Chapter 2
Learning from Animations: From 2D to 3D?

Stephan Schwan and Frank Papenmeier

2.1 Introduction

Although traditional animations usually present their content in a two-dimensional manner, there is a growing body of dynamic visualizations that make use of three-dimensional depictions. For the learning of science in particular, many topics demand comprehension of events unfolding in space, ranging from operations of sophisticated machines to chemical reactions of large organic molecules (cf. Jenkinson, 2017, this volume; McGill, 2017, this volume). Such topics often also require comprehension of complex three-dimensional objects based on their inspection from different sides, be they anatomical structures (cf. Berney & Bétrancourt, 2017, this volume) or archaeological artefacts, among others. This has raised questions as to whether the introduction of three-dimensional space in animations fosters learning and knowledge acquisition and how 3D animations should be designed in order to achieve their goals (Dalgarno & Lee, 2010).

The present chapter gives an overview of the role of three-dimensional animations in learning. It starts by characterizing three-dimensional animations as part of a larger transformation of animation production from conventional, drawing-based animations to animations that are generated by applying certain computational methods to numerical data sets. In the main part of this chapter, the scope of three-dimensional animations is described and classified, with three characteristics of three-dimensional animations and their implications for memory and learning being discussed in finer detail. First, the introduction of three-dimensional digital

S. Schwan (✉)
Leibniz-Institut für Wissensmedien, Tubingen, Germany
e-mail: s.schwan@iwm-tuebingen.de

F. Papenmeier
University of Tübingen, Tubingen, Germany
e-mail: frank.papenmeier@uni-tuebingen.de

depictions of objects and scenes opens up a broader range of animation possibilities than are available with traditional two-dimensional animations. These possibilities include both the dynamics of content (e.g., a machine in motion) and the dynamics of visual presentation (e.g., continuous movement of a virtual camera). Second, conceptual differences between three-dimensional monoscopic and stereoscopic presentations are described. In particular, monoscopic presentations (also called 2.5D, pseudo-3D, or synoptic) project three-dimensional content onto a two-dimensional plane (e.g., a computer screen). Because both of the viewer's eyes receive the same information, a truly three-dimensional impression is not possible. In contrast, in stereoscopic (3D) presentations, a viewer's left and right eye receive two slightly different views, creating the impression of true three-dimensional depth. Third, three-dimensional presentations also have a profound impact on users' modes of interaction: Besides being able to control the temporal aspects of an animation by actions such as starting, stopping, or rewinding, learners may also determine the animation's spatial characteristics by controlling position and movement of the virtual camera. The chapter ends by considering the implications of these innovations for animation research and practice.

2.2 From Two-Dimensional to Three-Dimensional Animations

Production of animations has undergone major transformations during the last two decades. Traditionally, animated cartoons were laboriously created on a frame-by-frame basis, similar to the painting of sequences of thousands of individual pictures in classical Disney cartoon films (Johnston & Thomas, 1981). With the advent of software tools such as Adobe Flash (first release in 1997), this process was substantially simplified and mechanized. Now, individual graphic objects could be digitally defined and animated. For example, by specifying the start and end point of an object's movement, the software could automatically determine and render the object's intermediate positions along a path. The appearance of objects in terms of size or color could also be easily transformed. By means of these possibilities, complex digital animations could be built out of simple graphic elements and commands. Yet, these software tools were based on the metaphor of visually presenting information on a flat plane, like a canvas or a sheet of paper. Accordingly, most animations built with these tools were two-dimensional in nature. In contrast, the current generation of software tools for animation design, like Unity or Blender, supplements previous ones by employing the metaphor of space that contains voluminous movable objects that extend in three-dimensions. Additionally, these tools not only allow the specification of three-dimensional objects and spatial layouts but also include so-called 'physics engines' that model the kinematics of objects according to predetermined physical laws.

With respect to space, recent animation tools allow digital objects to be defined as volumes (instead of two-dimensional shapes) on the basis of three-dimensional coordinates – width, height, and depth (cf. Jenkinson, 2017, this volume; McGill, 2017, this volume). Therefore, the projection screen now defines a window to a scene extending into depth. This move towards three-dimensionality has been accompanied by innovations in hardware technology that allow for a stereoscopic presentation, further enhancing viewers' impressions that they are perceiving actual spatial depth. With respect to dynamics, the digital definition of objects as sets of numerical values and coordinates allows for continuous transformations both of the objects themselves, (e.g., as morphs from one shape to another; Soemer & Schwan, 2012), of the movement of object parts, and of the whole objects' trajectories in three-dimensional space, according to predetermined mechanical or biological principles. With respect to camera position and perspective, recent animation tools allow for an easy definition of virtual cameras including their position, lens, and rotation, specifying both the distance and angle from which an animated scene is presented to the viewer. In addition, cameras may be set into motion, allowing for complex camera movements during the course of an animation. The net result is that animations made using these new facilities converge with films with their complex repertoire of established design principles (Bordwell & Thompson, 1979).

2.3   Types of Three-Dimensional Expository Animations

The observed trend towards using three-dimensional animations instead of two-dimensional ones has led to a broadened range of different animation types. In their meta-analysis of research on learning with animations, Höffler and Leutner (2007) had not yet referred to the distinction between two-dimensional and three-dimensional animations. Similarly, Ploetzner and Lowe (2012) have provided a detailed and comprehensive classification of animations used in research until then, where again the question of animations' spatial structure played only a minor role. This reflects the fact that to date, the animated material used in empirical research has not made much use of three-dimensional opportunities. But in order to accommodate probable future developments, further differentiations of the taxonomy's spatial dimension seem to be necessary. These differentiations relate to both the three-dimensional structure and the three-dimensional presentation of animations. While the three-dimensional structure concerns the addition of a third dimension to animations as well as the placement and motion of the virtual camera towards the three-dimensional scene, the three-dimensional presentation concerns the distinction between monoscopic (2.5D) and stereoscopic (3D) presentations of space.

2.4   Three-Dimensional Structure and Dynamic Camera Viewpoints

Most often, conventional expository animations have been used to show dynamically unfolding events from a fixed, stationary point of view. Examples range from piano mechanics (Boucheix, Lowe, Putri, & Groff, 2013; Lowe & Boucheix, 2017, this volume) and pendulum clocks (Fischer, Lowe, & Schwan, 2008), to organic systems (de Koning, Tabbers, Rikers, & Paas, 2007) and intercellular processes (Huk, Steinke, & Floto, 2010) to biological movement patterns (Imhof, Scheiter, Edelmann, & Gerjets, 2012; Lowe, Schnotz, & Rasch, 2011). Following Tversky, Morrison, and Bétrancourt (2002), one can ask for which content and under which conditions the addition of a third dimension to animations conforms to the principles of congruence and apprehension. According to these principles, the structure and content of an external representation should correspond to the desired structure and content of the internal representation (congruence principle) and be readily and accurately perceived and comprehended (apprehension principle).

Compared to two-dimensional depictions, adding a third dimension heightens the complexity of an illustration because spatial relations between the various elements have to be coded on three instead of two axes. Three-dimensional depictions also typically introduce occlusions and foreshortening (i.e. optical distortions of objects extending along the depth axis). In some circumstances, occlusions may help to understand otherwise unavailable spatial relationships. On the other hand, both occlusions and foreshortenings carry the danger of making understanding of the presentation more difficult for the learner. Thus if, for example, the elements of a mechanical system can be neatly arranged in two dimensions without loss of information, a two-dimensional depiction of the mechanical system in motion should be preferred over a three-dimensional one for ease of comprehension. In contrast, if the three-dimensional arrangement of the relevant elements in operation carries important information that cannot by easily depicted in two-dimensional graphics, and if this knowledge of this spatial structure should form an important element of the learner's mental model, a three-dimensional visualization should be preferred over a two-dimensional one for reasons of congruence.

When using three-dimensional animations, care must be taken to identify a viewpoint from which the relevant elements in operation can best be seen, avoiding occlusions and extreme foreshortenings. This often leads to an oblique viewpoint, showing the animated events not from a frontal perspective, but instead at an angle between about 30° to 60° degrees (Fischer, Lowe, & Schwan, 2008; Huk et al., 2010). This corresponds to the notion of canonical views as introduced by Palmer, Rosch, and Chase (1981; see also Blanz, Tarr, & Bülthoff, 1999). Compared to other viewing perspectives, canonical ones maximize the number of an object's visible surfaces and the visibility of its characteristic parts. Therefore, objects presented from canonical views are more accurately and easily identified than from other, non-canonical views.

To date, three-dimensional animations of dynamically unfolding events are a rare exception in the empirical literature on learning with animations. Following Tversky et al.'s (2002) principle of apprehension, at least for conventional animations with a fixed viewpoint, researchers seemingly have tended to avoid the additional complexity of a three-dimensional depiction in favor of an easier to grasp two-dimensional variant. Accordingly, few studies have systematically compared such two-dimensional and three-dimensional animations of similar content, finding at best mixed evidence for an advantage of the latter one (Huk, 2006; Huk et al., 2010). Still, under what circumstances – that is, for what content, what learning tasks, and what kind of learners – three-dimensional animations with a fixed viewpoint may better support the learners than two-dimensional ones is an open question that has to be addressed in future studies. For example, studies with static material indicate that three-dimensional depictions may be particularly suited for shape identification and discrimination but not for identification of relative positions in space (St. John, Cowen, Smallman, & Oonk, 2001). Also, adequate interpretation of three-dimensional depictions seems to require a high level of spatial ability (Huk, 2006; Huk et al., 2010; Khooshabeh & Hegarty, 2010).

It can be argued that if a dynamic event is simple enough or 'flat enough' to be intelligible from one stationary viewpoint, making the depiction three-dimensional will add little or nothing to its comprehensibility. In contrast, if a dynamic event is more complex and the interplay of its elements takes place not only in a flat plane, but extends into space, three-dimensionality may substantially enhance intelligibility.

Comprehensibility of an event taking place in space may also be facilitated by introducing a dynamic change of viewpoint. Accordingly, going three-dimensional has introduced a second important class of animations that show objects or scenes from changing viewpoints instead. Further, changing viewpoints are not only used for depicting dynamic events but also for depicting spatially extended static objects or scenes. Here, the impression of dynamics is not due to a moving or changing object but instead due to the observer's viewpoint (brought about by the <u>virtual camera</u>) moving through three-dimensional space. With regard to expository animations, such a moving viewpoint may serve a number of different purposes. Accordingly, several types of expository camera movements can be distinguished, including movement for completeness, for establishing connections, for regulating the focus of attention, and for decorative purposes.

<u>Camera Movement</u> for Completeness In many educational contexts, learners have to develop an appropriate mental representation of complex, three-dimensional objects, be they molecules, anatomical structures, or reconstructions of archaeological artifacts or buildings. In all these cases, inspecting the target object from one side alone may not be sufficient to fully understand its elements and their spatial relations because from a given viewpoint, relevant parts may be located on a hidden side or be occluded by other elements. Also, relative to a given viewpoint, visibility of surface planes extending into depth may suffer from foreshortening. To avoid these problems and allow the learner to make a comprehensive inspection of the

object, an animation may present a 360° circular movement of the camera around an object.

Basic research has demonstrated that mental representation of objects and scenes is largely viewpoint dependent (Diwadkar & McNamara, 1997; Tarr, 1995); that is, viewers do not normally tend to develop an abstract, viewpoint independent representation but instead store a set of individual views. On later occasions, in order to identify the object or scene from a novel view, viewers start with the stored view that most closely matches the novel one and try to align both views by mental rotation. The more discrepant the two views are, the longer this process takes and the more error prone it becomes. (Diwadkar & McNamara, 1997). Therefore, the more different viewpoints of an object or scene to-be-learned that are presented to a learner, the more flexible his or her resulting mental representation will be. This finding holds not only for static objects and scenes, but also for events that dynamically unfold in space. Here again, presenting the event from different viewpoints facilitates identification from novel perspectives, indicating a more flexible mental representation (Garsoffky, Schwan, & Hesse, 2002).

Because a particularly dense variety of views result from a continuous movement around an object or a scene, providing an animation that offers such movement conforms to Tversky et al.'s (2002) congruence principle. Also, because continuous change of viewpoint around an object or a scene is in accordance with everyday experience, it can be assumed to conform to Tversky et al.'s (2002) apprehension principle as well. However, as research from the field of anatomy learning has demonstrated, these assumptions hold only for learners with sufficient spatial ability (Garg, Norman, Spero, & Maheswari, 1999; Nguyen, Nelson, & Wilson, 2012). Both Garg et al. (1999) and Nguyen et al. (2012) found that learners with low levels of spatial ability benefited from the presentation of a small set of key views (similar to canonical views) as opposed to a large, comprehensive set of views interconnected via continuous and uniform camera movements. It seems that low spatial ability learners find difficulties in combining the dense set of views into an integrated mental representation, possibly due to the transience of viewpoint-specific information that imposes high processing demands on working memory.

In order to reduce processing demands while providing learners with animations encompassing a larger sample of views, thus balancing completeness of presentation with required processing resources, several design options come to mind. One solution could be to substantially slow down the speed of camera movement, while another, discussed by Garg, Norman, Eva, Spero, and Sharan (2002), could be to provide a small set of key views and let the camera "wiggle" around these views within a range of about 10° to provide some additional three-dimensional information. A third option, which will be discussed further below, is to give the learners the opportunity to interactively control speed and trajectory of viewpoint position.

Finally, while for some topics circular movements around an object or scene tend to provide learners with a more complete impression of the content, other topics such as astronomy, geography, or archaeology require continuous movement of camera along a linear path (for an example from astronomy see Eriksson, Linder, Airey, & Redfors, 2014). To date, little is known whether the provision of continuous

camera movements following a given trajectory instead of a set of distinct, but over-lapping views indeed leads to a better understanding of the respective content.

Camera Movement for Viewpoint Optimization Many instances of dynamically unfolding events can be decomposed into a sequence of individual steps. Think, for example, of the assembling of a machine along a production line or of the process of digestion along the gastrointestinal tract. Learning about and understanding such events requires building a mental model based on the comprehension of the individual steps and linking them according to underlying principles of causality (Lowe & Boucheix, 2008; Narayanan & Hegarty, 2002). While for some events or processes a single viewpoint may suffice for all steps to be intelligible to a viewer, other events may require shifts of viewpoints during the presentation's time course in order to present each step from an optimal perspective. This may be achieved by moving the virtual camera along a predefined path, stopping at certain moments at particular points that provide viewers with a privileged sight of the current step of the event.

Empirical evidence for this design strategy comes from studies that demonstrate the processing advantages of canonical views both for individual objects (Palmer, Rosch, & Chase, 1981; Blanz, Tarr, & Bülthoff, 1999) and for ongoing events (Garsoffky, Schwan, & Huff, 2009). Compared to other views, canonical views provide an optimal perspective on an object or scene, as manifested by viewers' preferences and also by memory advantages. In case of events, views perpendicular to an event's main axis of change or movement have been shown to be beneficial for processing and therefore to qualify as canonical views (Garsoffky et al., 2009). Because the main axis of movement may shift during the course of an event canonical views should shift accordingly. In conventional films, switching from one canonical view to another is typically achieved by abrupt viewpoint changes in form of film cuts. This is partly due to the fact that for real world film recordings continuous camera movements are difficult to create. In contrast, numerical definition of objects and events via digitalization allows for creating animations in which even complex predefined camera movements are easily implemented. Therefore, although both film cuts and continuous camera movements have become equally viable options for building animations, several empirical comparisons have provided evidence in favor of camera movements. For example, some learning topics require observers to simultaneously pay attention to several moving objects, like molecules in a chemical reaction, or players' moves on a playing field. Here, basic research has demonstrated that continuous movement of observers' viewpoint does hardly impede the attentional tracking of several moving objects (Meyerhoff, Huff, Papenmeier, Jahn, & Schwan, 2011), while film cuts do (Huff, Jahn, & Schwan, 2009). Also, a study of Garsoffky, Huff, and Schwan (2007) showed that memory for a complex dynamic event (an animated scene from a basketball game) was significantly higher for continuous compared to abrupt in-between changes of viewpoint induced by film cuts.

While camera movements for completeness typically deal with static objects or scenes, camera movements for viewpoint optimization can include both camera

motion *and* motion of objects or object parts. Therefore, learners have to disentangle both types of movement in order to comprehend the mechanism or event to be learned. Findings from Liu et al. (2005) indicate that during perception, viewers are successful in separating even extreme movements of whole scenes (due to camera pans or rotations) from relative movements of objects within that scene. But, on the other hand, in these studies, tracking multiple objects is so demanding that only little scene related information is processed and elaborated (Jahn, Papenmeier, Meyerhoff, & Huff, 2012), casting doubts on the appropriateness of such types of animations for learning. Also, while changing viewpoints during an event sequence may provide an optimized view for each step of the event, fostering comprehension of individual event steps, it also implies that different steps are seen from different viewpoints, possibly making it more difficult for the learner to appropriately link these steps causally in his or her mental model. Therefore, in terms of the Animation Processing Model proposed by Lowe and Boucheix (2011, 2017, this volume), viewpoint optimization by camera movement may facilitate parsing of the event into discrete steps (Phase 1) and the local processing of these steps (Phase 2), but may prove detrimental for connecting those steps into a causal chain (Phase 3). However, to our knowledge, to date no empirical research from the field of instructional design has addressed the topic of learning dynamic content from dynamically changing viewpoints.

Camera Movement for Regulating Focus of Attention   Even if arranged on a flat plane perpendicular to the line of sight (as often in the case of conventional animations), complex animations often include multiple entities that require attention from the viewer. A growing body of literature has shown that due to the transience of animations, learners may tend to overlook some relevant elements or dynamics because they are distracted by other more perceptually salient parts of the animation (Lowe & Schnotz, 2014). In order to guide learners' attention through an animation that requires multiple attentional foci, several cueing options have been developed and empirically tested, including, for example, arrows, shading, or color coding (de Koning, Tabbers, Rikers, & Paas, 2009). Virtual cameras allow for another, yet empirically largely unexplored cueing alternative, namely, change of camera distance from medium long shots (showing the whole scene) to close-ups (showing one particular detail of the scene), either by means of a camera track or by zooming-in. In cinematography, use of camera distance for guiding viewers' attention has a long tradition, and so-called analytical editing of scenes, by which an event is decomposed into various single shots that are shown from different distances, can be considered one of the keystones of Hollywood cinema (Bordwell & Thompson, 1979).

Compared to arrows or color coding, reduction of camera distance could operate more unobtrusively. Also, it not only guides learners' attention to a relevant part of the animation, but also presents this part in an enlarged manner, showing more details and simultaneously keeping other possible distracting elements out of the frame. On the other hand, attention guidance via reduction of camera distance is less precise because arrows or color coding more clearly indicate which of the pictorial

elements are intended to be looked at. Also, in a close-up, only a restricted section of the whole event is displayed, implying that some important contextual information may be missing that would otherwise be represented in memory (Papenmeier, Huff, & Schwan, 2012). Relating these considerations to the Animation Processing Model proposed by Lowe and Boucheix (2011, 2017, this volume), regulating focus of attention by camera movement may again facilitate parsing of the event into discrete steps due to the regular variations of distance from far to close and vice versa (Phase 1), and may also facilitate local processing of these steps because of its closer framing and its pictorial enlargement (Phase 2). On the other hand, due to the loss of "the whole picture" of the event, it may prove detrimental for connecting these steps into a causal chain (Phase 3). But once again, to our knowledge, little empirical research from the field of instructional design exists on the topic of guiding attention by variations in virtual camera distance. First empirical evidence on the cueing functions of zoom-ins comes from a study that was recently conducted by Glaser, Lengyel, Toulouse, and Schwan (in press). Taking three-dimensional reconstructions of ancient Roman buildings as the to-be-learned subject matter, these authors found that compared to static views and zoom-outs, learners in the zoom-in condition looked at the central part of the scene significantly longer, indicating that zoom-ins may indeed serve an attention focusing purpose.

Camera Movement for Decorative Purposes Finally, using camera movements to transform static depiction of scenes to animations with dynamically changing visual information is often used as a strategy to catch and hold viewers' attention in informal learning contexts. For example, museums and exhibitions today make heavy use of screens and displays for expository purposes. However, in the museum context, such displays have to compete with other exhibits for visitors' attention (Schwan, Lewalter, & Grajal, 2014). Building on evidence that dynamic visual stimuli attract more attention than static ones (Mital, Smith, Hill, & Henderson, 2011), many displays in museums come in the form of visualizations which are animated by complex camera movements, for example, as "fly-throughs" of reconstructed excavation sites in archaeological exhibits. Similar arguments also apply to science documentaries on TV or on the Internet. Here again, filmmakers tend to avoid static digital pictures in favor of dynamic ones in order to hold viewers' attention and prevent them from zapping to other competing channels. But besides their attention catching and holding purposes, the camera movements often seem to be only partly motivated by further, more learning-related intentions, similar to the ones discussed above. Therefore, they bear a strong resemblance to the use of decorative pictures and seductive details in multimedia learning material (Magner, Schwonke, Aleven, Popescu, & Renkl, 2014; Rey 2012). Yet, the implications of such decorative uses of camera movements in animations still await further empirical investigation.

2.5   Three-Dimensional Presentation: Adding Stereoscopic Cues

In the field of instructional design, the term "3D" is used in a broad sense to characterize representations that, in contrast to "2D", include a third axis of depth, thereby giving objects volume and defining the spatial layout of a given scene in three dimensions (cf. Jenkinson, 2017, this volume; McGill, 2017, this volume). However, pictorial representations (in contrast to haptic models, for example) are not truly three-dimensional but instead evoke only an impression of three-dimensionality on the basis of projection on a flat surface. To achieve this impression, they make use of a number of different pictorial cues. Perceptual psychology informs us that one large group of static pictorial cues operates monoscopically, requiring just one eye for the impression of depth in space (Vishwanath & Hibbard, 2013). These static depth cues include occlusion, size constancy, converging lines, and texture gradients. Motion parallax, which is the computation of relative distances due to observer movement, constitutes a further rather effective monoscopic depth cue. Besides these monoscopic cues, recent technological advancements have opened up the possibility for the addition of stereoscopic depth cues. In these cases, the term "3D" does not mark the difference to 2D regarding an animations three-dimensional structure (e.g., Huk et al., 2010), but instead the difference between stereoscopic and monoscopic viewing (Carrier, Rab, Rosen, Vasquez, & Cheever, 2012; Khooshabeh & Hegarty, 2010). In order to avoid confusions, we propose to use the term "2.5D" for three-dimensional monoscopic presentations, while restricting the term "3D" for three-dimensional stereoscopic presentations.

While monoscopic presentations do not require advanced technology but can be viewed on ordinary screens (e.g., Berney & Bétrancourt, 2017, this volume; Jenkinson, 2017, this volume; McGill, 2017, this volume), stereoscopic viewing requires special equipment. Several different technologies have been developed for stereoscopic viewing (Mendiburu, 2009). Currently, most applications operate by a combination of a specific display technology together with the use of corresponding glasses. Typically, the screen displays two separate, slightly different pictures to each of the eyes, either simultaneously or in brief succession. Viewers mentally fuse the two pictures into a single percept that appears to be truly three-dimensional, with the strength of the 3D impression depending on inter-ocular distance between cameras and the distance between projection screen and viewer. Differences relate to the way these two pictures are separated, either by combining differently colored pictures and corresponding filtering glasses, alternating pictures and the respective shuttered glasses ("active glasses"), or using polarized light, again together with the respective filtering glasses ("passive glasses"). Also, so-called autostereoscopic displays have been developed that do not require additional glasses, but use prismatic screens projecting two slightly different pictures to the viewer's eyes instead. The various 3D technologies all have their advantages and disadvantages. Stereoscopic pictures viewed with active glasses have a brighter tone but viewing suffers from flickering pictures and the heavy weight of the glasses. In contrast, passive glasses

are more lightweight and do not show flicker but viewing suffers from darker pictures. Finally, for autostereoscopic screens no glasses are needed, but they have a very limited resolution and the 3D impression is strongly dependent on the particular viewing position in front of the screen.

Preparing expository animations for 3D presentation requires careful consideration of several detrimental effects resulting from the perceptual specifics of stereoscopic projection (Meesters, Ijsselsteijn, & Seuntiens, 2004; Mendiburu, 2009). These include cardboard effects (objects appear unnaturally flat), puppet theatre effects (objects appear miniaturized), image ghosting (objects appear to have a second shadow contour), and keystone effects (distortions of vertical parallaxes). But even when designed appropriately, 3D should not be considered more "natural" than other presentation techniques because it is presented on a flat surface and therefore still requires a dissociation of convergence and accommodation Together, these characteristics may contribute to feelings of visual fatigue, visual discomfort, eyestrain, and headaches, which has been reported for a substantial proportion of viewers (Lambooij, Fortuin, Heynderickx, & Ijsselstein, 2009; Ukai & Howarth, 2008).

Hence, from an instructional perspective, the question arises under which circumstances the introduction of 3D instead of 2.5D for purposes of learning and knowledge acquisition is justified, given the necessity of a complex technology (displays, glasses), the additional costs of the appropriate design of stereoscopic material, together with the dangers of visual fatigue or discomfort, and the fact that about 5–10% of the population suffer from stereo blindness (i.e. the inability to perceive stereoscopic projections as three-dimensional; Lambooij et al., 2009). Because stereoscopic presentation has been introduced only quite recently, empirical evidence is sparse and mixed at best.

In general, both 2.5D and 3D provide a third dimension that may be beneficial for building appropriate mental representations, particularly when extension in space is relevant for comprehension. But while going from 2D to 2.5D may add some important information, going from monoscopic 2.5D to stereoscopic 3D is a smaller step because monoscopic presentations already include a rich array of spatial cues. Accordingly, recent findings indicate that learners benefit from the addition of stereopsis only under specific circumstances. More particularly, in basic memory research, several studies have found a stereo advantage for recognition of static objects, especially if these objects are presented from novel views (Bennett & Vuong, 2006; Burke, 2005). This was the case even for displays with strong monocular depth cues (shading; Lee & Saunders, 2011). But on the other hand, for recognition of a large set of photos of natural scenes, Valsecchi and Gegenfurtner (2012) found a stereo advantage only for a small subset of pictures. This positive effect of stereoscopic presentations was even more restricted in cases of animated learning material. In a series of studies, Papenmeier and Schwan (2016) investigated the role of stereoscopy for memorizing complex molecule-like structures. They found that viewers did not benefit from stereoscopic presentation while learning the stimulus material. In contrast, however, if the depictions of molecules were presented stereoscopically in a subsequent memory test, learners outperformed participants who had to solve the memory test with monoscopic test items. This indicates

that stereoscopic information is not included in the memory representation that is built during the learning phase but that the benefit of stereoscopic information is restricted to phases of reactivating object memory for purposes of recognition.

The finding that memory and learning benefit only to a small degree from stereoscopic over monoscopic three-dimensional dynamic presentations is also corroborated by studies with material from various other fields. For example, in a path analysis of possible memory effects of stereoscopic versus monoscopic movie screenings, substantial effects on emotions and immersion but neither direct nor indirect effects on memory for the films' content were found (Carrier et al., 2012). Similarly, using dental anatomy as a learning topic, Khooshabeh and Hegarty (2010) could not find an advantage of stereoscopic animations for tasks of visualizing a cross section of molar teeth. For learning abdominal anatomy, Luursema, Verwey, Kommers, and Annema (2008) found that for novices provision of stereoscopic animations facilitated localization but not identification (naming) of the various anatomical parts. In accordance with these findings, two recent reviews of the effectiveness of stereoscopic displays in medicine come to similar conclusions (McIntire, Havig, & Geiselman, 2014; Van Beurden, Ijsselstein & Juola, 2012). In medical practice, stereopsis has been shown to improve diagnosis (e.g., 3D ultrasound visualizations) and decrease the time needed for minimally invasive surgery (MIS) procedures and, more generally, for tasks involving the manipulation of objects. In contrast, its uses for training and learning are less clear. Analyzing the results of 11 experiments for medical training and learning, McIntire et al. (2014) found that four experiments showed an advantage of stereopsis, four experiments found mixed results, while the remaining three experiments showed no difference between 2.5D and 3D learning material.

Overall, these findings suggest that the suitability of stereopsis for purposes of learning and knowledge acquisition is limited. Not only does a substantial part of the population suffer from stereo blindness and many users of stereoscopic glasses report having experienced eyestrain and headaches, but also the learning gains seem to be small and restricted to certain types of learning content that has a strong spatial component but lacks strong monocular depth cues (McIntire et al., 2014). Accordingly, comparing the suitability of stereopsis in chemistry education, Trindate, Fiolhais, and Almeida (2002) found benefits of stereoscopic presentations only for comprehension of crystalline structures, but not for phase transitions or orbital structures, indicating that possible advantages of 3D presentations are strongly topic dependent.

2.6   Adding Interactivity to Three-Dimensional Visualizations

A conventional animation often allows learners to control its temporal parameters in terms of starting/stopping, varying presentation speed from slow to fast motion, and also changing presentation direction from forward to backward and vice versa (Schwan & Riempp, 2004). While some conventional animations give learners

rudimentary control over its spatial characteristics by letting them switch between two different two-dimensional views (Meyer, Rasch, & Schnotz, 2010), the underlying numerical description of digital animations now substantially broadens possibilities for controlling the spatial parameters of three-dimensional animations by the users. But whereas control of temporal parameters can easily be done with predefined, fixed animations, user dependent variation of spatial parameters requires online computing of the animation and can therefore currently only be done on computer devices with sufficient processing power.

In general, user control provides the opportunity for an animation's characteristics to be adapted to a learners' individual cognitive needs (Schwan & Riempp, 2004). For example, giving learners the option to control the pace of multimedia learning material has been shown to facilitate learning and understanding (pacing principle; Hasler, Kersten, & Sweller, 2007; Mayer & Chandler, 2001; Wouters, Tabbers, & Paas, 2007). In the case of spatial characteristics, options for control encompass all parameters discussed in the previous sections, including continuous camera movements regulating distance through zoom-ins and zoom-outs as well as selection of appropriate, canonical viewpoints. This gives learners the freedom to freely explore a complex object or scene or even a dynamically unfolding event from different perspectives. Typically, learners spontaneously use these options, not only regarding an animation's temporal characteristics (Schwan & Riempp, 2004) but also regarding its spatial characteristics. In particular, changing the angle of view and zooming in/out have been found to be prominent types of interactivity that are heavily used in 3D environments (Yuan, Calic, & Kondoz, 2012).

However, on the other hand, having control over the virtual camera places some additional burden on the learners because they have to appropriately plan and execute changes in camera position. In comparison to predefined system controlled trajectories of the camera, this may lead both to more extraneous cognitive load and also to the danger of choosing suboptimal camera positions (Keehner, Hegarty, Cohen, Khooshabej, & Montello; 2008). Therefore, the benefits of freely exploring a three-dimensional animation in a self-guided manner may be outweighed by its cognitive costs. This may be the reason why most empirical studies that have directly compared system-controlled (non-interactive) and user-controlled (interactive) three-dimensional animations have either found no differences between the two conditions or even advantages of the system-controlled versions (Keehner et al., 2008; Khooshabeh & Hegarty, 2010; Nguyen et al., 2012; Papenmeier & Schwan, 2016).

Whether learners indeed benefit from interactively controlling the spatial parameters of a three-dimensional animation depends on a number of factors. First, successful control of three-dimensional animations seems to require an above average level of spatial abilities (Garg et al., 2002; Huk, 2006). Learners with low spatial abilities may experience high cognitive demands because interactive control requires additional planning and monitoring of content-related activities over and above the cognitive demands that result from building an appropriate spatial mental representation. Second, learners need to have appropriate strategies for controlling the spatial parameters of the visualization. In particular, they should be able to use

an animation's control options to identify and focus on canonical viewpoints that provide the most informative perspectives on a given object or scene (Garg et al., 2002; Keehner et al., 2008). While in the Keehner et al. (2008) study about one half of the learners were able to spontaneously identify these key views, a substantial portion of learners failed to do so, indicating that they lacked the necessary strategies. But it should be kept in mind that in most studies, participants were not familiar with interactive, three-dimensional animations. Instead, it was the first time they had such interactive 3D systems and they had been given only a brief introduction into the system. Therefore, further research should investigate whether training or routinely practicing such tasks for an extended period of time would enable users to develop appropriate strategies for dealing with this type of visualizations. Additionally, almost all of the studies have investigated the role of interactivity for animations of the "complete view of static objects" type (mostly with anatomical topics). An even more demanding type of animations presents dynamic events in which canonical views change during its course. As discussed above, canonical viewpoints may shift during the course of event, requiring a time-dependent planning of the moves of the virtual camera, most probably overwhelming even learners with high levels of spatial abilities. Under these conditions, system-controlled three-dimensional animations would be expected to better facilitate learning than user-controlled types.

Additional measures may also help learners to control the spatial parameters of three-dimensional animations in better ways. In particular, the cognitive costs of executing position changes and movements of the virtual camera may be reduced by the use of devices that allow for a natural interaction with 6 degrees of freedom, like 3D mice or Wii controllers instead of keyboards or 2D mice (Yuan et al., 2012). Reducing the cognitive costs of planning is probably more difficult to achieve. Also, a better spatial orientation of the viewers can be achieved by including a visible coordinate system that updates according to the users interactions with the animation (Stull, Hegarty, & Mayer, 2009). Further, current technology also allows for systems of graded interactivity where learners can choose between different levels of interactivity, depending on their prior knowledge and their cognitive prerequisites. Instead of offering novices the whole range of possible interactions, such systems could, for example, restrict viewpoint positions to a set of meaningful ones and let users switch between them.

2.7   Conclusions and Outlook

Three-dimensional animations can be seen to embody the fundamental transition from sketching to computing that has taken place in recent years. This transition, which is still underway, has profound implications for the development of digital learning material. Being based on numerical descriptions, learning content can be visualized in many different ways – from simple two-dimensional wireframes to detailed stereoscopic renderings. Learning content can also be computationally

transformed and its appearance can be flexibly controlled and modified by the learners. This may even go beyond pure graphic visualizations, opening up possibilities for haptic interactions with 3D prints (Preece, Williams, Lam, & Weller, 2013).

Within this broad range of options, going 3D does not simply add a third dimension to conventional animations, but instead complements them by animations that show static objects or scenes from changing viewpoints. Here, the impression of dynamics is not due to a moving or changing object or scene, but instead due to a moving viewpoint of the observer. Certainly, both principles can be combined, resulting in animations with changing objects or events from changing viewpoints. Also, the notion of interactivity is broader in the context of three-dimensional animations. While traditional animations focus on allowing learners to control the pace of an animation, interactive three-dimensional depictions often allow learners to control their relative viewing position as well; that is, they may interactively approach or retreat, zoom in and out, rotate around an event, or pursue even more complicated trajectories.

From a psychological perspective, these opportunities have implications for learning and understanding. In general, in comparison to two-dimensional representations, animated three-dimensional representations are both more detailed and more complex, with implications for three relevant learning issues. First, three-dimensional animations allow for a precise definition of viewpoint trajectories that may guide the viewers' attention to relevant parts of objects or events; that is, instead of providing learners with a fixed perspective, viewpoints can be flexibly adapted in terms of viewing angle and distance during the course of an animation. Additionally, camera movement may serve a range of different purposes, including completeness of view, optimizing viewpoints, guiding attention, or simply making the presentation more appealing.

Hence, questions of pedagogically, perceptually, and cognitively guided selection of appropriate viewpoints arise (Garsoffky, Schwan, & Huff, 2009). While extension into depth, changing distances, and moving viewpoints are relatively new approaches in the design of instructional animations, they have a long tradition in other fields, particularly in cinematography. Furthermore, filmic design principles have received some attention from empirical research on cognition and perception of film in recent years (Smith, Levin, & Cutting, 2012; Schwan, 2013). Therefore, while the boundaries between animation and film get more and more blurred (McClean, 2007), research findings from cognitive film studies may provide some guidance for animation design as well.

Second, 3D provides a third dimension that may be beneficial for building up appropriate mental representations, particularly when extension in space is relevant for comprehension. However, the term "3D" should be differentiated into monoscopical three-dimensional presentations ("2.5D") and stereoscopic 3D-presentations. But whereas going from 2D to 3D opens up the field for a much broader range of animations because not only events or moving objects but also the continuous changes of viewpoint brought into effect by movements of the virtual camera come into play, introducing stereoscopic 3D does not add much to the instructional options of animation beyond providing an additional depth cue.

Accordingly, recent results show that learners benefit from addition of stereopsis only under specific circumstances, indicating that stereoscopic information supports the construction of mental representations only in the absence of other depth cues such as depth from motion (Papenmeier & Schwan, 2016).

Third, 3D also adds more degrees of freedom for learner control and can be combined with touch, gesture, or head-motion based interfaces instead of mouse or keyboard. In accordance with assumptions of embodied cognition, coupling complex 3D presentations with the possibility for haptic manipulation and haptic feedback has been shown to enhance learning and deepen understanding (Bivall, Ainsworth, & Tibell, 2011). But while more natural, increases in 3D interactivity may also have its costs in terms of increased requirements for appropriately planning and monitoring content-related activities.

Taken together, from a conceptual perspective, existing taxonomies have to be complemented and differentiated with regard to these new forms of animation. Taking the taxonomy proposed by Ploetzner and Lowe (2012) as a starting point, the spatial characteristics of animations should include not only 2D, but also 2.5D (three-dimensional monoscopic) and 3D (three-dimensional stereoscopic) presentations. The taxonomy should also be complemented with a distinction between event dynamics and viewpoint dynamics, offering many new opportunities for future research on the role of animations for learning and knowledge acquisition.

References

Bennett, D. J., & Vuong, Q. C. (2006). A stereo advantage in generalizing over changes in viewpoint on object recognition tasks. *Perception & Psychophysics, 68*, 1082–1093.

Berney, S., & Bétrancourt, M. (2017). Learning three-dimensional anatomical structures with animation: Effects of orientation references and learners' spatial ability. In R. Lowe & R. Ploetzner (Eds.), *Learning from dynamic visualization — Innovations in research and application*. Berlin: Springer (this volume).

Bivall, P., Ainsworth, S., & Tibell, L. A. E. (2011). Do haptic representations help complex molecular learning? *Science Education, 95*, 700–719.

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception, 28*, 575–599.

Bordwell, D., & Thompson, K. (1979). *Film art: An introduction*. New York: McGraw Hill.

Boucheix, J.-M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction, 25*, 71–84.

Burke, D. (2005). Combining disparate views of objects: Viewpoint costs are reduced by stereopsis. *Visual Cognition, 12*, 705–719.

Carrier, L. M., Rab, S. S., Rosen, L. D., Vasquez, L., & Cheever, N. A. (2012). Pathways for learning from 3D technology. *International Journal of Environmental & Science Education, 7*, 53–69.

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*, 10–32.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention guidance in learning from complex animation: Seeing is understanding? *Learning and Instruction, 20*, 111–122.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review, 21*, 113–140.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*, 302–307.

Eriksson, U., Linder, C., Airey, J., & Redfors, A. (2014). Who needs 3D when the universe is flat? *Science Education, 98*, 412–442.

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.

Garg, A. X., Norman, G. R., Eva, K. W., Spero, L., & Sharan, S. (2002). Is there any real virtue of virtual reality? The minor role of multiple orientations in learning anatomy from computers. *Academic Medicine, 77*, S97–S99.

Garg, A., Norman, G. R., Spero, L., & Maheswari, P. (1999). Do virtual computer models hinder anatomy learning? *Academic Medicine, 74*, S87–S89.

Garsoffky, B., Huff, S., & 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.

Glaser, M., Lengyel, D., Toulouse, C., & Schwan, S. (in press). Designing computer-based learning contents: Influence of digital zoom on attention. *Education Technology Research and Development*.

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*, 713–729.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722–738.

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition, 17*, 297–306.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning, 22*, 392–404.

Huk, T., Steinke, M., & Floto, C. (2010). The educational value of visual cues and 3D-representational format in a computer animation under restricted and realistic conditions. *Instructional Science, 38*, 455–469.

Imhof, B., Scheiter, K., Edelmann, J., & Gerjets, P. (2012). How temporal and spatial aspects of presenting visualizations affect learning about locomotion patterns. *Learning and Instruction, 22*, 193–205.

Jahn, G., Papenmeier, F., Meyerhoff, H. S., & Huff, M. (2012). Spatial reference in multiple object tracking. *Experimental Psychology, 59*, 163–173.

Jenkinson, J. (2017). The role of craft-based knowledge in the design of dynamic visualizations. In R. Lowe & R. Ploetzner (Eds.), *Learning from dynamic visualization — Innovations in research and application*. Berlin: Springer (this volume).

Johnston, O., & Thomas, F. (1981). *Disney animation: The illusion of life*. New York: Walt Disney Productions.

Kheener, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science, 32*, 1099–1132.

Khooshabeh, P., & Hegarty, M. (2010). Inferring cross-section: When internal visualizations are more important than properties of external visualizations. *Human-Computer Interaction, 25*, 119–147.

Lambooji, M., Fortuin, M., Heynderickx, I., & Ijsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology, 53*, 1–14.

Lee, Y. L., & Saunders, J. A. (2011). Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *Journal of Vision, 11*, 1–12.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., et al. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 235–247.

Lowe, R., & Boucheix, J. M. (2008). Learning from animated diagrams: How are mental models built? In G. Stapleton, J. Howse, & J. Lee (Eds.), *Diagrammatic representation and inference* (pp. 266–281). Berlin: Springer.

Lowe, R., & Boucheix, J.-M. (2017). A composition approach to design of educational animations. In R. Lowe & R. Ploetzner (Eds.), *Learning from dynamic visualization — Innovations in research and application*. Berlin: Springer (this volume).

Lowe, R., & Schnotz, W. (2014). Animation principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 513–546). Cambridge, MA: Cambridge University Press.

Lowe, R., Schnotz, W., & Rasch, T. (2011). Aligning affordances of graphics with learning task requirements. *Applied Cognitive Psychology, 25*, 452–459.

Luursema, J. M., Verwey, W. B., Kommers, P. A. M., & Annema, J. H. (2008). The role of stereopsis in virtual anatomical learning. *Interacting with Computers, 20*, 455–460.

Magner, U. I. E., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction, 29*, 141–152.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390–397.

McClean, S. T. (2007). *Digital storytelling*. Cambridge, MA: MIT Press.

McGill, G. (2017). Designing instructional science visualizations in the trenches: Where research meets production reality. In R. Lowe & R. Ploetzner (Eds.), *Learning from dynamic visualization — Innovations in research and application*. Berlin: Springer (this volume).

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays, 35*, 18–28.

Meesters, L. M. J., Ijsselsteijn, W. A., & Seuntiens, P. J. H. (2004). A survey of perceptual evaluations and requirments of three-dimensional TV. *IEEE Transactions on Circuits and Systems for Video Technology, 14*, 381–391.

Mendiburu, B. (2009). *3D movie making*. London: Routledge.

Meyer, K., Rasch, T., & Schnotz, W. (2010). Effects of animation's speed of presentation on perceptual processing and learning. *Learning and Instruction, 20*, 136–145.

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.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation, 3*, 5–24.

Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57*, 279–315.

Nguyen, N., Nelson, A. J., & Wilson, T. D. (2012). Computer visualizations: Factors that influence spatial anatomy comprehension. *Anatomical Sciences Education, 5*, 98–108.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Hillsdale: Erlbaum.

Papenmeier, F., Huff, M., & Schwan, S. (2012). Representation of dynamic spatial configurations in visual short-term memory. *Attention, Perception, & Psychophysics, 74*, 397–415.

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.

Ploetzner, R., & Lowe, R. (2012). A systematic characterization of expository animations. *Computers in Human Behavior, 28*, 781–794.

Preece, D., Williams, S. B., Lam, R., & Weller, R. (2013). "Let's get physical": Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education, 6*, 216–224.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*, 216–237.

Schwan, S. (2013). The art of simplifying events. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies* (pp. 214–226). New York: Oxford University Press.

Schwan, S., Lewalter, D., & Grajal, A. (2014). Understanding and engagement in places of science experience: Science museums, science centers, zoos and aquariums. *Educational Psychologist, 49*, 70–85.

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293–305.

Smith, T. J., Levin, D., & Cutting, J. E. (2012). A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science, 21, 107–113*.

Soemer, A., & Schwan, S. (2012). Visual mnemonics for language learning: Static pictures vs. animated morphs. *Journal of Educational Psychology, 104*, 565–579.

St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors, 43*, 79–98.

Stull, A. T., Hegarty, M., & Mayer, R. E. (2009). Getting a handle on learning anatomy with interactive three-dimensional graphics. *Journal of Educational Psychology, 101*, 803–816.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2*, 55–82.

Trindade, J., Fiolhais, C., & Almeida, L. (2002). Science learning in virtual environments: a descriptive study. *British Journal of Educational Technology, 33*, 471–488.

Tversky, B., Morrison, J. B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262.

Ukai, K., & Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays, 29*, 106–116.

Valsecchi, M., & Gegenfurtner, K. R. (2012). On the contribution of binocular disparity to the long-term memory for natural scenes. *PlosOne, 7(11)*, e49947.

Van Beurden, M. H. P. H., IJsselsteijn, W. A., & Juola, J. F. (2012). Effectiveness of stereoscopic displays in medicine: A review. *3D Research, 3*, 1–13.

Vishwanath, D., & Hibbard, P. B. (2013). Seeing in 3-D with just one eye: Stereopsis without binocular vision. *Psychological Science, 24*, 1673–1685.

Wouters, P., Tabbers, H. K., & Paas, F. (2007). Interactivity in video-based models. *Educational Psychology Review, 19*, 327–342.

Yuan, H., Calic, J., & Kondoz, A. (2012). Analysis of user requirements in interactive 3D video systems. *Advances in Human-Computer Interaction, 2012*, 1–11. ID 343197.