Fixed Effects in the Generalized Linear Mixed Model for Number of Fixations Per Image

| | *b* | *SE* | *z* |
|---|---|---|---|
| **(Intercept)** | 2.645 | 0.072 | 36.910 |
| **History degree (yes)** | -0.196 | 0.103 | -1.904 |
| **Image number** | -0.112 | 0.033 | -3.383 |
| **Task complexity (easy)** | -0.087 | 0.100 | -0.873 |
| **History degree (yes): Image number** | 0.097 | 0.047 | 2.063 |
| **History degree (yes): Task complexity (easy)** | 0.097 | 0.143 | 0.680 |
| **Image number: Task complexity (easy)** | 0.013 | 0.048 | 0.266 |
| **History degree (yes): Image number: Task complexity (easy)** | -0.041 | 0.067 | -0.613 |

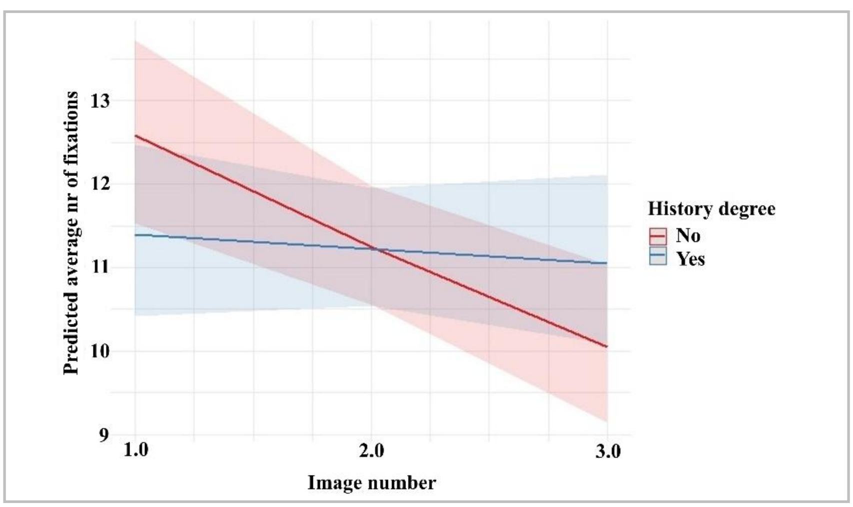


**Figure 4.** Predicted Average Number of Fixations per Image Number for Novices and Historians

**Table 4.** Fixed Effects in the Linear Mixed Model for Average Fixation Durations

| | *b* | *SE* | *t* |
|---|---|---|---|
| **(Intercept)** | 5.257 | 0.058 | 91.437 |
| **History degree (yes)** | 0.081 | 0.081 | 1.001 |
| **Image number** | 0.116 | 0.022 | 5.343 |
| **Task complexity (easy)** | 0.118 | 0.067 | 1.779 |
| **History degree (yes): Image number** | -0.062 | 0.031 | -2.021 |
| **History degree (yes): Task complexity (easy)** | -0.015 | 0.094 | -0.162 |
| **Image number: Task complexity (easy)** | -0.071 | 0.031 | -2.307 |
| **History degree (yes): Image number: Task complexity (easy)** | 0.035 | 0.044 | 0.794 |

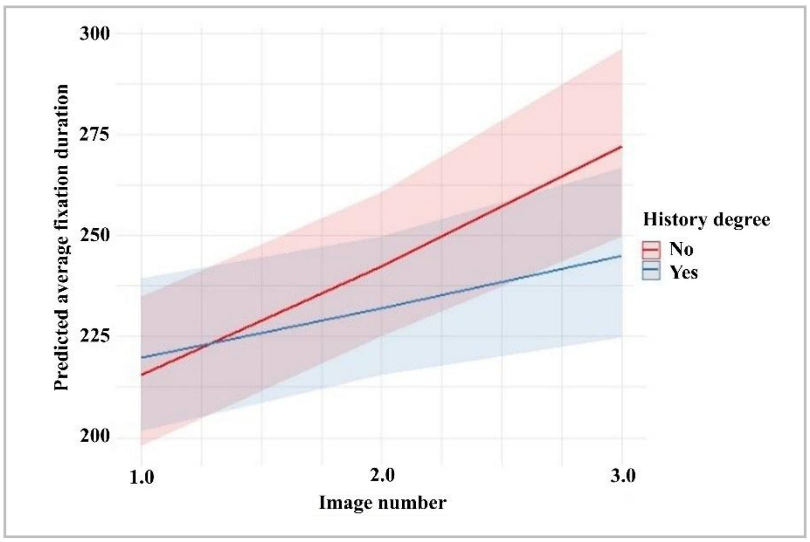


**Figure 5.** Predicted Average Fixation Duration per Image Number for Novices and Historians

**Table 5.** Fixed Effects in the Generalized Linear Mixed Model for Number of Visited Areas of Interest

| | *b* | *SE* | *z* |
|---|---|---|---|
| **(Intercept)** | 2.006 | 0.096 | 20.789 |
| **History degree (yes)** | 0.033 | 0.134 | 0.246 |
| **Image number** | -0.119 | 0.046 | -2.578 |
| **Task complexity (easy)** | -0.027 | 0.134 | -0.201 |
| **History degree (yes): Image number** | 0.030 | 0.064 | 0.467 |
| **History degree (yes): Task complexity (easy)** | -0.156 | 0.188 | -0.832 |
| **Image number: Task complexity (easy)** | 0.078 | 0.063 | 1.231 |
| **History degree (yes): Image number: Task complexity (easy)** | 0.032 | 0.089 | 0.361 |

## Effects of Domain Knowledge and Task Complexity on Interpretation

Table 6 shows the characteristics of the interpretations of series topics made by the historians and novices. When examining topic quality, it was observed that the easy series resulted in a ceiling effect—all participants were able to provide a topic that covered all images, irrespective of history degree. For the complex series, we observed a tendency toward the positive effect of expertise—historians were more likely to provide topics that covered all images of a series than novices (Pr[>|z|] = 0.067; see Table 7).

With regard to the topic being presented as a historical process, no significant effect of history degree was observed for the easy series (see Table 8). However, for the complex series, history degree was found to play a notable role—historians were significantly more likely to describe the complex series as a historical process (Pr[>|z|] < 0.01; see Table 9).

Overall, it was observed that history degree affected the quality of responses, but only when the task was complex. In other words, the historians showed a tendency to verbally integrate the images in even the complex series, while also describing it as representing a historical process.

**Table 6.** Characteristics of the Topics Chosen by Novices ($n$ = 40) and Historians ($n$ = 40) for the Easy and Complex Series

| | **Easy (School Children)** | | **Complex (Doctors and Infants)** | |
|---|---|---|---|---|
| | **Topic covers all images** | **Topic does not cover all images** | **Topic covers all images** | **Topic does not cover all images** |
| **Novices** | 40 | 0 | 11 | 29 |
| **Historians** | 40 | 0 | 19 | 21 |
| | **Topic presented as a historical process** | **Topic not presented as a historical process** | **Topic presented as a historical process** | **Topic not presented as a historical process** |
| **Novices** | 22 | 18 | 7 | 33 |
| **Historians** | 29 | 11 | 22 | 18 |

**Table 7.** Fixed Effects in the General Linear Model for Probability of Topic Covering All Images: The Complex Series (Doctors and Infants)

| | ***b*** | ***SE*** | ***z*** |
|---|---|---|---|
| **Intercept** | 0.97 | 0.35 | 2.74 |
| **History degree (yes)** | -0.87 | 0.48 | -1.83 |

**Table 8.** Fixed Effects in the General Linear Model: Probability of the Topic Being Described as a Historical Process. The Easy Series (School Children)

| | ***b*** | ***SE*** | ***z*** |
|---|---|---|---|
| **Intercept** | 0.20 | 0.32 | 0.63 |
| **History degree (yes)** | 0.77 | 0.48 | 1.62 |

**Table 9.** Fixed Effects in the General Linear Model: Probability of the Topic Being Described as a Historical Process. The Complex Series (Doctors and Infants)

| | *b* | *SE* | *z* |
|---|---|---|---|
| **Intercept** | -1.55 | 0.42 | -3.73 |
| **History degree (yes)** | 1.35 | 0.52 | 2.58 |

## Discussion

Building on expertise research and scene perception studies, we investigated the characteristics of visual expertise in the domain of history. More specifically, we studied the visual processing and task performance exhibited by historians and novices in response to an image-integration and interpretation task.

We hypothesized that owing to their ability to quickly activate appropriate schemata (VanLehn, 1989; Võ, 2021) through the efficient use of expert memory (Ericsson & Kintsch, 1995), professional historians would require shorter fixation durations than novices during image inspection (Gegenfurtner et al., 2011; Jarodzka et al., 2017). Our findings were partly in agreement with this hypothesis. Historians exhibited shorter fixation durations, but only in the case of the last (third) images of a series. Meanwhile, the average fixation frequencies displayed a different pattern—novices exhibited a larger number of fixations than historians for the first images (as in Humphrey & Underwood, 2009; Underwood et al., 2009), but less so for the third images of the series. This finding is partly consistent with those of Humphrey and Underwood (2009) in their image recognition task. However, the authors of that study considered the overriding effect of domain knowledge to be stable over time. In contrast, the present study, in which images had to be compared to one another, points out that the effect of domain knowledge is moderated by task requirements. To summarize, historians demonstrated consistency in their visual processing throughout the task, inspecting all three images somewhat similarly, with respect to fixation frequency and durations (e.g., van der Gijp et al., 2017). In contrast, novices were more affected by the number of an image—after an initial search for possible topics when inspecting the first image, they became more selective when deciphering the later images.

We also expected the historians to exhibit more focused (i.e., less spatially distributed) visual processing of images owing to their more efficient holistic perception (Kundel et al., 2007; Sheridan & Reingold, 2017). However, no such general expertise effect was observed. In the case of all participants alike, the inspected area was the largest for the first images of the series and diminished toward the last image. With regard to task complexity, we expected the increase in cognitive demand to result in fewer and longer fixations for all participants (Walter & Bex, 2021), with expertise likely moderating this effect. However, we noticed only a general effect of increased cognitive demand for visual processing on average fixation durations. Irrespective of domain knowledge, the average fixation durations increased from the first to the third image in each series, especially for the complex series. The lack of expertise effects in these measures may be partly attributed to the use of authentic and, therefore, visually complex historical images in which the artists' compositional choices come into play, thus guiding the visual processes of all viewers alike. In other words, as the order of presenting the images of one series was kept stable (in order to enable the identification of historical processes and create comparable conditions for all participants), the finding might—at least in part—reflect the effects of image characteristics. Another explanation is that, regarding the spatial distribution of fixations, the task requirements overcame some of the possible expertise effects: perhaps the participants' attempts at interpreting and integrating the images in the given 4-s time window compelled them to search the first images for cues on potential topics, due to which their focus in the later images might have been more on items relevant to their active schemata. A longer viewing time might have made other visual expert (or novice) strategies more manifest.

We also examined the effects of expertise on the participants' task performance, that is, the quality of the topics interpreted from the historical image series. Although our hypotheses about historians' and novices' visual processing were met only in part, historians' expertise still resulted in improved response quality for the more challenging image-integration task. The historians found a way to describe the complex series as a representation of a historical process and, unlike novices, they also showed a tendency to categorize the series under an appropriate topic (see also, e.g., Wineburg, 1991; Leinhardt & Young, 1996; Rouet et al., 1997; Greene & Yu, 2014; Lévesque et al., 2014). For the easy task, all participants performed at the ceiling level. In other words, historians were able to contextualize even the more complex series and place it within a meaningful *bigger picture* (see, Wineburg, 1991; Lévesque et al., 2014). Notably, this result was expected—historians' highly developed memory skills and ability to rapidly move between different competing schemata aided them in contextualizing the series even when rapidly presented with a challenging set of images (VanLehn, 1989; Ericsson & Kintsch, 1995).

To summarize our findings, domain expertise enables the effective use of appropriate schemata also in visual tasks of historical interpreting, and such effects are manifest also on the eye-movement level. The results showed that compared to novices, historians exhibited more consistent gaze behavior (in terms of average fixation duration and fixation frequency) for each series of images, indicating a systematic approach to interpreting them. Furthermore, novices' fixation durations increased, and the number of fixations tended to decrease from the first to the final image, suggesting that their gaze became more focused only for the later images. Expertise also affected task performance: historians tended to provide more appropriate topic answers and interpreted the topics as part of a historical process, indicating that they were able to situate even more complex sets of images within a meaningful bigger picture. Thus, even in the ill-structured domain of history, domain expertise induces differences in the visual processing and interpretation of domain-related images compared to lack of such expertise.

This study is not without limitations. For one, to be able to make conclusions that can be generalized beyond this specific experiment, we focused only on macro-level eye-movement measures and did not pay attention to those that should be even more affected by exact stimulus features. After the initial steps reported here, we hope further research can move on to explore the nuances of the interplay between history expertise and scene perception in more detail. For another, although some stimuli used in our experiment were piloted beforehand (Puurtinen et al., 2015), the participants were so varied in their interpretation of one of the series that it could not be considered for further analyses. This demonstrates the challenges arising from the use of authentic images in quasi-experimental studies. Nevertheless, we consider the visual task of naming a series topic after a brief presentation of three images to be quite suitable for the purposes of our experiment for several reasons. First, investigations into the effects of expertise on visual task performance often encounter the problem of the task being not accessible to novices. In our case, the open-ended nature of the task and the used historical imagery ensured that it was meaningful and accessible to historians and novices alike. For the novices, responding to the task did not require the use of highly specialized language or making sense of visual materials that they would not otherwise encounter. Second, the task left room for various response and visual strategies, which allowed for a detailed examination of the extent and the ways in which expertise is associated with the different characteristics of the responses and visual processing. This was also considered in the data-driven coding of the responses. Third, the limited viewing time provided to the participants suited the purposes of this study on the potential effects of domain knowledge on visual processing. We opted for strategic but selective viewing.

Through our study, we sought to take a step toward finding a balance between experimental control and domain relevance in expertise studies during scene perception tasks (see, Võ, 2021). We based the present work on two types of research literature—expertise studies and scene perception studies—hoping to demonstrate the benefits of integrating the two strands of literature. The process of addressing the gaps in one kind of research through insights and ideas drawn from the theoretical approaches and methodological choices prevalent in another research domain seems to be useful in explaining skillful visual processing and task performance more fully, especially in the context of underexplored domains that offer little empirical evidence to lean on. From this juncture, the avenues for expanding the cross-disciplinary characterization of visual expertise in history are numerous. For instance, one could apply historical imagery in a series of tasks characterized by different kinds of instructions and goals while also accounting for viewers' expertise and its impact on visual processing. One could also complement qualitative research on skills required in the history domain with theoretical and methodological insights from scene perception and eye-tracking studies. We hope that our work inspires other researchers to experiment with their research designs and approaches.

## Endnote

1. The third image in the FW series was changed for the present experiment due to changes in copyright permissions. The original idea of the series topic was maintained.

## Authors' Declarations

The authors declare that all data and code used in this study have been made available at https://osf.io and can be accessed at https://osf.io/uba92/overview?view_only=acaa62ad0c9a4e0cb6c0c2f14010f671 (Puurtinen et al., 2024). This study was not preregistered.

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that the research reported in this article is in accordance the Ethical Principles of the *Journal of Expertise*.

## Acknowledgements

This research was funded by the Research Council of Finland (decision number 340381). Eye-tracking equipment and software were provided by Turku EyeLabs (www.utu.fi/turkueyelabs). The authors thank all study participants and project collaborators at the different data collection sites for their time and effort, as well as project worker Henri Harju for help in the data collection process.

## ORCID iDs

Marjaana Puurtinen
https://orcid.org/0000-0003-1972-2268

Johanna K. Kaakinen
https://orcid.org/0000-0002-2970-826X

Halszka Jarodzka
https://orcid.org/0000-0003-2312-4703

## References

Baron, C. (2012). Understanding historical thinking at historical sites. *Journal of Educational Psychology, 104*(3), 833–847. https://doi.org/10.1037/a0027476

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bleichmar, D., & Schwartz, V. R. (2019). Visual history: The past in pictures. *Representations*, *145*, 1–31. https://www.jstor.org/stable/10.2307/26616020

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, *14*(3), 29–29. https://doi.org/10.1167/14.3.29

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*(6–7), 790–811. https://doi.org/10.1080/13506280902793843

Drew, T., Võ, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science*, *24*(9), 1848–1853. https://doi.org/10.1177/0956797613479386

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), 1–15. https://doi.org/10.1167/9.3.6

Cole, C. (1998). Information acquisition in history Ph.D. students: Inferencing and the formation of knowledge structures. *Library Quarterly, 68*, 33–54

Cole, C. (2000). Name collection by Ph.D. history students: inducing expertise. *Journal of the American Society for Information Science*, *51*(5), 444–455. https://doi.org/10.1002/(SICI)1097-4571(2000)51:5%3C444::AID-ASI5%3E3.0.CO;2-Y

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037//0033-295x.102.2.211

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge University Press.

Foulsham, T. (2015). Scene perception. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 257–280). MIT Press. https://doi.org/10.7551/mitpress/10033.003.0013

Gegenfurtner, A., Boucheix, J.-M., Gruber, H., Hauser, F., Lehtinen, E., & Lowe, R. K. (2020). The gaze relational index as a measure of visual expertise. *Journal of Expertise, 3*(1), 32–40. Available at https://www.journalofexpertise.org.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Gegenfurtner, A., Gruber, H., Holzberger, D., Keskin, Ö., Lehtinen, E., Seidel, T., Stürmer, K., & Säljö, R. (2024). Towards a cognitive theory of visual expertise: Methods of inquiry. In G. Ritella, A. Rajala, C. Damşa, & J. Brouwer (Eds.), *Re-Theorising Learning and Research Methods in Learning Research* (1st ed., Vol. 1, pp. 142–158). Routledge. https://doi.org/10.4324/9781003205838-10

Greene, J. A., & Yu, S. B. (2014). Modeling and measuring epistemic cognition: A qualitative re-investigation. *Contemporary Educational Psychology*, *39*(1), 12–28. https://doi.org/10.1016/j.cedpsych.2013.10.002

Gottlieb, E., & Wineburg, S. (2012). Between *Veritas* and *Communitas*: Epistemic switching in the reading of academic and sacred history. *The Journal of the Learning Sciences, 21*(1), 84–129. https://doi.org/10.1080/10508406.2011.582376

Goulding, J. (2021). Historical thinking online: An analysis of expert and non-expert readings of historical websites. *Journal of the Learning Sciences, 30*(2), 204–239. https://doi.org/10.1080/10508406.2020.1834396

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements* (pp. 537–562). Elsevier. https://doi.org/10.1016/B978-008044980-7/50027-6

Humphrey, K., & Underwood, G. (2009). Domain knowledge moderates the influence of visual saliency in scene recognition. *The British Journal of Psychology*, *100*(2), 377–

398. https://doi.org/10.1348/000712608X344780

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research, 10*(1), 1–18. https://doi.org/10.16910/jemr.10.1.3

Kaakinen, J., Hyönä, J., & Viljanen, M. (2011). Influence of a psychological perspective on scene viewing and memory for scenes. *The Quarterly Journal of Experimental Psychology, 64*(7), 1372–1387. https://doi.org/10.1080/17470218.2010.54887

Kainulainen, M., Puurtinen, M., & Chinn, C. A. (2019). Historians and conceptual change in history itself: The domain as a unit of analysis. *International Journal of Educational Research, 98*, 245–256. https://doi.org/10.1016/j.ijer.2019.09.001

Kainulainen, M., Puurtinen, M., & Chinn, C. A. (2022a). Historians' criteria for appraising historiography: An interview study. In C. A. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences 2022* (pp. 989-992). Available at Proceedings - ISLS Annual Meeting 2022.

Kainulainen, M., Puurtinen, M., & Chinn, C. A. (2022b). Aims in the practice of historiography: An interview study with Finnish historians. *Historical Encounters*, *9*(1), 156–175. https://doi.org/10.52289/hej9.109

Kainulainen, M., Puurtinen, M., & Chinn, C. A. (2025). Regrounding inquiry-based learning in history: A study of historians' epistemic processes. *Cognition and Instruction*, 1–25. https://doi.org/10.1080/07370008.2025.2503193

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2012). Looking in the same manner but seeing it differently: Bottom-up and expertise effects in radiology. *Applied Cognitive Psychology*, *26*(6), 854–862. https://doi.org/10.1002/acp.2886

Kundel, H. L., & Nodine, C. F. (1983). A visual concept shapes image perception. *Radiology, 146*(2), 363–368. https://doi.org/10.1148/radiology.146.2.6849084

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, *242*(2), 396–402. https://doi.org/10.1148/radiol.2422051997

Leinhardt, G., Stainton, C., Virji, S. M., & Odoroff, E. (1994). Learning to reason in history: Mindlessness to mindfulness. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 131–158). Lawrence Erlbaum Associates.

Leinhardt, G., & Young, K. M. (1996). Two texts, three readers: Distance and expertise in reading history. *Cognition and Instruction*, *14*(4), 441–486. http://www.jstor.org/stable/3233783

Lévesque, S., Ng-A-Fook, N. & Corrigan, J. (2014). What does the eye see? Reading online primary source photographs in history. *Contemporary Issues in Technology and Teacher Education*, *14*(2), 101–140. https://www.learntechlib.org/primary/p/130219/

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(4), 565–572. https://doi.org/10.1037/0096-1523.4.4.565

McGrew, S. (2022). Internet or archive? Expertise in searching for digital sources on a contentious historical question. *Cognition and Instruction 40*(4), 488-516. https://doi.org/10.1080/07370008.2021.1908288

Monte-Sano, C. (2017). Bridging reading and writing: Using historians' writing processes as clues to support students. In G. Andrews, & Y. Wangdi (Eds.), *The role of agency and memory in historical understanding: Revolution, reform, and rebellion* (pp. 247–265). Cambridge Scholars Publishing.

Monte-Sano, C., & Thomson, S. (2022). It's not that simple: Re-thinking historical writing tasks based on insights from disciplinary experts. *The History Teacher (Long Beach, Calif.)*, *55*(3), 391–418.

Nye, A., Hughes-Warrington, M., Roe, J., Russell, P., Peel, M., Deacon, D., Laugeson, A., & Kiem, P. (2009). Historical thinking in higher education: Staff and student perceptions of the nature of historical thinking. *History Australia, 6*(3), 73.1–73.16. https://doi.org/10.2104/ha090073

Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (2015). Reading ahead: Adult music students' eye movements in temporally controlled performances of a children's song. *International Journal of Music Education: Research, 33*(1)*,* 36–50. https://doi.org/10.1177/0255761413515813

Puurtinen, M., Kaakinen, J., & Jarodzka, H. (2024). *Data and code for historians' and novices' visual processing and interpretation of historical images* Open Science Framework. https://osf.io/uba92/overview?view_only=acaa62ad0c9a4e0cb6c0c2f14010f671

Puurtinen, M., Nivala, M., & Virta, A. (2015). Visual sources and historical thinking in higher education. *Nordidactica – Journal of Humanities and Social Science Education, 4*, 1–20. Available at https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A904133&dswid=4370

R Core Team (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rouet, J. F., Fabart, M., Anne Britt, M., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, *15*(1), 85–106. https://doi.org/10.1207/s1532690xci1501_3

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. L. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Lawrence Erlbaum Associates.

Sheridan, H., & Reingold, E. M. (2017). The holistic processing account of visual expertise in medical image perception: A review. *Frontiers in Psychology*, *8*, 1620. https://doi.org/10.3389/fpsyg.2017.01620

Schneider, J., & Zakai, S. (2016). A rigorous dialectic: Writing and thinking in history. *Teachers College Record (1970)*, *118*(1), 1–36. https://doi.org/10.1177/016146811611800102

Shanahan, C., Shanahan, T., & Misischia, C. (2011). Analysis of expert readers in three disciplines: history, mathematics, and chemistry. *Journal of Literacy Research 43*(4), 393–429. https://doi.org/10.1177/1086296X11424071

Shreiner, T. L. (2014). Using historical knowledge to reason about contemporary political issues: An expert-novice study. *Cognition and Instruction*, *32*(4), 313–352. https://doi.org/10.1080/07370008.2014.948680

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, *1*(1), 7–27. https://doi.org/10.1068/i0382

Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, *17*(6–7), 812–834. https://doi.org/10.1080/13506280902771278

van der Gijp, A., Ravesloot, C. J., Jarodzka, H., van der Schaaf, M. F., van der Schaaf, I. C., van Schaik, J. P. J., & ten Cate, T. J. (2017). How visual search relates to visual diagnostic performance: A narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education: Theory and Practice, 22*(3), 765–787. https://doi.org/10.1007/s10459-016-9698-1

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527–579). MIT Press.

Voet, M., & De Wever, B. (2017). History teachers' knowledge of inquiry methods: An analysis of cognitive processes used during historical inquiry. *Journal of Teacher Education, 68*(3), 312–329. https://doi.org/10.1177/0022487117697637

Voss, J. F., & Wiley, J. (2006). Expertise in history. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 569–584). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.033

Võ, M. L.-H. (2021). The meaning and structure of scenes. *Vision Research, 181*, 10–20. https://doi.org/10.1016/j.visres.2020.11.003

Walter, K., & Bex, P. (2021). Cognitive load influences oculomotor behavior in natural scenes. *Scientific Reports*, *11*(1), 12405–12405. https://doi.org/10.1038/s41598-021-91845-5

Woodman, G. F., & Luck, S. J. (2015). Using working memory to control attentional deployment to items in complex scenes. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 173–198). MIT Press.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87. https://doi.org/10.1037/0022-0663.83.1.73

Wineburg, S. S. (1994). The cognitive representation of historical texts. In G. Leinhardt, I. L. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 85–136). Lawrence Erlbaum Associates.

Wineburg, S. S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, *22*(3), 319–346. https://doi.org/10.1016/S0364-0213(99)80043-3

Wolf, T. (1976). A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research, 5*(2), 143–171.

Received: 3 April 2025
Revision received: 17 October 2025
Accepted: 27 October 2025

# Appendix A

## Details and Copyright Information of Images Used in the Experiment

| Series | Scene no. | Details and copyright information |
|---|---|---|
| Practice | 1 | Village Street with Two Carts. Artist Aegidius Sadeler II (Netherlandish, Antwerp 1568–1629 Prague), after Jan Brueghel the Elder (Netherlandish, Brussels 1568–1625 Antwerp), *n*.d. Publisher Marcus Sadeler (German, Munich before 1614–in or after 1650). The Elisha Whittelsey Collection, The Elisha Whittelsey Fund, 1949. Retrieved in 2022 from: https://www.metmuseum.org/art/collection/search/382734 CC 0 |
| | 2 | Gezicht op Heeg. Print maker Carel Frederik Bendorp (I), after drawing by Jan Bulthuis, 1786–1792. Retrieved in 2022 from: http://hdl.handle.net/10934/RM0001.COLLECT.730617 CC 0 |
| | 3 | Grafton Street. Dublin. Creator unknown, 1900/1910. Girona City Council. Retrieved in 2022 from: https://www.europeana.eu/en/item/2024914/photography_ProvidedCHO_Ajuntament_de_Girona_353163 CC 0 |
| Easy (School Children) | 1 | The Village School, from "Illustrated London News." Artist After Alfred Rankley (British, 1819–1872 London), 1856. Gift of Donato Esposito, 2013. Retrieved in 2022 from: https://www.metmuseum.org/art/collection/search/626349 CC 0 |
| | 2 | A Country School. Artist Edward Lamson Henry, 1890. Yale University Art Gallery. Retrieved in 2022 from: https://artgallery.yale.edu/collections/objects/49982 CC 0 |
| | 3 | Interieur van de Openbare School voor Buitengewoon Lager Onderwijs Kruisdwarsstraat 6 te Utrecht: Schoolklas met kinderen. Photographer unknown, 1915–1925. 121116 / Het Utrechts Archief. Retrieved in 2022 from: https://hetutrechtsarchief.nl/beeld/43E08B339340519480A35DD60CB1DF51 CC 0 |
| Semi-complex (Female Workers) | 1 | Flax Barn at Laren. Artist Max Liebermann (1847–1935), 1887. Nationalgalerie. Retrieved in 2022 from https://nat.museum-digital.de/index.php?t=objekt&oges=420206 CC BY-NC-SA 3.0 DE |
| | 2 | Carding, Drawing and Roving. Textiles: Women and children working at large cotton manufacturing machines. Wellcome Collection. Retrieved in 2022 from https://library.artstor.org/asset/24885239 CC BY 4.0 |
| | 3 | The Munitions Girls. Artist Alexander Stanhope Forbes, 1918. Science Museum, London.<br>Retrieved in 2022 from: https://collection.sciencemuseumgroup.org.uk/objects/co65065/the-munition-girls-oil-painting CC BY-NC-SA 4.0 |
| Complex (Doctors and Infants) | 1 | Vaccinating the Baby. Artist Ed Hamman. Publisher Wm. Wood & Co., New York, 1890. Retrieved in 2022 from: http://resource.nlm.nih.gov/101447863 CC 0 |
| | 2 | La vaccination gratuite à Paris. Artist Jules Scalbert, 1890. Retrieved in 2022 from: http://resource.nlm.nih.gov/101392775 CC 0 |
| | 3 | Le Tubage. Artist Georges Chicotot, Georges, 1904. AP-HP/musée - F. Marin. Published with permission |

# Appendix B

## Series Topics and Descriptions Based on Data-Driven Thematic Analysis

| Series | Topic and description |
|---|---|
| Easy (School Children) | Child's everyday life<br><br>The response mentions childhood, children's life, children's upbringing, or children's life at home and at school as the overarching theme that integrates the images.<br><br>Going to school<br><br>The response mentions going to school, being in school/preschool, a classroom, or being in a classroom as the theme of the images. *Note: this theme centers on the Finnish concept of "koulunkäynti," which is usually translated as "schooling." However, this compound word also holds the very practical meaning of "going to school." Thus, the word does not necessarily imply a higher level of meaning related to, for instance, educational systems.*<br><br>Teaching and learning<br><br>The response mentions learning, teaching, a teaching situation, or studying as the theme expressed by the images. If the response mentions both teaching or learning and going to school or the classroom, choose this theme over "Going to school."<br><br>Education and educational systems<br><br>The response presents education, education system, or school as an institution as the theme that integrates the images. Responses in which the development of educational systems is described in terms of specific historical institutions (e.g. "kansakoulu" and "mamsellikoulu," which refer to earlier forms of relatively organized education in Finland) should also be included under this topic. |
| Complex (Doctors and Infants) | A child at the doctor's<br><br>The response mentions children's doctors, children's healthcare, a child at the doctor's, or a child's medical procedure or operation as the theme that integrates the images. The description of the procedure can be fully or partly inaccurate (e.g., dental care).<br><br>Medicine<br><br>The response refers to doctors, healthcare, or a medical procedure or operation as the theme that connects the images. The description of the procedure can be fully or partly inaccurate (e.g. about dental care).<br><br>People's life<br><br>The response mentions children, mothers, fathers, or specific groups of people or their lives as the unifying theme that connects the images.<br><br>Family life<br><br>The response mentions family, family life, new family members, family-related events, or family culture as the theme that integrates the images. If the response mentions both family life and people and their lives, choose this theme over "People's life."<br><br>Childcare<br><br>The response refers to childcare or living with children as the theme that integrates the images. If the response mentions family life and/or people and their lives as well as childcare, choose this theme over "People's life" and "Family life." (continued on next page) |

| Series | Topic and description |
|---|---|
| Complex (Doctors and Infants), continued | People's life, including a visit to the doctor<br>The response mentions children, mothers, fathers, or specific groups of people or their lives as the theme that integrates the images. In addition, the response mentions a doctor, a hospital, an examination, or a visit to the doctor, but not in a way indicating that it is a clear theme that integrates all images.* |
| | Family life, including a visit to the doctor<br>The response refers to family, family life, new family members, family-related events, or family culture as the theme that connects the images. In addition, the response mentions a doctor, a hospital, an examination, or a visit to the doctor, but not in a way indicating that it is a clear theme that integrates all images.* |
| | Childcare, including a visit to the doctor<br>The response states childcare or living with children as the theme that integrates the images. If the response also mentions people's lives or family life, choose this theme. In addition, the response mentions a doctor, a hospital, an examination, or a visit to the doctor, but not in a way indicating that it is a clear theme that integrates all images.* |

* Examples of expressions indicating a visit to the doctor that are considered as addressing only some of the images of a series, and not all of them: "on the other hand," "also," "the last image."

# Appendix C

## Coding Example: Written Responses to the Task "What Do You Think the Series is About?"

| Series | Response | Topic | Topic captures all scenes | Historical process |
|---|---|---|---|---|
| Easy (School Children) | Teaching situations at home and at school. In the first image you are likely at home, in the others at school. In all images mixed groups [gender-wise] and women as teachers. (Novice, participant EOH1-H72) | Teaching and learning | Yes | No |
| | The historical development of education from individual teaching to classroom teaching. (Historian, participant EOH1-H17) | Education and educational systems | Yes | Yes |
| Complex (Doctors and Infants) | The series was about three different families in an ancient historical era. Each family had a similar familial living situation, where [there is] a child of a certain age, and the father and mother are taking care of the child. (Novice, participant EOH12-H33) | Family life | No | No |
| | The change of children's medical care/examinations. Male doctors perform external examinations for children in a medical sense. (Historian, participant EOH1-H28) | A child at the doctor's | Yes | Yes |