

Chapter 3

Digital Pictures, Videos, and Beyond: Knowledge Acquisition with Realistic Images

Stephan Schwan

Abstract While the distinction between pictures and texts is well established on theoretical grounds and has attracted much research, the differences and commonalities between realistic depictions and its real-world counterparts have received much less attention. This chapter aims to contribute to closing this gap by systematically comparing life-like images to real-world events in terms of commonalities and differences in visual appeal as well as in perception and mental processing. Based on the notion of a “dual character” of digital images, both closely resembling reality but simultaneously being systematically different, several issues regarding processes of knowledge acquisition will be discussed, including: Are viewers aware of differences between real-world information and mediated information—and do they take them into account? Do realistic images require specific competencies for comprehension? Should the realism of visual representations be maximized for learning? How do viewers deal with the informational complexity and ambiguity of realistic images?

Keywords Realistic pictures • Authenticity • Event cognition • Video

Introduction

Take a brief look at a current school textbook for biology, physics, geography, or history, and you will find that nearly 50 % of the page space is occupied by visual depictions the majority being realistic images (i.e., drawings or photographs that resemble real-world referents; Lee, 2010; Yasar & Seremet, 2007). Also, use of realistic depictions in education is not restricted to static pictures in textbooks. Instead, according to a recent survey, moving images such as films and videos are the most frequent types of media used in German classrooms (Institut für Demoskopie Allensbach, 2013). This abundance of realistic depictions is not a new development. From the beginning, modern science and education has been coupled with usage of

S. Schwan (✉)

Leibniz-Institut für Wissensmedien, Schleichstraße 6, Tübingen 72076, Germany

e-mail: s.schwan@iwm-tuebingen.de

visual depictions as a means of storing and distributing knowledge—be it prints and engravings of plants and animals from distant countries, like Sybilla Merian's famous depictions of exotic insects or the portrayals of technological inventions and machinery in Diderot's *Encyclopedia* (Stafford, 1994). Ever since the nineteenth century, these early forms of illustrations have been increasingly complemented by advancements in technologies for recording and for mass distribution of images, such as lithography, photography, and filming.

Today, digitization has led to an even broader scope of realistic depictions. From satellites to CCTV, Google Street View and camera traps to webcams, dashcams, or action cams—almost all aspects of reality are portrayed and made available via large Internet repositories such as YouTube or Flickr. Also, scientific research routinely uses digital photographing and filming for documentation and explanation, again building large digital databases (e.g., the Europeana platform in the humanities) or channels for scholarly communication, like the Journal of Visualized Experiments (JoVE) in the science domain. Also, advancements in computer graphics today allow creation of life-like renderings and simulations of objects, scenes, processes, or events with unprecedented fidelity. Accordingly, all these types of digitized images have made their way into formal and informal education, as exemplified by advanced digital textbooks (e.g., Wilson's digital biology textbook *Life on Earth*), game-based learning scenarios, or current museum exhibitions on science or natural history (e.g., the Welcome Wing of the Science Museum in London).

Why do images (in the sense of realistic or iconic depictions of real-world phenomena) play such an important role in science and education? From the perspective of educational psychology, realistic images have usually been treated as one particular class of representational media, and accordingly, the main focus both of theorizing and of empirical research has been on comparing them with other types of representational media. In particular, the most influential models in the field have contrasted visual depictions (with realistic images as an important type) with texts, assuming a difference between depictive and descriptive modes of information presentation (Schnotz, 2002), which is reflected in different subsystems of cognitive processing (Mayer, 2001), in different working memory compartments (Baddeley, 2012), and in different mental representations in long-term memory (Paivio, 1986). Contrasting pictures with text is not only motivated by the fact that both are the dominant modes of disseminating information in our culture, but also by its fundamental differences in terms of representational characteristics (Schnotz, 2002). Texts are based on arbitrary signs, conform to a grammar specifying rules of combining words to larger meaningful chunks, and easily allow for abstractions, generalizations, negations, changes of tempus, or counterfactual arguments. In contrast, pictures are organized in a two- or even three-dimensional manner, do not possess definite basic components (like words), do not conform to syntactical rules but, on the other hand, typically provide the viewer with a denser and detailed array of information which does not follow a single explicit argument and instead allows for inspection regarding various different purposes and questions. Pictures as a certain class of representational media may be further decomposed into logical pictures, such as graphs and diagrams, and images, with the latter having a relationship of

resemblance to real-world phenomena. According to Peirce (1940), this relationship can be described as iconic because some visual regularities (like shape or color) of real-world phenomena are mapped onto corresponding visual regularities of the image. Depending on the amount of mapped regularities, the resulting images may range from simple black and white line drawings to film clips with a high visual fidelity.

Scholarly discussion of the role of images in education has often focused on their illustrative, “decorative” purposes. It is assumed that while realistic visualizations may make learning material more attractive, thereby possibly heightening students’ motivation and interest, simultaneously they may hinder acquisition of relevant knowledge, eventually distracting students from processing and elaborating the relevant learning matter (which is thought to be primarily provided by texts, graphs, and diagrams; Magner, Schwonke, Alevén, Popescu, & Renkl, 2014; Rey, 2012). Still, realistic visualizations should not be reduced to mere decoration. Instead, the role of images in knowledge acquisition is far more diverse: They may present visual details that are difficult to describe verbally, may make spatial relations be easily picked up perceptually, or may specify the minute changes of biological movement patterns.

While the distinction between pictures and texts is well established on theoretical grounds and has attracted a considerable amount of research, another relevant distinction has received much less attention, namely, the differences and commonalities between realistic depictions and their real-world counterparts; that is, as an alternative, the content of most images can not only be described in words, but can also be perceived and experienced in a direct, unmediated way (at least in principle). A chemical experiment may be shown as a video or it may be verbally described in a textbook; however, it may also be directly demonstrated in the classroom. Similarly, famous architectures, important historical sites, or geomorphologic interesting landscapes can be portrayed as images, described in words, but can also be inspected directly on location. Starting from the distinction between image and real world, both the theoretical perspective and the corresponding research questions change fundamentally. Now we may ask: How does our perceptual and cognitive apparatus deal with life-like pictorial representations? Are they processed in a similar way to real, unmediated percepts or do they require certain kinds of specific visual literacy? Is knowledge acquisition by means of realistic images comparable to real, unmediated visual experiences? How can the differences between both modes of experience be systematically exploited for designing appropriate learning material? Accordingly, possible theoretical underpinnings of this approach can be found in models of everyday perception (Gibson, 1979) and event cognition (Zacks & Tversky, 2001) instead of in theories of text comprehension or models of multimedia learning.

Current developments in digital technologies will make these issues even more relevant. Not only has the realistic appeal of many digital images become nearly perfect—from CGX (i.e., digital effects) in movies to immersive consumer technologies like Oculus Rift. Also, digital technology has left the standard computer cases and has started to inhabit many different devices and objects, from smart

watches to home heating, blurring more and more the borderline between reality and digital virtuality. Part of next generation educational tools will not be based on didactically motivated decisions between images and words or combinations of them, but on decisions between life-like visualizations and real-life experiences or, again, combinations of them. Therefore, in the following, I will try to sketch this alternative view of processing digital images for knowledge acquisition in more detail. In particular, I will explore certain questions about the implications of inspecting (static or dynamic) pictorial representations for learning, which look life-like but nevertheless systematically differ from real life in a number of relevant ways.

Real-Life Presentations and Life-Like Representations: Commonalities and Differences

One obvious purpose of realistic depictions is to reproduce a view of a certain real-world phenomenon in a permanent way, thus serving as a materialized, external kind of visual memory. Observing or scrutinizing this phenomenon is thereby no longer bound to its existence at a certain place and time, but instead becomes independent of it. In many circumstances, this is important, for example, if the phenomenon is singular and short-lived or if it cannot be inspected in real life due to its geographical distance—think of the pictures that NASA’s “New Horizon” sent from the outer parts of our solar system as the most extreme case. In other words, realistic images allow for cultural transmission of visual information across space and time.

Of course, images do not keep record of a relevant event, object, or any other real-world phenomenon in its entirety but are mostly restricted to its visual appearance, lacking information about its nonoptical aspects, for instance, its acoustics, its odor, or its tactile qualities. On the other hand, images may also result from transforming information that exceeds the scope of the human visual system, as is the case for X-ray, fMRI, or infrared images, as well as microscopic or telescopic depictions. Additionally, with regard to their visual appearance, images may greatly vary in the visual details that they preserve, ranging from high-resolution photographs to simple line drawings, from black and white renderings to nuanced color reproductions, or from single view static depictions to dynamic portrayals that capture a phenomenon’s changes across time and from different viewpoints. But no matter how restricted the visual fidelity of an image may be, usually it will still keep an iconic relationship to its referent scene.

Although, strictly speaking, iconicity implies a resemblance to real-world referents, the origin of a given image need not necessarily stem from a direct optical source. While in photographs and films a visual array is retained through optical and chemical processes, other types of images are constructed in a more indirect way by drawing, painting, or use of digital tools. As a consequence, such processes of construction may even portray scenes that have no current real-world counterpart in a realistic manner, as in the case of archaeological reconstructions or imagined future scenarios.

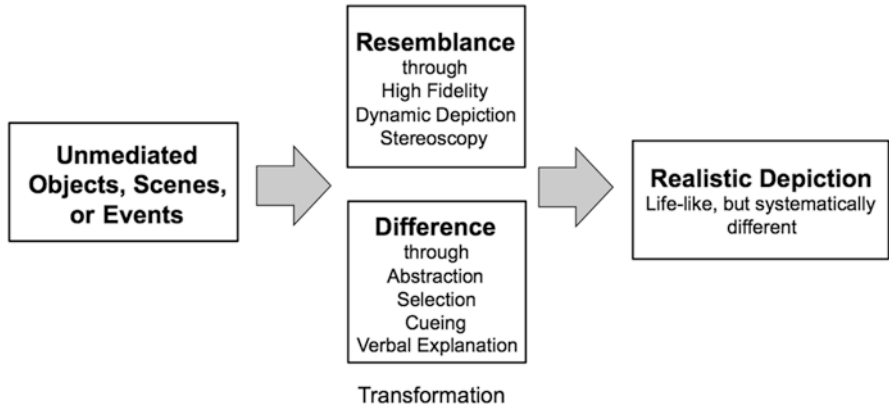


Fig. 3.1 Transforming unmediated objects, scenes, or events into realistic depictions

Due to the mentioned differences, digital images may, on the one hand, be legitimately considered as impoverished surrogates of real-world entities. But, on the other hand, by transforming a given phenomenon into a realistic pictorial representation, the status of the phenomenon is changed as well. From its pure existence, it is transformed into a document that may serve various epistemic purposes, including scientific reasoning and teaching. In particular, because of their loose coupling with reality, images may be tailored according to educational purposes. This is something that could not be accomplished under real-life conditions, for example, by selecting and simplifying content, by adding further layers of information, or by cueing learners’ attention. From this “dual character” of digital images (see Fig. 3.1), both closely resembling reality but simultaneously being systematically different, several questions arise regarding processes of knowledge acquisition that will now be discussed in turn.

Are Viewers Aware of the Differences Between Real-World Information and Mediated Information—And Do They Take Them Into Account?

First of all, do viewers readily notice the difference between a real, unmediated object, scene, or event and its pictorial representation—and do they take this difference into account for their information processing behavior? Like in Magritte’s famous painting of a pipe, entitled “Ceci n’est pas une pipe,” viewers should be aware that an image of an object has a visual resemblance to the real object, but lacks its functionality; that is, the picture may be looked at and inspected visually, but the depicted object cannot be used according to its real-world purposes. We addressed the cognitive implications of this difference in a number of studies in a

museum setting (Hampp & Schwan, 2014, 2015; Schwan, Bauer, Kampschulte, & Hampp, *in press*). We designed several display cabinets at the Deutsches Museum in Munich in which we either placed real objects or corresponding life-size photographs, together with additional material (texts, graphs) informing about nanotechnology and health technology. Although highly similar in their optical information, photographs were inspected less intensively than their real counterparts. Also, after a delay of about 1 h, visitors remembered fewer details of the exhibits if they had seen them as photographs and not as real objects. In line with these findings, a recent study by Sareen, Ehinger, and Wolfe (2015) showed that viewers differentiate ontologically even when inspecting photographs. Sareen et al. presented photos showing scenes of rooms filled with objects that also contained large mirrors reflecting these objects. They found that within the photographs, viewers paid more attention to the objects than to the objects' reflections in the mirror. Taken together, these results indicate that the perceived ontological status of a presentation (real object vs. photography) serves as a metacognitive cue that may modulate the amount of cognitive resources devoted to its processing. Surely, this presupposes that the ontological status of a given presentation can be easily perceived. While this is normally the case even for stereoscopic, three-dimensional representations such as virtual realities (because they rely on salient technical equipment), the ontological status may be blurred for three-dimensional, material reproductions of objects. Here, a new research field for the role of perceived authenticity for information processing and learning opens up.

Do Realistic Images Require Specific Competencies for Comprehension?

As discussed above, images should not be conceived as simple reflections of the real world, but differ from it in a more fundamental way with the introduction of new forms of depiction that have no real-world counterparts. This is particularly true for moving images because for them a repertoire of stylistic means have been developed, including film cuts, zooms, or slow motion, among other things, that provides perceptual experiences substantially deviating from conditions of natural vision. Hence, the question arises whether (static or moving) images require some additional visual literacy beyond the competencies used for natural real-world perception and cognition. Early studies by Hochberg and Brooks with children (1962) and Hudson (1967) with members of cultures that lack images have demonstrated that photographs or even line drawings of familiar objects normally are correctly perceived and identified. In contrast, other drawing conventions such as the inclusion of a horizontal line or placing distant objects in the upper part of a picture and showing them at smaller scale are often misinterpreted by viewers who are unfamiliar with pictures (Hudson, 1967). This indicates that, for appropriate interpretation of drawings, principles of natural perception are not sufficient and have to be complemented by some initial experience with pictorial representations.

Similar arguments also apply to perception of moving images. On the one hand, films are more realistic than static pictures because they additionally preserve temporal characteristics of scenes and events. But this is complemented by a set of cinematic techniques which introduce some substantial differences to conditions of real-world perception. Thus, as for static images, the issue of film-specific competencies is of relevance. We addressed this topic in two studies that we conducted with adults unfamiliar with film, living in a difficult-to-access mountain region in southern Turkey (Schwan & Ildirar, 2010; Ildirar & Schwan, 2015). In individual sessions at their homes, they were shown a set of short video clips, each containing a different type of common cinematic techniques such as a shot-reverse-shot, a temporal gap, an eye line match, or an establishing shot. The cinematic techniques were classified according to the relation between adjacent shots. Shots were linked either by visual, causal, or conceptual overlap. In the case of visual overlap, substantial parts of the scene (e.g., a salient object or person) were shown in both shots. In the case of causal overlap, shots were linked by sequences of activities (not necessarily implying visual overlap), while in the case of conceptual overlap, shots were linked on the basis of semantic relations (e.g., the front of a mosque followed by a prayer inside the mosque). By asking the viewers to describe each video immediately after presentation, we found that the viewers unfamiliar with film had no problem describing the individual shots, indicating that they had understood the objects and activities shown in the videos, but they were often unable to link the shots appropriately. Thus, only a small subset of filmic devices was intelligible to them, whereas a control group of viewers familiar with film and having a similar cultural background gave appropriate descriptions for the whole set of videos. Surprisingly, it was not visual overlap between shots that primarily contributed to immediate comprehension. Instead, shots linked by a sequence of familiar activities were most intelligible to those unfamiliar with film, suggesting that in moving images, the existence of a kind of familiar “story line” helps film novices to comprehend filmic techniques that are at odds with conditions of natural perception.

While filmic means constitute comprehension obstacles for film novices, experienced viewers typically do not show comprehension problems regarding cinematic techniques. In contrast, due to the high amount of time that viewers spend watching films or TV, viewers become so familiar with filmic means that they tend to go unnoticed. Therefore, another facet of visual literacy is to become aware of and critically reflect on filmic means in persuasive contexts like TV ads or propaganda films. This issue was addressed by Merkt and Sochatzy (2015) in two experiments. They found that ninth graders had problems to spontaneously identify persuasive visual film techniques such as the use of low or high camera angles that let persons appear to be powerful or powerless, respectively. Both by training and by cueing specific cinematic techniques during film presentation, the identification rate increased and also transferred to new films without such cues. Thus, in terms of knowledge acquisition from static or moving images, visual literacy goes well beyond the basic skills of identifying pictorial elements and events, and it also includes awareness of filmic techniques and their manipulative power.

Should Realism of Visual Representations Be Maximized for Learning?

At first sight, maximizing realism seems to be the natural strategy for design of images because the more life-like a pictorial representation is, the more it can serve as a substitute for real-life entities. But while a maximum of realism may be indicated for purposes of documentation, research has demonstrated that for learning and knowledge acquisition it might not be the best option. In particular, instead of presenting objects, scenes, or events in rich detail, abstraction by highlighting relevant aspects while leaving out irrelevant or accidental ones may make images serve better for learning (Gerjets, 2017). Therefore, in some studies viewers of simple line drawings outperformed viewers of photorealistic depictions in terms of learning and understanding, both for static (Dwyer, 1968) and dynamic visualizations (Scheiter, Gerjets, Huk, Imhof, & Kammerer, 2009).

Similarly, while films or animations preserve the temporal qualities of a procedure, an activity, or an event with high fidelity, this gain in temporal realism may be outweighed by the transience of the presentation, making it difficult to identify and process its individual steps (Tversky, Bauer Morrison, & Betrancourt, 2002). Depending on the specific learning task, this interplay of opposing factors may either favor the use of dynamic visualizations (like films or animations), or static ones (e.g., comic strip-like sequences of pictures). Accordingly, Lowe and Schnotz (2014) emphasize the fit between the requirements of the learning task and the preservation of corresponding dimensions of pictorial realism. For instance, for comprehending a sequence of several clearly distinguishable steps, a display of the temporal transitions is often not necessary. Therefore, for this type of learning task, several studies have shown that sets of static pictures can be at least as effective for learning as dynamic depictions (e.g., Hegarty, Kriz, & Cate, 2003). However, in other cases such as learning to reproduce a certain pattern of continuous movements, the specifics of temporal transitions require a higher temporal fidelity, thus making dynamic visualizations a more appropriate form of learning material.

A further facet of task appropriateness relates to the congruence of format of learning and format of testing. In a recent series of experiments, we asked participants to learn a set of kanji signs (Soemer & Schwan, 2016). We systematically varied presentation mode (static, static sequential, animated), task requirements (identifying the sign, knowing the stroke order, knowing drawing direction of the individual strokes), and testing mode (static, static sequential, animated). In the experiments, congruence of presentation mode and testing mode (i.e., static-static, static sequential—static sequential) was shown to have the strongest impact on learners' testing performance, well above compatibility of presentation mode and task requirements. Besides important practical consequences, the theoretical implication of these findings is to extend the notion of realism beyond the resemblance between real-world situation and visual presentation in the learning phase to also include the resemblance between visual representation and perceptual circumstances during testing.

Taken together, research both from line drawings and from animations demonstrates that in the realm of learning, pictorial realism is not a value in itself but must be considered in the light of specific learning goals and their information requirements. This is also true for recent advancements in realistic depictions such as stereoscopic presentations (Schwan & Papenmeier, [in press](#)). Stereoscopic presentations heighten realism by adding binocular disparity as a further depth cue. While this has been shown to be beneficial for training of complex manual tasks requiring eye-hand-coordination (e.g., medical surgery tasks), advantages for other types of learning content are still under debate. In a series of experiments, participants were presented molecule-like objects, either stereoscopically or monoscopically, which they afterwards had to recognize as accurately and as fast as possible, again either stereoscopically or monoscopically (Papenmeier & Schwan, [2016](#)). We found that learners benefited from stereoscopic presentation in the test phase, while in the learning phase, presenting the molecule-like objects as a continuously rotating animation turned out to be as effective as presenting them stereoscopically. Hence, while stereoscopic presentation enhances realism, its contribution beyond more traditional types of presentations (like the animations in the present case) seems to be limited.

Can Systematic Deviations from Realism Help Comprehension?

Above it has been shown that differences between real-world states and realistic depictions should not be regarded as deficiencies that have to be overcome by advanced technologies that provide a more and more perfect illusion of reality. Instead, in terms of comprehension and knowledge acquisition, deviations from reality may even be purposefully exploited for optimizing the content to be learned for perceptual and cognitive processing. In the past years, this topic has been systematically explored in our lab particularly for realistic dynamic visualizations such as animations and films. In particular, we were interested how the range of design options that dynamic visualizations provide for the portrayal of real-world activities and events may be used for fostering comprehension.

In a first set of experiments, we investigated how the structure of unfamiliar events or activities can be made more salient for viewers. Observers tend to spontaneously segment real-world activities like troubleshooting a machine or assembling a device into a series of discrete segments, separated by event boundaries. Identification of event boundaries and structuring an event accordingly has been shown to be an important prerequisite for event comprehension (Hard, Lozano, & Tversky, [2006](#)). This may pose a problem for viewers who are confronted with a new and unfamiliar event. By analyzing several educational movies produced for classroom presentation in Germany, we found that learners preferably placed event boundaries at the occurrence of formal filmic features such as film cuts (Schwan,

Garsoffky, & Hesse, 1998). In a laboratory study (Schwan, Garsoffky, & Hesse, 2000), we found that placing film cuts at natural event boundaries made the boundaries more salient to viewers who were not familiar with the activity. Additionally, use of film cuts at event boundaries increased recall of the event sequences shown in the films. These findings indicate that by informed use of film techniques, comprehension can be fostered by highlighting the structure of unfamiliar events or activities.

Natural event boundaries can also be used to make learning more efficient by producing event summaries instead of presenting an event over its whole course. From basic research on event cognition, it is known that content at event boundaries is processed more deeply and remembered better than content at non-boundary points in time (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). This indicates that observers tend to preferably select and memorize event boundaries as a kind of compact characterization of the corresponding event segment. Accordingly, preselecting these boundaries in event portrayals may serve as an effective event summary that condenses an event to its most important parts while leaving out irrelevant or redundant aspects. This hypothesis was confirmed in an experiment in which viewers were either shown complete records of events, event summaries consisting of film shots around event boundaries, or event summaries consisting of film shots around non-boundaries (Schwan & Garsoffky, 2004). We found that viewers of event-boundary summaries recalled largely the same event parts as the viewers of the complete event, whereas viewers' recall of the non-boundary summaries corresponded to a much lesser extent to the recall found in the complete event condition. These results indicate that by systematic deviation from real-world conditions, video recordings may forestall cognitive selection processes, thereby making learning more efficient.

Event presentations can also be optimized in terms of the visual perspective from which they are shown. It has been demonstrated that not only static spatial layouts, but also dynamically changing object constellations are mentally represented in a viewpoint-dependent manner (Garsoffky, Schwan, & Hesse, 2002). Also, not all viewpoints are equally well suited for recognition and recall. Instead, the so-called canonical viewpoints that maximize visibility of an object's or event's characteristic features have been shown to possess cognitive processing advantages over noncanonical views (Garsoffky, Schwan, & Huff, 2009). Again, this opens up a number of possibilities for designing realistic images for learning. In real-life presentations, viewing conditions on an object or phenomenon are often suboptimal due to a number of restrictions (e.g., distance is too large, object is partly occluded by other viewers, object is seen from an oblique viewing angle). In contrast, by appropriate choice of viewing distance and angle, images can present objects in an optimized manner.

For events, matters are more complicated because appropriate viewing position may frequently change during the course of event. Again, in real life, these changes are often difficult to carry out, while in videos staging, film techniques, and postproduction allow for adapting viewing position through an event's course. Yet, as

another set of experiments in our lab has shown, frequently changing viewpoints come also with some cognitive costs. In particular, abrupt changes of viewing position by film cuts is correlated with loss of spatial orientation and comprehension of spatial configurations, compared to static or continuously changing viewpoints (Garsoffky, Huff, & Schwan, 2007; Huff, Jahn, & Schwan, 2009; Meyerhoff, Huff, Papanmeier, Jahn, & Schwan, 2011). However, producers of instructional films or animations can counteract these problems by adhering to certain principles of film design. In particular, since the early times of Hollywood cinema, the so-called continuity editing rules have been established that tend to make the transitions between shots as unnoticeable (and thereby as intelligible) to the viewers as possible. Part of the continuity editing system is the centerline rule that regulates the viewing positions of adjacent shots, stating that changes in viewing perspective are easily processed as long as the camera stays on the same side of the main axis of action across the cut. In a recent study, we could demonstrate that viewers indeed spontaneously rely on this rule, shortcutting elaborate alignment processes in favor of a simple spatial heuristic which helped them keep spatially oriented across cuts at minimal cognitive processing demands (Huff & Schwan, 2012a). Again, the findings of these studies demonstrate how systematic deviations from perceptual conditions of real-life (in this case by use of film cuts introducing “unnatural” abrupt changes of viewing position) may be utilized for making the spatial structure of ongoing events comprehensible to learners.

As a final example, consider the temporal characteristics of events, which sometimes unfold at a speed which is difficult to handle perceptually or cognitively. This is true on both sides of the temporal scale, encompassing events that unfold at a very high or at a very slow speed (think of high speed collisions on the one hand or growth of plants on the other). Again, creating depictions that systematically deviate from a natural time scale may substantially facilitate learners’ comprehension of the underlying processes and mechanisms. In one study, we had learners watch for ten minutes a video showing the inner workings of a mechanical pendulum clock, either in real time or in fast motion (Fischer, Lowe, & Schwan, 2008). We found that viewers of the fast motion depiction better understood the basic principles of pendulum clocks in terms of the regulating role of both the clock’s weight and its anchor mechanism. This was because by speeding up the presentation, the operation of these elements became more salient to the participants, which in turn helped them to make more correct inferences about the underlying physical forces at play. In a second study, presenting the clockwork at a higher speed proved to be more effective for comprehension than highlighting the relevant elements of the clock’s mechanism by color coding (Fischer & Schwan, 2010).

Overall, the results of the described studies indicate that models of perception and cognition may inform design options that foster memory and comprehension by systematically introducing certain deviations from realism. These deviations include, among others, additional formal structures by cuts, optimization of viewing position by abrupt changes of perspective, or use of slow or fast motion to make dynamic changes during events more salient and comprehensible.

How to Deal with the Informational Complexity and Ambiguity of Realistic Images?

In a seminal study, Yarbus (1967) had viewers look at a picture of a family scene with different goals such as forming an impression of the depicted persons or understanding the activities taking place. Depending on the task at hand, the viewing patterns of the participants were quite different, indicating that their course of processing the picture did substantially differ. Put in more general terms, realistic images, both static and dynamic, usually contain an abundance of elements and details and are open for various different interpretations, thus leaving it up to the viewer which information to extract to answer a specific question or to solve a specific task. Yet, while images are inherently “goal-free” at first sight, particularly in educational contexts, they serve as tools for visual communication based on a specific didactic intention. Therefore, producers of learning materials face the task of guiding the viewers’ attention to those elements and attributes of an image that they consider relevant for the current learning goal. Also, cueing attention is even more pressing for dynamic images with their transient, rapidly changing visual content.

While multimedia research has focused mainly on overt forms of cueing important pictorial elements, including graphic signs such as arrows, color coding, shadowing or overlay of expert eye movements (see van Gog, 2014 for an overview), a number of design strategies already discussed in the previous section allow for more unobtrusive, covert means of guiding viewers’ spatiotemporal distribution of attention, including simplification of content by line drawings instead of photorealistic depictions or summarizing events by leaving out its non-boundary parts. Also, the mentioned principles of continuity editing in films have been interpreted as instruments for attention guidance (Smith, 2012) but have not been systematically related to learning and knowledge acquisition to date. Additionally, strategies of camera movement, like zooming-in or panning, are frequently found in educational movies (for instance, in the form of the so-called Ken Burns effect, where camera movements are used to visually explore static historical documents such as prints or photographs), but with few exceptions (e.g., Salomon & Cohen, 1977), the analysis of their effects on learning and understanding still awaits systematic empirical research.

New technologies also open up innovative strategies for scaffolding viewers’ attention. For instance, autostereoscopic displays allow viewers to switch between different pairs of similar images by slight movements of head (similar to a vexing image) without any necessity for recalibrating attention (e.g., by means of a saccade). In one study, we asked participants to solve a structural task with pairs of different visualizations of complex proteins, namely, a stick-and-ball model and a wireframe model of the protein (Huff, Bauhoff, & Schwan, 2012). The pairs of visualizations were either presented side-by-side, overlaid, or via vexing image display. We found that particularly viewers with low spatial abilities benefitted significantly from the vexing image condition, which helped them to identify corresponding parts of the molecules, thereby minimizing detrimental split attention effects.

Finally, adding written or spoken explanations is a further common strategy of shaping viewers’ processing and interpretation of (static or moving) images, rang-

ing from audio guides in museums to narrators in educational films and teacher's explanations in classrooms. Regarding the interplay of text and pictures, several models have been proposed, both in cognitive psychology (Schooler, & Engstler-Schooler, 1990; Yee & Sedivy, 2006) and in educational psychology (Mayer, 2001), that build a theoretical basis for addressing issues of verbal guidance. For instance, verbal overshadowing research indicates that giving a verbal explanation after presentation of an image may decrease the accuracy of memory for pictorial details because the more abstract verbal description interferes with the more concrete visual representation (Schooler, & Engstler-Schooler, 1990). We confirmed and extended this effect in several studies, showing that memory for visual details of an event decreased if its observation was followed by a verbal description (verbal overshadowing), but in contrast, giving a verbal description before observing an event facilitated memory for visual event details instead, presumably because the verbal description serves as an abstract scheme in which visual event details can subsequently be integrated (verbal facilitation; Huff & Schwan, 2008, 2012b, see also Eitel & Scheiter, 2015, Scheiter, Schüler, & Eitel, [this volume](#)).

Most often, verbal descriptions and explanations are given not prior or after, but concurrently to viewing a picture. Here, models of multimedia learning (Mayer, 2001) provide a well-established framework of analysis. By assuming separate processing channels for visual and audio information, they posit that, in general, presenting both pictorial and verbal information should lead to better learning and understanding than relying on just one type of information (multimedia effect) and that learning also benefits from complementing pictures with spoken instead of written information (modality effect).

While traditional formulations of the multimedia model treat pictures as illustrative text supplements and ask primarily how the addition of pictures may foster text comprehension, we were interested in the complementary issue, treating text as a complement to images and asking how the addition of text may foster processing of realistic images (Glaser & Schwan, 2015). Working with (fictitious) depictions of reconstructed historical buildings accompanied by audio-guide-like explanations, we found that mentioning a pictorial element in the text led viewers immediately to turn their attention to that element, as indicated by a highly synchronous fixation pattern across participants. Also, in a subsequent memory test, these pictorial elements were better recalled than pictorial elements that were not mentioned in the audio explanation. Overall, the results of the Glaser and Schwan (2015) study indicate that the multimedia effect can be extended, now stating that a combination of pictures and texts fosters learning better than either text or picture alone. Additionally, the findings suggest that accompanying pictures with text provide a twofold advantage both by guiding attention to relevant pictorial elements and by linking these pictorial elements with additional text information. Also, a similar extension from text to images was demonstrated for the modality principle (Dutz & Schwan, 2014). In an experimental art exhibition, the artworks were accompanied by explanations either as labels placed beside each work, as digital text on an iPad, or as spoken text in an audio guide. In line with the modality principle, we found that memory for pictorial details of the artworks was best for audio guide, compared to both text on a label or text on an iPad.

Taken together, these results demonstrate that besides covert visual design principles and overt graphical cues, accompanying (spoken) text does play a major role for guiding viewers' attention in complex realistic pictures. Yet, while the addition of spoken text to pictures has been shown to be beneficial for learning, casual observation from various audio guides in museums indicates that, depending on its linguistic features, texts may fulfill this purpose to a higher or lesser degree. Future research will have to determine in more detail what kinds of texts (in terms of organization, formulation, etc.) will be suited best for helping viewers to scrutinize and interpret a given complex image.

Realistic Images and the Active Learner

A last question pertains to overt learning activities that realistic images may afford, which is of particular importance for moving images as in educational movies, TV documentaries, or video-based lectures from the Internet. Typically lasting several minutes to even an hour or more, they each provide a high amount of densely packed and transient content that the learner has to deal with. Whereas voluminous static media such as books or comics allow readers to inspect their content at will, regulating pace and sequence of their information intake by decreasing or increasing reading speed, rereading difficult parts, or skipping back and forth between pages, possibilities for active regulation of information is significantly more restricted for traditional forms of moving images. For example, besides starting and stopping from time to time, educational films are typically screened in classrooms via VCR without many intervening activities by the teacher or students (Hobbs, 2006).

Digital videos offer new possibilities for individualized, active learning. They allow viewers to regulate their information intake in ways similar to reading books, including stopping the presentation, changing presentation speed (analogous to decrease or increase reading speed), or viewing parts of the video several times (analogous to rereading a text passage several times; Merkt, Weigand, Heier, & Schwan, 2011). Thus, viewers of digital videos are offered advanced opportunities for information acquisition that observers of real-life situations typically lack: Most often, a real-life event cannot be easily slowed down, stopped, or repeated at will. This advantage of interactively viewing a realistic event over its noninteractive observation has also been demonstrated experimentally. Using nautical knots as an example, learners of interactive videos demonstrating the tying of various nautical knots spontaneously used the available features (such as stop, rewind, slow, or fast motion) and thereby outperformed learners of respective noninteractive videos in efficiency of learning to tie the respective knot (Schwan & Riempp, 2004).

Matters get more complicated if we turn from simple recordings of real-life events or activities (like tying a nautical knot) to realistic audiovisual presentation that depict more complex matters, as with explaining historical developments by use of authentic news footage or demonstrating principles of physics with filmed experiments while making extensive use of filmic design features (such as changes

of perspective, skipping irrelevant event episodes, and so on) discussed in the previous sections. First, it can be argued that such dynamic audiovisual material has already been optimized for learning by its authors, reducing the necessity for interactive control. In line with this notion, we found that for TV documentaries of principles of physics, learners spontaneously built causal bridging inferences during viewing, despite the videos' fast pace and transience of information (Tibus, Heier, & Schwan, 2013).

Second, it can also be argued that complex educational films require interaction opportunities that go beyond local regulation of information intake by control of presentation speed and allow for direct access of information within a given video. In order to address this issue, we implemented additional tools for information access into an educational film about post-war Germany that were analogous to those in textbooks such as the table of content and alphabetical register (Merkt et al., 2011). Both in the laboratory and in a classroom setting, we found that students used these interactive options only to a small degree. Accordingly, while the interactive features did help students to quickly locate certain bits of information in the film, it did not substantially improve the quality of students' essays about the film's topic (Merkt et al., 2011). In a second study, students' lack of appropriate usage strategies could be identified as one important reason for these findings (Merkt & Schwan, 2014). Thus, if interactive features like a table of contents and an alphabetical register were available and students were trained in active use of videos, these features would not only be used to a substantial degree but also would increase the quality of the students' essays about the film's historical content.

Taken together, the reported findings indicate that introducing possibilities for interaction into realistic moving images is not as simple as it seems at first sight. On the one hand, interactivity gives learners new ways of controlling a video's flow of information, thereby adapting it to her individual cognitive needs. On the other hand, control options not only presuppose some knowledge and skills for appropriate use, but also require additional mental resources for planning and execution (Scheiter, 2014). Also, in contrast to real-life events, realistic moving images may already be designed for learning and knowledge acquisition without the necessity of leaving optimizing information presentation up to the learner herself. For example, by extending film shots or by introducing explicit brief pauses, information density can be reduced and effects of transience minimized without requiring viewers to plan and execute pauses themselves. Further research on the interplay of shaping information presentation by film design versus by the learners' individual activity is needed, in particular, against the backdrop of a proliferation of video material on the Internet.

Conclusion

Due to the digitalization of everyday life, today we face a continuous blurring of the distinction between the real and the virtual. As part of this process, digital types of realistic images including photographs, videos, or virtual reality renderings play an

increasingly important role in learning scenarios. This is particularly true for informal learning settings such as museums, television, or the Internet, where realistic visualizations offer the opportunity to present content in a vivid, motivating, and comprehensible way (Glaser, Garsoffky, & Schwan, 2009, 2012; Schwan, Grajal, & Lewalter, 2014; Töpfer, Glaser, & Schwan, 2014). While research on differences and commonalities between text and pictures has a long-standing tradition in educational psychology, this chapter aimed to outline a complementary perspective of comparing realistic depictions with conditions of real-life experience. Despite all its similarities, digital images (both static and moving) are not simply reflections of reality, but should instead be seen as purposefully designed modes of presenting information in a comprehensible way. This can be achieved by numerous strategies, ranging from carefully staging objects or events to use of cinematographic techniques to addition of guiding cues or supplementary explanations. This blending of realism with didactical design not only leads to unique forms of learning material that stand in-between real-life experiences and symbolic forms of information presentation (like texts or graphs), but opens up a number of fundamental research questions regarding the relationship between vivid life-like experiences and processes of learning and knowledge acquisition.

References

- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review Psychology*, 63, 1–29.
- Dutz, S., & Schwan, S. (2014). *New media use in art exhibitions: Enriching or annoying?* Paper presented at the 27th Annual Visitor Studies Association Conference, Albuquerque, USA.
- Dwyer, F. M. (1968). Effect of varying amount of realistic detail in visual illustrations designed to complement programmed instruction. *Perceptual and Motor Skills*, 27, 351–354.
- Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, 27, 153–180.
- Fischer, S., Lowe, R. K., & Schwan, S. (2008). Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Applied Cognitive Psychology*, 22, 1126–1141.
- Fischer, S., & Schwan, S. (2010). Comprehending animations. Effects of spatial cueing versus temporal scaling. *Learning and Instruction*, 20, 465–475.
- Garsoffky, B., Huff, M., & Schwan, S. (2007). Changing viewpoints during dynamic events. *Perception*, 36, 366–374.
- Garsoffky, B., Schwan, S., & Hesse, F. W. (2002). Viewpoint dependency in the recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 1035–1050.
- Garsoffky, B., Schwan, S., & Huff, M. (2009). Canonical views of dynamic scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 17–27.
- Gerjets, P. (2017). Learning and Problem Solving with Hypermedia in the 21st Century: From Hypertext to Multiple Web Sources and Multimodal Adaptivity.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Glaser, M., Garsoffky, B., & Schwan, S. (2009). Narrative-based learning: possible benefits and problems. *The European Journal of Communication Research*, 34, 429–447.

- Glaser, M., Garsoffky, B., & Schwan, S. (2012). What do we learn from docutainment? Processing hybrid television documentaries. *Learning and Instruction, 22*, 37–46.
- Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology, 107*, 1006–1018.
- Hampf, C., & Schwan, S. (2014). Perception and evaluation of authentic objects: Findings from a visitor study. *Museum Management and Curatorship, 29*, 349–367.
- Hampf, C., & Schwan, S. (2015). The role of authentic objects in museums of the history of science and technology: Findings from a visitor study. *International Journal of Science Education Part B: Communication and Public Engagement, 5*, 161–181.
- Hard, B. M., Lozano, S. C., & Tversky, B. (2006). Hierarchical encoding of behavior: Translating perception into action. *Journal of Experimental Psychology: General, 135*, 588–608.
- Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*, 325–360.
- Hobbs, R. (2006). Non-optimal uses of video in the classroom. *Learning, Media and Technology, 31*, 35–50.
- Hochberg, J., & Brooks, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *The American Journal of Psychology, 75*, 624–628.
- Hudson, W. (1967). The study of the problem of pictorial perception among unacculturated groups. *International Journal of Psychology, 2*, 89–107.
- Huff, M., Bauhoff, V., & Schwan, S. (2012). Effects of split attention revisited: A new display technology for troubleshooting tasks. *Computers in Human Behavior, 28*, 1254–1261.
- Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition, 17*, 297–306.
- Huff, M., & Schwan, S. (2008). Verbalizing events: Overshadowing or facilitation? *Memory & Cognition, 36*, 392–402.
- Huff, M., & Schwan, S. (2012a). Do not cross the line: Heuristic spatial updating in dynamic scenes. *Psychonomic Bulletin & Review, 19*, 1065–1072.
- Huff, M., & Schwan, S. (2012b). The verbal facilitation effect in learning to tie nautical knots. *Learning and Instruction, 22*, 376–385.
- Ildirar, S., & Schwan, S. (2015). First-time viewers' comprehension of films: Bridging shot transitions. *British Journal of Psychology, 106*, 133–151.
- Institut für Demoskopie Allensbach (2013). Digitale Medien im Unterricht (Digital media in the classroom). Retrieved from http://www.ifd-allensbach.de/uploads/tx_studies/Digitale_Medien_2013.pdf
- Lee, V. R. (2010). Adaptations and continuities in the use and design of visual representations in US Middle School science textbooks. *International Journal of Science Education, 32*, 1099–1126.
- Lowe, R., & Schnotz, W. (2014). Animation principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 513–546). Cambridge, MA: Cambridge University Press.
- Mager, U. I., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based environments. *Learning and Instruction, 29*, 141–152.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge, MA: Cambridge University Press.
- Merkt, M., & Schwan, S. (2014). Training the use of interactive videos: Effects on mastering different tasks. *Instructional Science, 42*, 421–441.
- Merkt, M., & Sochatzy, F. (2015). Becoming aware of cinematic techniques in propaganda: Instructional support by cueing and training. *Learning and Instruction, 39*, 55–71.
- Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning & Instruction, 21*, 687–704.
- Meyerhoff, H. S., Huff, M., Papenmeier, F., Jahn, G., & Schwan, S. (2011). Continuous visual cues trigger automatic spatial target updating in dynamic scenes. *Cognition, 121*, 73–82.
- Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York: Oxford University Press.

- Papenmeier, F., & Schwan, S. (2016). If you watch it move, you'll recognize it in 3D: Transfer of depth cues between encoding and retrieval. *Acta Psychologica, 164*, 90–95.
- Peirce, C. S. (1940). *The philosophy of Peirce: Selected writings*. New York: Harcourt.
- Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*, 216–237.
- Salomon, G., & Cohen, A. A. (1977). Television formats, mastery of mental skills, and the acquisition of knowledge. *Journal of Educational Psychology, 69*, 612–619.
- Sareen, P., Ehinger, K. A., & Wolfe, J. M. (2015). Through the looking-glass: Objects in the mirror are less real. *Psychonomic Bulletin & Review, 22*, 980–986.
- Scheiter, K. (2014). The learner control principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 487–512). Cambridge, MA: Cambridge University Press.
- Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*, 481–494.
- Scheiter, K., Schüler, A., Eitel, A. (this volume). Learning from Multimedia: Cognitive Processes and Instructional Support.
- Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*, 101–120.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36–71.
- Schwan, S., Bauer, D., Kampschulte, J., & Hampf, C. (in press). Presentation equals representation? Photographs of objects received less attention and are less well remembered than real objects. *Journal of Media Psychology*.
- Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology, 18*, 37–55.
- Schwan, S., Garsoffky, B., & Hesse, F. W. (1998). The relationship between formal filmic means and the segmentation behavior of film viewers. *Journal of Broadcasting & Electronic Media, 42*, 237–249.
- Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition, 28*, 214–223.
- Schwan, S., Grajal, A., & Lewalter, D. (2014). Understanding and engagement in places of science experience: Science museums, science centers, zoos, and aquariums. *Educational Psychologist, 49*, 70–85.
- Schwan, S., & Ildirar, S. (2010). Watching film for the first time: How adult viewers interpret perceptual discontinuities in film. *Psychological Science, 21*, 970–976.
- Schwan, S., & Papenmeier, F. (in press). Learning from animations: From 2D to 3D? In R. Plötzner & R. Lowe (Eds.), *Learning from dynamic visualizations: Innovations in research and application*. New York: Springer.
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning & Instruction, 14*, 293–305.
- Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections, 6*, 1–27.
- Soemer, A., & Schwan, S. (2016). Task-appropriate visualizations: Can the very same visualization format either promote or hinder learning depending on the task requirements? *Journal of Educational Psychology*.
- Stafford, B. M. (1994). *Artful science. Enlightenment entertainment and the eclipse of visual education*. Cambridge, MA: MIT Press.
- Tibus, M., Heier, A., & Schwan, S. (2013). Do films make you learn? Inference processes in expository film comprehension. *Journal of Educational Psychology, 105*, 329–340.
- Töpper, J., Glaser, M., & Schwan, S. (2014). Extending social cue based principles of multimedia learning beyond their immediate effects. *Learning and Instruction, 29*, 10–20.
- Tversky, B., Bauer Morrison, J., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262.
- van Gogh, T. (2014). The signaling (or cueing) principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge, MA: Cambridge University Press.

- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Yasar, O., & Seremet, M. (2007). A comparative analysis regarding pictures included in secondary school geography textbooks taught in Turkey. *International Research in Geographical and Environmental Education*, 16, 157–187.
- Yee, E., & Sedivy, J. C. (2006). Movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1–14.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127, 3–21.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: S mind-brain perspective. *Psychological Bulletin*, 133, 273–293.