Special Section: Research Article

Markus von der Heyde*

Self-Encounter in Virtual Reality in Robot-Based Telepresence

https://doi.org/10.1515/icom-2018-0017

Abstract: Have you ever met yourself? Have you met your past? This report is meant to make a phenomenon known in which VR users at a break-in-presence do not fall back into the lab environment. However, we are not yet able to provide tangible evidence and systematic research about it. Setup: We describe a virtual reality application which originally was intended to provide control for a search and rescue robot. Due to a design requirement to use very limited resources, we developed a sparse representation of the past of the robot. The user encounters the past path of the robot in VR as a collection of 360° photo-spheres which each captures one instant. Multiple users of the application can individually review all past pictures. The most recent picture represents the current perspective of the robot. In addition, each user can interact with virtual objects, e. g., control the robot. Observation: According to perceptual research, breaks-in-presence might occur after sensory conflicts. An encounter of one's self in VR introduces a perceptual and cognitive conflict. Users were able to realign with their own episodic memory and did not fall back into the lab environment as a result of this new type of break-in-presence.

Keywords: break in presence, self perception, virtual reality

1 Introduction

This article is an unconventional way of describing a potentially new phenomenon, which occurred in a particular situation where humans observed a VR representation of their own past. We have chosen to describe the observation by explaining the design principles of the setup in which it occurred and linking it to known frameworks and ideas from perceptual research.

The structure of this report is as follows: The occurrence of a potentially new type of break-in-presence probably relies on the technological choices of the setup in which it occurred. We therefore introduce first the design choices of the technological background (VR and robotics). Next we summarize the main perceptual frameworks related to the issues at hand. Finally, the report describes the phenomenon and some observers' reactions – which is a little like describing colours in black and white print.

***Corresponding author: Markus von der Heyde,** vdH-IT, Weimar, Germany, e-mail: info@vdh-it.de, ORCID: https://orcid.org/0000-0002-6026-082X

2 Technological Background

The two technologies coming together to create the phenomenon described here are virtual reality (VR) and robotics. We understand virtual reality (VR) as computer-generated stimuli presented to users in multiple modalities. Robots use actuators to drive physical interaction. Tele-Robotics is one obvious combination of VR and robotics. The following paragraphs summarize the necessary context for the report of the observed perceptual effects in part two.

2.1 Virtual Reality

Virtual Reality (VR) and Virtual Environments (VE) are often understood as artificial environments where all aspects of perception are generated by a computer. A broad overview of the field is assembled by Jerald [9] in a state-of-the-art reference on technology, applications, and perceptual background.

Two aspects of VR are essential for this paper: First, the perceptual qualities should be generated or transmitted by a computer, but do not necessarily consist of pure mathematical renderings as a duplicate of real stimuli. Second, human observers always perceive their environment simultaneously with all senses; therefore, it consists of sensory qualities from all modalities at the same time. Consequently, even those sensory qualities not included in a simulation are present to the user and will be integrated into one coherent perception of the world.

2.2 Robots and Robotics

Robots[1] classically are driven by actuators to interact with the world in general. Under the control of computers they perform series of (complex) actions automatically. For this article, we use a very broad definition: a robot needs to have one or more actuators (motors/brakes of any kind) to control motion (accelerate or slow down) controlled by a computer. While the application described in this paper used a specific six-leg walking robot, we believe the observed phenomenon to be independent from the type of robot. Controlling a remote camera without show the actual robot is the key in the context of this article.

2.3 Tele-Robotics and VR

The combination of VR as a synthetic environment and an actuator system driven by the same computer has been done, and many applications are possible. If the action is distributed across a considerable distance, we speak of tele-robotics in combination with VE/VR. Many applications have been known since the early '80s [12], when both technologies were introduced in universities and research facilities at high expense.

Today a basic Tele-Robotics system with all necessary components like the one used for our experiments can be built for less than 5000 USD. These basic set-ups can stimulate psychological discourse and offer insights for teaching, learning, and research within universities and start-up companies. The true costs are no longer the investments in technology, but in staff.

3 Concepts of Perception

3.1 Emerging Reality as a Phenomenon of Consciousness

When humans perceive the environment, they rarely are aware of their assumption of coherence and constancy. Sacks [16, p. 178ff] connects various aspects from neuroscience literature to motivate the assembly of single moments into one "apparent continuity". Perceptual continuity is the default [22]. Human observers believe in the existence of this reality if "enough" of what is perceived aligns with their expectations. This might also be linked to the notion of the threshold phenomenon of consciousness in Crick and Koch's framework [5]. The sense of being in a certain context or environment – a phenomenon called "presence" [7] – potentially emerges through our subconscious processes which are not violated by disturbing events.

Breaks-in-presence (BIP) in VR often occur in situations where violations of the expected happen [9, e. g., p. 46]. The term "break-in-presence" was introduced by Slater and Steed [20] in the context of measuring presence as a phenomenon. Like waking up from a dream, the user becomes aware of being "just in a VR simulation" and jumps back to the environment where he or she physically is.

3.2 Perception in Sparse Environments

At the same time, when humans consciously concentrate on specific qualities of the environment, other features and changes go by completely unnoticed. Change blindness has been intensely studied after the first investigation by Simons and Levin [17]. One cause of change blindness might be the ability of the brain to fill in missing but plausible information. The brain – as simulated by machine learning and algorithm predictions – is able to fill in nearly any kind of coherent percept, if the provided information is ambiguous [13]. A non-visual example is that of mis-hearings, which most people have experienced. As Sacks points out [16, p. 126], we do not hear scrambled characters or syllables; rather, "parts of the brain manage to construct real words or phrases, even if they are absurd".

The perceptual representation of an environment can therefore be sparse in the sense of fidelity and quality, as long as no mismatches which are too obvious capture the attention of the observer. But what is "too obvious" in terms of everyday life or a certain application? This certainly depends on the task and intentions of the user. It also depends on the overall perception of all modalities. Therefore, the processes of sensory integration are important for the understanding of perception in real life, especially in sparse and ambiguous conditions.

3.3 Sensory Integration

Let us consider the observer in real life environments as well as in virtual ones to be conscious human being, who tries to make sense of all perceptions at all times. In VR this particularly means all simulated and non-simulated modalities are taken into account. VR research focusing

1 For a general purpose introduction, see https://en.wikipedia.org/wiki/Robotics

only on the computer-generated modalities but trying to understand the human observer has to fail to some degree, at least on natural results of spatial learning.

If we consider all senses of the observer, we need to also consider all layers of information in the percept. Many researchers have proposed models for sensory integration following the maximum-likelihood hypothesis [6]. In other words, humans perceive the expected, most likely joint interpretation of all sensory input. Naturally, we have to acknowledge all cognitive and unconscious biases for the interpretation.

3.4 Augmentation

In the case of the simulated worlds in VE/VR, we present some information generated by the computer and overlay that on top of whatever is already present. The simulated information might be suitable to replace some parts of the environment which would otherwise be perceived.

A common example is the head mounted display (HMD) used in VR which covers part of the visual field and is able to present a rendered scene (or any other image). The remaining visual field is often left black, in hopes of not presenting any disturbing or disorienting information to the user. However, it is well known that peripheral vision provides significant input to at least velocity perception and vection [3]. Additionally, it is known that the presented cues within the HMD do not match the real visual input: the accommodation-vergence conflict is one prominent example which is part of VR sickness (see also [9, p. 160] and [10]).

In cases of augmented reality (AR) headsets, the image reaching the eye is automatically a combination of the "real" environment and an overlay of a virtual additive. This shifts the level of augmented information towards the true environment, but does not constitute a new category of devices. Billinghurst and colleagues [1] compiled an extensive summary describing AR as a seamless blending between reality and virtual additions that have been developed during the last 50 years. Recent technologies have introduced the notion of a "powerful and unobtrusive layer of digital information on top of the real world"[2] to pinpoint the difference between intended augmented reality and other mixed reality applications.

3.5 Augment All Modalities

For visual cues, augmented reality, mixed reality, virtual environments, and virtual reality always display a combination of the actual reference frame of the real environment and anything we augment in it by devices. The same is also true for other senses: Any headphone leaves some of the outside aural world perceptible while presenting a computer-generated sound (or recording) on top of this. Often in psychophysical experiments, the outside world is therefore hidden by white noise or more ecologically valid flowing water sounds, which are less annoying and cover the laboratory noises sufficiently.

Tactile or haptic devices like the phantom do the same: some part of the tactile percept is replaced with a computer generated stimulus (see also [1, p. 144f]). Motion simulations use linear sleds or six degrees of freedom platforms to move the observer. This generates a level of augmentation for the vestibular senses of the inner ear. Nonetheless, the gravity and physical forces of the masses involved act within the external reference frame. The augmentation here is often called "motion cueing", and leaves the users with ambiguous percepts when the simulation deviates from the physical space or volume of the device (see [14] for a perceptually oriented approach minimizing the perceptual error). Even in everyday life we are confronted with virtual augmentation of gustatory or olfactory cues: supermarkets and restaurants use artificial odours to enhance the percept of good taste and longing for food (see [19] for a critical review). Therefore, we can generalize that AR/VR/VE are combinations of augmented percepts in multiple modalities.

Some research and start-ups (e. g., Magic Leap[3]) are trying to produce a perfect level of augmentation. The ultimate goal is to replace all information with a realistic simulation. This strategy is promising a new "Virtual Reality" which ultimately ends at the laws of physics. An alternative strategy cheats on human perception as much as possible and tries to get away with it [8, 2, 14]. Comparing usability and effort, the second approach seems more rewarding: What is the minimal level of augmentation we can present and still elicit the emerging sense of being there? What is needed to interact in a meaningful way?

2 See https://metavision.com/augmented-not-mixed-reality/

3 See https://www.magicleap.com/

4 Example of a Sparse Application

How we can present only the necessary information in an AR/VR/VE context and still maintain a natural behaviour? This general question can only be answered in the context of a specific application. Depending on the task, different levels of information presented in different modalities might be required. Reading, for example, is uncomfortable in current standard HMDs, but reading is fine on most HD displays. But radiologists need to use even higher density and extended colour resolution to recognize conspicuity on their displays making this a requirement beyond reading capability.

What level of sparseness – the art of leaving out information which is normally not used – can we propose for a wider range of applications? Is there an effective way of separating information across time and content? Is there an elegant way of providing interaction without fully 3D modelled environments?

This section introduces a system which is designed with these concepts in mind. We aimed to scale down all technical requirements wherever possible and aimed at a valid behaviorally relevant perception of the user.

4.1 Scenario

We designed a robot/VR system for a search and rescue scenario. We gathered requirements from professionals and went through multiple design stages [18]. The goal was that a robot should be able to enter buildings with debris and other obstacles under direct remote control of a human operator. The environment would potentially change due to further collapsing structures. All current and past representations should therefore be able to be perceived at the operator's choosing. A natural orientation of the operator was expected to be crucial for the efficiency of finding and rescuing injured people.

4.2 Representation

In major catastrophes it is expected that technology, communication bandwidth, and energy will be limited. Online streaming of video or other high density data was excluded due to considerations of maximum distance, storage capacity, and computational power requirements.

The core idea of our system is based on 360° images and their location in 3D space. The design choice was based on the natural ability to act in the required way (e. g., search for survivors) in one single photo-sphere. The obligatory "snap" of the spatial orientation towards the visual representation [15] is known to work for those representations.

The position of the pictures was estimated by the active robot motion and the path integration of acceleration sensors. Additional sensors were designed to detect slippage on loose ground. While path integration from pure vision was an option, this technique was not used.

4.3 Display and Interaction

The visual representation – or, more precisely, augmentation – contained primarily the 360° images. In contrast to other interfaces (e. g., Google Street View), we restrained from using any teleport interaction and aimed for pure walking (see Figure 1). The application selected the two closest photo-spheres from the past path of the robot based on the current user position. Between the two spheres, a smooth blending enhanced the apparent motion effect. Interestingly, we found that the users could deviate considerably from the original path and still accept the percept as a convincing environment.

At the same time, we were able to display additional standard 3D objects inside the photo-sphere as needed: for example, the past path, the current position, or the configuration of the robot. The required information was separately stored and virtually introduced as 3D objects. This enabled full interactivity apart from the static 360° pictures.

Naturally, the actual background consisting of simple photo-spheres did not change in terms of parallax or other visual cues. However, adding foreground objects in VR (e. g., a representation of the true 3D path of the robot) introduced dominant parallax cues, and other shortcomings became less prominent. We conclude that augmentation can be used to suppress unwanted visual cues, as long as relevant dominant cues are present.

While moving along the past path of the robot, the user interacted with the current representation of the robot and chose a location and perspective based on free walking. Steering the robot to new locations extended the VR scene in real time, adding new 360° views every other second.

4.4 Summary of the Sparse Application

Adjusting the level of augmentation for different cues within the visual modality allowed us to generate a sparse

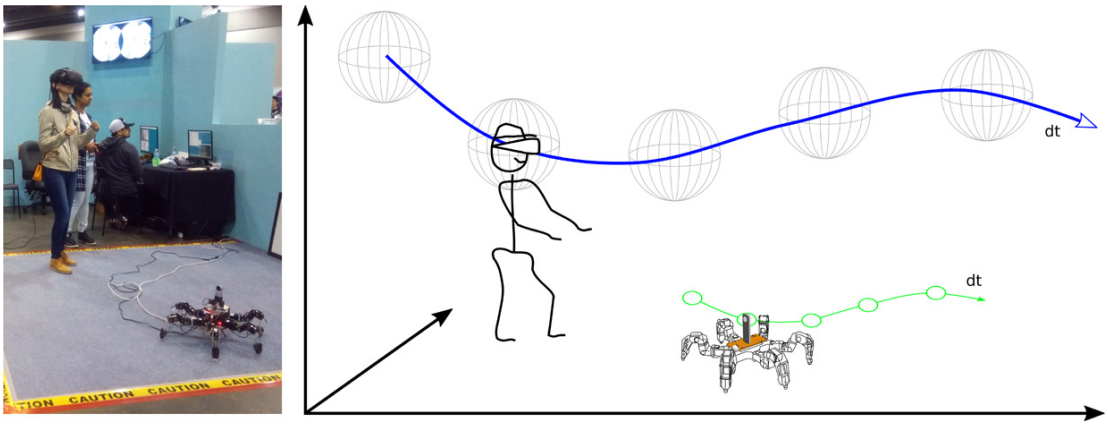

Figure 1: The visual photo-spheres were taken along the path of the robot. The visual display blends the two closest photo-spheres based on the current user position and orientation.

display for the search and rescue robot in combination with real time interaction and natural orientation, even without a true 3D scene in VR.

The recorded past of the robot acts like the pensieve in J. K. Rowling's *Harry Potter* series.[4] We recorded in essence a sparse 360° photo stream of the positions of the robot. The application enabled users to freely move through this chain of bubbles, deviate from the path, and select the time they wished to review by position. In addition, users could interact with the added virtual objects, e. g., controlling the robot's current representation.

5 Self-Perception

Reviewing of the recorded sessions is also possible, as it was required by the original search and rescue application. Observers reported feeling body-less when visiting a scenario of snapshots. The experience – although one can spot oneself in the static images – is far from the situation in front of a mirror. A mirror always reacts continuously and immediately. The camera system, however, allows a variable time delay controlled by the user. Presenting only single images instead of a continuous stream of images makes our representation also sparse in the time dimension.

The series of frozen pictures is less than a video, since only a sparse representation of one picture per second is provided. On the other hand it felt interactive, since one could not only turn around within one static photosphere, but move by walking back and forth along the original path of the robot. Therefore one had the impression of "controlling time" by walking. The VR application enabled seamless blending between pictures which matches the concept of Crick and Koch [5] ("a visual series of snapshots with motion painted on them") as described by Sacks [16] in the overall context of blending single pictures in a stream of conscious impressions.

However, an unexpected effect occurred for those observers seeing themselves within the recorded VR. Here again a conflict forces the human to find a plausible interpretation. While one sees the picture of one's own body, the episodic memory[5] of this instance comes back and becomes very vivid. A key element in our observations which links this experience to episodic memory is that one can easily describe what happened before or after this certain situation. One observer commented, "It feels like an interactive movie". The primary reference frame was neither the laboratory nor the simulated VR, but the actual episodic memory of the individual observer.

Since the observer is already immersed in the scene, all visual cues (and potentially auditory cues as well) match the episodic memory from this moment in time. The combination of one's own memory and the scene revisited in VR is much more intense than a standard video. Although interactions with virtual robots do not change the past perspective of the robot, the virtual interaction seems to enhance the experience.

In the moment of self-encounter, the observer seems to have an active choice between the recorded pictures from

4 Magical object first introduced in "Harry Potter and the Goblet of Fire".

5 The features of episodic memory are summarized in Table 1.1 by Tulving [21, p. 11]. Later, the term also was included by LeDoux and Brown [11] in the higher-order theory of emotional consciousness.

the robot's perspective and the individual episodic memory as the current reference frame. More research is needed to identify if it would be possible to act in either of them. Hypothetically, if the reference frame of the episodic memory could be chosen deliberately, observers should be able to change the subjective perspective and point towards other reference points in the room context.

While something like a break-in-presence happens, the observer is left with the choice between two plausible reference frames: either the one from the physical lab or the one from one's own memory. The latter is only possible due to the strong immersion in the matching scene of the episodic memory. Both are an alternative to the past view of the robot the user previously had immersed themselves in.

Anecdotally, the observed effect was stronger in situations where the person had done something noticeable in the recorded pictures, e. g., putting on the headset at the start of the simulation or kneeling down to the robot. Pictures with a neutral pose did not draw as much attention. Overall, this indicates that the episodic memory of the situation can only be used as a strong reference if there was something worth remembering.

The situation also was very distinct from an effect of embodiment, as the observer was not in the position of viewing the body from within a virtual body. Surprisingly, the conscious conflict of seeing the body from outside triggered the new type of break-in-presence.

To what extent the sparseness of the simulation contributed to the "option" to choose between body-less observation and reliving the episodic memory is not yet known. The "framework for classifying representations of humans in physical and digital space" by Bailenson and colleagues [4] could be applied to design research experiments defining necessary and sufficient conditions for the phenomenon to occur. Further research is needed to address this and many other questions around the phenomenon of encountering one's own episodic memory in VR.

6 Summary

This paper described a robotic application in which a real time VR scenario is constructed by a collection of positioned 360° pictures. The representation is selected to be sparse, in the sense that only necessary information is stored and displayed to users interactively. The users (possibly multiple ones at the same time) can review the recorded scene or follow the path of the robot online in a search and rescue mission.

Visiting the recording of the past reminds one of the magic pensieve in the *Harry Potter*-series. The self-encounter during this visit elicits a flashback of one's own episodic memory. It could be seen as a special case of a break-in-presence, but the observers did not jump back into the physical world outside VR. On the contrary, it enabled an active choice towards the situation of the episodic memory.

This first report of this new phenomenon has not been confirmed by basic research. However, it should stimulate further considerations when planning VR applications with and without avatars matching the actual body of the users.

Acknowledgment: The author thanks Bernhard E. Riecke and his team at the iSpace Lab of Simon Fraser University, David Clement of Wavesine Solutions, and Saeed Dyanatkar and his team at the Emerging Media Lab of the University of British Columbia for fruitful discussions on VR and robotics. Without the robotics team at Archiact Interactive and the generous support by its founder Derek (Jinlin) Chen, this work would not have been possible.

References

[1] Billinghurst, M., Clark, A., & Lee, G. (2015). A Survey of Augmented Reality. Foundations and Trends® in Human–Computer Interaction, 8(2–3), 73–272. https://doi.org/10.1561/1100000049

[2] Bruder, G., Interrante, V., Phillips, L., & Steinicke, F. (2012). Redirecting Walking and Driving for Natural Navigation in Immersive Virtual Environments. IEEE Transactions on Visualization and Computer Graphics, 18(4), 538–545. https://doi.org/10.1109/TVCG.2012.55

[3] Berthoz, A., Pavard, B., & Young, L. R. (1975). Perception of linear horizontal self-motion induced by peripheral vision (linearvection) – Basic characteristics and visual-vestibular interactions. Experimental Brain Research, 23(5), 471–489. https://doi.org/10.1007/BF00234916

[4] Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. Presence: Teleoperators and Virtual Environments, 15(4), 359–372.

[5] Crick, F., & Koch, C. (2003). A framework for consciousness. Nature Neuroscience, 6, 119–126. https://doi.org/10.1038/nn0203-119

[6] Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion.

Nature, 415(6870), 429–433. https://doi.org/10.1038/415429a
[7] Heeter, C. (1992). Being There: The Subjective Experience of Presence. Presence: Teleoperators and Virtual Environments, 1(2), 262–271. https://doi.org/10.1162/pres.1992.1.2.262
[8] von der Heyde, M., & Riecke, B. E. (2001). How to cheat in motion simulation–comparing the engineering and fun ride approach to motion cueing (Technical Report No. 089). Tübingen: Max-Planck-Institut für biologische Kybernetik. Retrieved from http://www.cyberneum.de/fileadmin/user_upload/files/publications/pdf635.pdf
[9] Jerald, J. (2016). The VR Book: Human-Centered Design for Virtual Reality. New York, NY, USA: Association for Computing Machinery and Morgan & Claypool. ISBN 978-1-970001-12-9.
[10] Keshavarz, B., Riecke, B. E., Hettinger, L. J., & Campos, J. L. (2015). Vection and visually induced motion sickness: How are they related? Frontiers in Psychology, 6(413). https://doi.org/10.3389/fpsyg.2015.00472
[11] LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. Proceedings of the National Academy of Sciences, 114(10), E2016–E2025. https://doi.org/10.1073/pnas.1619316114
[12] Minsky, M. (1980). Telepresence. OMNI Magazine. Retrieved from http://web.media.mit.edu/~minsky/papers/Telepresence.html
[13] Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. Nature, 317, 26.
[14] Pretto, P., Venrooij, J., Nesti, A., & Bülthoff, H. H. (2015). Perception-Based motion cueing: A cybernetics approach to motion simulation. In Recent Progress in Brain and Cognitive Engineering (pp. 131–152). Springer. ISBN 978-94-017-7238-9.
[15] Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2005). Visual cues can be sufficient for triggering automatic, reflexlike spatial updating. ACM Transactions on Applied Perception (TAP), 2(3), 183–215. http://doi.acm.org/10.1145/1077399.1077401
[16] Sacks, O. (2017). The River of Consciousness (first edition). New York, Toronto: Alfred A. Knopf. ISBN 978-0-385-35256-7.
[17] Simons, D. J., & Levin, D. T. (1997). Change blindness. Trends in Cognitive Sciences, 1(7), 261–267. https://doi.org/10.1016/S1364-6613(97)01080-2
[18] Stepanova, E. R., von der Heyde, M., Kitson, A., Schiphorst, T., & Riecke, B. E. (2017). Gathering and Applying Guidelines for Mobile Robot Design for Urban Search and Rescue Application. In Human-Computer Interaction. Interaction Contexts (Vol. 10272, pp. 562–581). Vancouver, Canada: Springer, Cham. https://doi.org/10.1007/978-3-319-58077-7_45
[19] Spence, C. (2015). Leading the consumer by the nose: on the commercialization of olfactory design for the food and beverage sector. Flavour, 4(1), 31. https://doi.org/10.1186/s13411-015-0041-1
[20] Slater, M., & Steed, A. (2000). A Virtual Presence Counter. Presence: Teleoperators & Virtual Environments, 9(5), 413–434.
[21] Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In The missing link in cognition: Origins of self-reflective consciousness (pp. 3–56).
[22] Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. Nature, 414(6861), 302–305. https://doi.org/10.1038/35104551

Bionotes

Markus von der Heyde
vdH-IT, Weimar, Germany
info@vdh-it.de

Dr. Markus von der Heyde received his PhD in Computer Sciences from the University of Bielefeld for his work at the Max Planck Institute for Biological Cybernetics Tübingen in 2000. His approach to adopt biological principles into distributed computer applications in order to enhance stability and robustness was applied in DFG and EU funded research projects. Between 2003 and 2011 he served as ICT director of Bauhaus University in Weimar and focussed on topics such as information security, service management, strategy and governance. Since 2011 he is CEO of vdH-IT and management consultant specializing in IT governance and digital transformation in higher education. In cooperation with various partners he has conducted the German CIO studies since 2014. Since 2016 he has organized the HEI CIOs congress in Germany. He supports ZKI, GI, EUNIS and EDUCAUSE, and serves as a program committee member as well as a proposal reviewer for conferences and the scientific community. From 2016 to 2018 he also was Head of Research and Development in an AR/VR start-up in Vancouver and combined biologically motivated robotics and perceptually oriented VR design. In 2018 he was appointed Adjunct Professor in the School of Interactive Arts and Technology (SIAT) at Simon Fraser University (SFU), Vancouver. Recently he conducted the Open Data Repository Landscape Analysis for the Swiss National Science Foundation (SNSF), adding science policies and strategic planning on a national level to his portfolio. See more publications and details on https://www.researchgate.net/profile/Markus_Von_Der_Heyde3.