Videography as Design Nexus is a doctoral thesis in Innovation and Design on how to improve the usabilities of recorded video instructions via design-erly means. This thesis highlights the videographer’s craft, how he/she handles technological constraints, records procedural instructions and “edits in the head” while recording. In so doing the videographer must consider perceptual affordances and live-action video medium specificities with the video-user as both a biological and cultural being in mind.

In this study, several live-action videos, including story-driven films and typical procedural video instructions serve as Research Through Design (RTD) prototypes that permit critical inquires with regards to their assumed communication efficacies. These efficacies are explored and assessed in tests that feature actual users and audiences in more or less ecologically valid trial set-ups, as well as in blind tests that compare different variants of video styles. A number of these tests also include eye-tracking as a method in order to identify how live-action video instructions’ usability may be conditioned by users’ basic visual decoding strategies. Moreover, YouTube data is gathered and analyzed as a complementary way of circumscribing a video’s usability aspects in the online, screen-based, learning context.

The conclusion of this research undertaking is that medium specificities of live-action videos influence how such videos are appreciated and experienced by users in the moment of interacting with them. For example, this study shows how degrees of perceived verisimilitudes, in part, depend on camera-handling techniques and choice of recording gear. Other findings imply that live-action video instructions are not so sensitive to unfocused viewing behavior, and that the display of discontinuity in recorded videos results in users experiencing increased cognitive load.

Thus, this study addresses both first- and second-order understandings of instructional video communication. It does so by demonstrating how both sublime image quality factors and the video instructions’ overall communication congruence impact on the ability of the video medium to support procedural knowledge and purposeful information processing of users. We also gain some concrete recommendations with regards to how to manage actual live-action video production constraints as videographer or designer of recorded video instructions, for the betterment of the video-user.

Per Erik Eriksson is a Ph.D. candidate in Innovation and Design at the School of Innovation, Design and Engineering, Mälardalen University, Sweden. He holds a Ph. Lic. in Innovation and Design at Mälardalen University, an MA in Communication at Stanford University, USA and a BFA at Massachusetts College of Art and Design, USA. Eriksson is a former TVproducer, editor and documentary filmmaker. Since 2006, he holds a position as Lecturer in Moving Image Production at Dalarna University, Sweden. He is currently Director of the Short Story Audiovisual Design Program at Dalarna University. Eriksson is also part of the ISTUD research profile, as well as the Audiovisual Research group at Dalarna University, with a research focus on audiovisual knowledge in design processes.
VIDEOGRAPHY AS DESIGN NEXUS

CRITICAL INQUIRES INTO THE AFFORDANCES AND EFFICACIES OF LIVE-ACTION VIDEO INSTRUCTIONS

Per Erik Eriksson
2018

School of Innovation, Design and Engineering
VIDEOGRAPHY AS DESIGN NEXUS
CRITICAL INQUIRES INTO THE AFFORDANCES AND EFFICACIES OF LIVE-ACTION VIDEO INSTRUCTIONS

Per Erik Eriksson

Akademisk avhandling

som för avläggande av filosofie doktorsexamen i innovation och design vid Akademin för innovation, design och teknik kommer att offentligen försvaras fredagen den 14 september 2018, 10.00 i Raspen, Mälardalens högskola, Eskilstuna.

Fakultetsopponent: Professor Fredrik Lindstrand, Konstfack

Akademin för innovation, design och teknik
Abstract

This thesis is about live-action instructional videos (LAVs). By addressing design problems with respect to the how-to video genre, the thesis asks fundamental questions about mediated instructional communication efficacies and the factors that either obstruct or augment them.

The analysis presented in this thesis is based on the notion that videography is a design nexus and key focal point of the connections that make live-action video instructional efforts possible. This Design Nexus is explored by defining and illuminating key ontological dimensions, medium specificities and the video users’ cognitive capacities. This is to acknowledge that the users of instructions in this thesis are center stage, both as biological and cultural beings.

The methods used in this thesis and its associated papers are eye-tracking, video observations, questionnaires, self-reports, focus group interviews and YouTube analytics. Hence, both numerical data and non-numerical data are analyzed in this study.

The results of the analyses indicate that pre-production planning is key in live-action video instructional endeavors, but not at the expense of the videographer’s status as designer. Moreover, the analyses show that users’ cognitive processing and visual decoding depend on the power of the live-action format to show actual human behavior and action. Other presented evidence seems to infer that LAV-instructions are a little less demanding if users apply a focused decoding style when interacting with them. Nevertheless, physiological engagement of this kind is likely not to fully compensate for users’ psychological engagement.

This thesis contributes to a more comprehensive understanding of humans’ abilities to interpret the actions of others via medial means. By relating this to video medium-specific affordances, this thesis also furthers important efficacy distinctions and boundary conditions. This understanding is considered important for live-action video makers and designers of visual instructions as well as scholars who need to develop better methods to assess users’ behavioral engagement when they interact with digital instructional media.

ISSN 1651-4238
Acknowledgements

First, I’d like to thank Professor Yvonne Eriksson, my advisor, for her spirited, but kind, and formidably well-informed guidance. It was under her supervision that I made the most progress toward my academic goals. I also thank my second advisor, Professor Arni Sverrisson, who, among other good deeds, made sure I was able to become a doctoral student in the first place. I also thank my colleague and fellow researcher, PhD Thorbjörn Swenberg, for his extremely sensible take on various scientific matters. Without him, this research journey would have been much more bumpy.

Thanks to PhD Marie Sjölinder (of SICS) for well-motivated and specific feedback when I was carrying out the finishing touches.

At DU, I thank Professor Lars Rönnegård, PhD Xiaoyun Zhao, PhD Åsa Pettersson, PhD Johnny Wingstedt, Andrew Scott, Sören Johansson, PhD Cecilia Strandroth, Ari Pereira, Anita Carlsson, Joakim Hermansson, Marit Nybelius Stub and PhD Thomas Florén for their ongoing intellectual support. I also thank Ninni Jonsson Wallin, Sylvia Ingemarsson, Joachim Bergenstråhle, Diana Walve, PhD Sven Hansell, Martin Sandström, Totti Allard, Therese Hercules, Henrik Stub and Håkan Holmberg for being considerate colleagues. Thanks also to Head of School of Humanities and Media Maria Görts, Docent Tomas Axelsson, Professor Emerita Kerstin Öhrn and Acting Pro-Vice-Chancellor Jonas Stier who are among some of the most important people in making Dalarna University a great place for media production research.

At MDH, I have had the good fortune of getting to know PhD Ulrika Florin, PhD Jennie Schaeffer, PhD Carina Söderlund, PhD Koteshwar Chirumalla, PhD Anna-Lena Carlsson and Jonathan Lundin. Thanks for setting the tone of my doctoral experience.

Special thanks to my parents, Roland and Agneta. It is my good fortune to have been born their son.

However, my deepest gratitude goes to my wife, Charlotte. Never once did she question whether this journey with all its sacrifices would be worthwhile. She knew it was important to me. Equally important was the cheering on from my children, Yasmin, Vidar, Love and Edith. You are most loveable.
Dedication
For Sten Ekwall.
Thank you for opening my eyes to the real meaning of ecology.
List of Papers

This thesis is based on the following papers/articles, which are referred to in the text by their ordering capital letters. All of the papers/articles are peer-reviewed.


Reprints were made with permission from the respective publishers.
Contents

1. Introduction ....................................................................................... p. 1
   The Rise of the How-To Video Genre 1
   The Live-Action Format 3
   The Video-User 4
   Problem Statement 5
   Aims and Research Questions 6
   Contributions 7
   Delimitations 8
   Research Basis and Conceptual Linkages 8
   Outline of Thesis 10

2. Theory and Previous Research ...................................................... p. 11
   The Realist Theoretical Approach 11
   Videography as Design Nexus 12
   Design Principles and Cognitive Load Theory 14
   Live-action Videography and Modality 19
   Direct Perception and Affordances 21
   Human-Centered Design 25
   Medium Theory and Human-Computer Interaction 28

3. Methods .................................................................................. p. 32
   Overall Research Approach 32
   Sampling and Sample Sizes 34
   The Qualitative Research Approach 34
   Choice of Methods and Methodological Reflections 35
   Questionnaires and Self-reports 36
   Video Observation and Learner Tests 38
   Eye-tracking 40
   YouTube Analytics 41
   The Focus Group Interview 43

4. Results and Analyses ................................................................... p. 45
   Videography-Generated Medium Specificities 45
   Leveraging LAV-instructions’ Efficacies 47
   Obstructing LAV-instructions’ Efficacies 49
   Implications of Findings 50

5. Discussion and Conclusion ............................................................ p. 53
   The Results - An Overview 53
   Instructional Live-Action Video Design Prescriptions 58
   Concluding Remarks 61
   Limitations and Future Research 62
   References 63

Papers A-G and Additional Materials (Appendix) ......................... p. 75
   Paper A
   Paper B
   Paper C
Paper D
Paper E
Paper F
Paper G
Questionnaires and Additional Information
1. Introduction

This thesis reflects an ongoing intellectual mission grounded in practice as well as in scientific research. This mission is to formulate an answer to how one might design instructional videos in order to improve their usefulness.

To me, this question has been a haunting one since the early 2000’s. I was then involved in an instructional video project associated with a company called Palm Pilot, situated in Silicon Valley, California. At the time, Palm Pilot needed their sales staff to quickly get a firm grasp of what wonderful things this device (a forerunner to smartphones) could do for its users. Someone at the company had decided that the solution to this information need was to produce a short video to showcase all of the device’s great new features and instruct on how to use them. However, the film’s producer had no idea of how to accomplish this. There was no design plan worth the name, and I - the videographer - had certainly no clue since I considered myself a documentary filmmaker who did not want to be associated with the not-so-glamorous instructional film genre. However, the producer and I ended up spending a few hours shooting the video at the company’s headquarters with an articulate man showing the device and talking about it to the camera. The following day, I spent countless grueling hours editing the video, desperately trying to figure out how to edit the material to convey what was meant to be conveyed. The extremely indecisive and grumpy producer was hanging over my shoulder, saying, over and over again, “do this instead!”, something which had become the mantra of our hapless collaboration and was unfortunately of no help. Thus, the video was stuck in a communication limbo.

Shortly thereafter, I was assigned to edit one instructional video sequence of the Hampton Hotels DVD-series. Hampton Hilton were then rapidly expanding their hotel franchise network and needed to ensure that all of the staff of the various franchises could serve a good, and, above all, consistent, Hampton Hotel continental breakfast. The solution was a high production value DVD that basically communicated a lot of “do this, but do not do that”. However, it did so in a humorous, cheerful and campy manner. Possibly, this instructional strategy backfired, as, a few years later, I learned that the DVD had been painful to watch – at least, according my next-door neighbor. She worked at a local Hampton hotel and was very critical of the pedagogical failures of the video (and we never spoke again). Thus, several negative experiences, such as these, resulted in my wondering from time to time, to paraphrase W.J.T. Mitchell (2005), why is it that some instructional visuals do not always achieve what they “want”?

However, in my personal life, I have frequently consulted how-to, or do-it-yourself, videos on YouTube. In some circumstances, I prefer instructional videos showing “real life” over static pictures or abstract diagram-based procedural instructions. Therefore, to cut a long story short, intrigued by the question of how information design can best make complex information unambiguous, how-to videos have become the object of study in my research. And now in the age of YouTube, as it turns out, I am certainly not alone in being interested in the communication potential of video instructions. There are
more how-to videos being made than ever before, and there is little doubt that video instructions have become a popular learning and support tool in the education and training arenas (Clark & Mayer, 2016).

In addition to being a researcher, recently I have continued to explore this genre as a videographer, editor, producer and director. For example, I have recorded, produced and co-produced several instructional videos for organizations and companies, such as Granngården, XL-bygg, Swedol, Socialstyrelsen and Svenska ishockeyförbundet.

Over the years, I have come to realize that, in spite of the recorded video format’s seemingly commonsensical usability aspects, this instructional format is a complex medial object that is partly a decision support system, partly an audiovisual information system. While the instructional format might be very useful, it may also be completely useless, unless its expressional qualities are fully understood and adequately controlled by the videographer. Hence, the videographer is an “information manager” (Bonsiepe, 1999, p. 59).

1.1 The Rise of the How-To Video Genre

In this image-crazed world, we are constantly bombarded with audiovisual arguments for alternative truths. These are arguments that make viewers wary of rational thought, that entertain them, politicize them or enable them to enjoy delusion. Alongside such intricately crafted fictions, there are audiovisual objects that work in completely different ways. These have the capacity to extend our bodily and cognitive functions and demystify the complexities of everyday reality. They are highly unfussy and utilitarian. They are about truths. These thoughts are based on what I have called “how-to video.”

According to the 2017-survey of Internetstiftelsen i Sverige (ISS), YouTube is growing in popularity. This is especially true for relatively young audiences. 86 percent of young Swedish people aged between 12-15 years use YouTube every day. They are among the over one-and-half billion YouTube-users (2018) worldwide, expected to become almost two billion by 2021 (Youtube.com/press 2018). The keywords “how to” generate more than half a billion hits (consulted in January 2018). Citizens, students, educators and businesses increasingly opt for the video medium as an instructional, visual means of presenting and interacting with complex information. They do so, it seems, because of the commonsensical notion that the video medium affords the possibility for the video user to quickly grasp key time and velocity information aspects, since it allows for a more direct audiovisually-primed engagement and immersion than static media. Virtual Reality (VR) and Augmented Reality (AR) mediated media surely do so too, and are in this respect assumed to represent an apex. But in comparison, as of yet, video communication means are a lot more popular, accessible, affordable and easy to use in training and education. However, this popular appeal is not entirely new. Instructional films have been produced and used at the very least since the early 1940s. For example, during World War 2, the father of the ecological theory of direct perception, James J. Gibson, was involved in making instructional films for US-Army fighter pilots.

At the heart of many contemporary video-based instructional communication efforts is the notion that (printed) words are not the perfect communication means for instruction, partly due to language not being universal (Eriksson, Johansson & Björndal, 2011). The rise of the instructional video genre is also due to certain consumption-related and technological factors. In terms of the demand side, there is the consumer-driven growth of video instructions that complements or replaces older instructions which up until now were often associated with a certain expense. Such videos often employ experts who communicate a specific skill or method. A variant of this are video instructions that replace or complement older paper-based instructions on how to use and operate a particular (consumer) product. Such videos may or may not include humans; they tend to be transmedial and are often high in production value. Another large source of video instructions is found within the educational sector,
such as videos specifically designed for e-learning. Software training videos are a common type (Swarts, 2012; Swarts & Morain, 2012; van der Meij & van der Meij, 2013). Yet another common type are professionally made industrial videos that may be found on YouTube, where they tend to be part of various marketing efforts and are high in production value. Otherwise these videos may exist on websites, or on company-specific local networks, and tend to be of low production value. In brief, generally speaking, there are three common types of how-to videos:

. Those that are part of formal educational/training settings
. Those that are part of formal corporate settings
. Those that are informal, DYI (do-it-yourself) - videos, produced by amateurs.

On the supply side, the extremely rapid advances in easy-to-use and versatile media technologies have played an important role, not least, HD-video capable smartphones with high-resolution displays or powerful DSLR-cameras. Inexpensive, sometimes even free, editing and post-production software, in conjunction with powerful computers, has also greatly facilitated the production and editing of instructional videos. The ability to upload instructional videos easily and rapidly has provided the final momentum, since the Internet provides the premier distribution channel for reaching millions of potential clients at the touch of a screen.

In conclusion, the communication and distribution advantages outlined above have resulted in a rapidly increasing catalogue of video instructions: training videos, standard operating procedural videos, cooking videos, lessons learned videos, language instruction videos, assembly videos, tutorial videos, sport and recreational how-to videos, software tutorial videos, video lectures, reference manual videos, and so on.

1.2 The Live-Action Format

Although it is difficult to know for certain, it is very likely that the majority of “how-to” videos are of the Live-Action format (LAV-format). This particular video format is the object of study in this thesis. And henceforth, in this thesis, I will refer to this video artifact as “LAV”. This is a visual “echo object” that performs unique “cognitive work” on its users (Stafford, 2007). It is also a video format that is intrinsically linked to the activity of videography, just as it is to most documentaries and fiction films (Cutting, 2005). This format represents a distinct visual ontology with unique material properties, addressed in this thesis in terms of medium specificities. Compared to computer-generated images, live-action ontologies are photographic in nature and synchronous-sounding. However, similar to computer-generated images, these ontologies are digital. In digital LAV-ontologies the notion of verisimilitude and the reproduction of realistic movement, audio, color, geometrically valid, space configurations, etc., are key. Therefore, to the average person, at present, Live-Action videography is the most typical example of an audiovisual reproduction of the stuff that makes up reality. This partly explains the fascination with fake LAV-footage (also known as AI-enhanced videos). Hence, the notions of indexicality and visual evidence cannot easily be disregarded when it comes to the LAV-format (Eitzen, 2005).

In spite of live-action videography being a well-known design activity, in academia the LAV-format is frequently confused with and/or grouped together with other kinds of video formats that are fundamentally different. Even worse, it is sometimes completely disregarded as a unique kind of video format. Perhaps this is due to the LAV-format blending most seamlessly with synthesized imagery. Perhaps elitist pretentions that hold refined abstractions at an apex in instructional content, such as is often considered in the field of information design, render it unworthy of serious scholarly inquires (cf. Isaacs, 2008). As a consequence, at least in academia, it appears that this instructional format is the least understood. Indeed, this is strange since this format in its celluloid form marks a
groundbreaking communication invention. In any case, this invention, in particular in the form of fiction and documentary films, has certainly permeated and conditioned technology-mediated communication ever since it came into commercial use in the late 1800’s.

A LAV is a video that is the result of a videographer pressing the record button on an image recording device, thereby recording action live as it unfolds. Of course, there are other LAV-formats that also show “reality” as it unfolds (sports, events, reality-based entertainment, news, live-streamed happenings etc.). However, this thesis focuses on LAVs that are recorded and edited, not merely transmitted live as in live broadcasts/live streaming. This highlights my view that the term “live-action”, derived from film production discourse, provides a good, concise, commonsensical starting point for a more systematic study of how-to videos’ efficacies. It does so since, by definition, it excludes other formats that represent fundamentally different ontologies that are not live, such as computer-generated animations, what Anderson & Hodgins call “synthesized” images (2005, pp. 61-66).

This is also to underscore that the LAV-format, regardless of genre, has similar affordances, since all LAV-formats (documentaries, fiction films, how-to videos, etc.) are shot on cameras. The recording technologies used in LAV-instructional video making are essentially tools of the TV/Film-trade. These tools have long histories not easily disregarded, especially in terms of their stylistic imprints (Eriksson, 2013; Salt, 2009). Hence, in many cases, cinematic tools’ transformative powers are carried over to, and take on, material form in instructional LAVs. This phenomenon is thoroughly explored in my Licentiate thesis, Videoography as Production Nexus (Eriksson, 2013).

1.3 The Video-User

LAV-instructions render complex information unambiguous with user needs in mind. If good video-mediated instructional practices are to be adequately defined, the video user must be consulted in one way or another. This is the one primary, albeit indirect, insight of my Licentiate thesis (2013) that explores how creativity is conditioned by the videographer and his/her associated technical systems.

In this thesis, the term user primarily connotes a conceptual construct, a kind of symbolic representation, although it features real users too (students and professionals). More concretely, the term represents a relevant and comprehensive label for all the different persona data categories (i.e. probability samples), as well as the more or less randomly collected samples that are the sources for the empirical evidence discussed in this thesis.

Here, user is also at times interchangeable with the term audience or viewer. However, in this thesis I mostly also employ the term user in a rhetorical way to emphasize what I consider critical usability aspects of live-action videos, as opposed to their aesthetic and immaterial dimensions. Consequently, I regard this user as a human who needs the instruction to be useful, and who is driven by inner, but also external, motivational aspects when interacting with the instruction. In this sense, the user is also a learner. It is essential that positive learning outcomes are at stake for the user. The user’s motivation primarily springs from the need to most effortlessly and conveniently accomplish a task in a practical and unique context. Therefore, it can be expected that information search behavior among users probably differs since information seeking and acting upon instructions is conditioned by sociocultural factors. Thus, the relevance of the sought-for information is relative (Söderlund & Lundin, 2017). Nevertheless, application of the term user highlights that all users share similar concerns, primarily that an instruction ought to function in a way that renders it useful. Alternatively, no user would think that an instruction they consider to be bad is useful, although, possibly, they would find it amusing.
The term user employed throughout this thesis also connotes other common critical denominators, for instance, how users of visual instructions have very similar cognitive architectures and working memory capacities. This is fortunate, since this commonality essentially renders standardized technology-mediated instructional communication possible and instructional best-practices relevant (Mayer, 2005; Sweller, 1988). Otherwise, designers of instructions would have to resort to designing millions of tailored alternatives which would be practically impossible to achieve. This is especially true considering peoples’ different ability levels, diverse decoding styles and different cognitive styles (cf. Höfler et al., 2017). Thus, a user-informed design research perspective facilitates the identification of common denominators, baseline requirements and sound best practices.

Here, the term user also suggests that I do not approach subjects as primarily unique viewers/listeners, although the subjects that have generated the data available for analysis in this thesis certainly are also unique viewers/listeners. Instead, I am more interested in how user agency plays out in more or less ecological learning scenarios. I do this well aware that such agency may or may not be present in actual practice, which is why I refrain from claiming that the videos presented in this thesis are consistent with user-centered design or user-experience-design (UX-design), since such design approaches tend to take user agency for granted.

By stating the above, I also acknowledge a growing awareness of a problematic divergence between the “design project” and the “user project” (Erthoff & Marshall, 2008, p. 427). Indeed, it is puzzling that although educational psychology research into the technology-mediated instructional arena has been extensive, especially considering so-called multimedia, there are still many unusable instructions in fleeting existence. On the other hand, there are assuming inferior instructions of low production value that work surprisingly well for their users.

1.4 Problem Statement

Scientific and practitioner literature abounds with multimedia design principles. Such principles are for the most part considered the basic means to augment learning efficiency in knowledge-transfer situations that employ videos. However, in this strain of literature, there is little attempt to theorize in a comprehensive way, how, why and when these principles apply to LAV-instructions. Primarily, the problem is that previous instructional assessment studies on instructional videos are inconclusive, what Boucheix and Forestier regard unsystematic (2017). One broad explanation for this unsatisfactory situation is the varying specificities that pertain to the intrinsic load of a specific learning situation (Spinillo, 2017). Another explanation, a little more to the point, is the plethora of research approaches in the fields of educational psychology, technical communication and information design that advocate different media format definitions. Unfortunately, many of these media definitions are unnecessarily vague. Most frequently used are the terms multimedia, animations and video. Dynamic, transient and film clips are also used to describe various time-based medial formats that can be both animations and live-action videos, and various hybrid forms of the two that may, or may not, include graphics, stills and, of course, audio. This has led to a difficulty in comparing results and, in turn, to understanding what the efficacies of live-action format really are, as noted by Höfler and Leutner (2007).

Inconclusive results and unintelligible definitions notwithstanding, this thesis has developed from a wider and evolving tradition of research on instructional videos. In such research there is evidence indicating that the video format has the capacity to enhance content comprehension, by, for instance, reducing cognitive load (Lowe & Schnotz, 2008; Boucheix & Forestier, 2017; Swenberg, 2017), motivating learning (Rieber, 1991), effectively communicating real time reorientations (Tversky & Morrison, 2002), and, in line with this, by supporting procedural knowledge and learning skills with respect to dynamic processes (Boucheix & Forestier, 2017; Berney & Betrancourt, 2016; Betrancourt, 2005; Boucheix & Schneider, 2009; Tversky, Morrison, & Betrancourt, 2002; Höfler & Leutner,
The video format also facilitates quick procedural knowledge and good enough performance (Ganier & de Vries, 2016). On the other hand, there are some key caveats regarding the assumed efficacies of instructional transient media. For example, it generally presents information too quickly. Motion might blur important details and does not easily invite reinspection. This is a phenomenon often discussed in terms of “the transient effect” or “transience effect” (Sweller, Ayres & Kalyuga, 2011; Boucheix & Forestier, 2017; Leahy & Sweller, 2011; Paas et al., 2007; Wong et al., 2012).

In summary, then, the circumstances in which LAV-instructions might be useful and what design strategies leverage their possible usefulness are rather unclear. Adding to this general feeling of inconclusiveness are relatively recent research findings that contradict Just and Carpenter’s eye-mind assumption (1980). These findings show high levels of visual attention that should indicate an increase in cognitive processing do not result in superior learning outcomes (Boucheix et al., 2013; De Koning et al., 2010; Kriz & Hegarty, 2007; van Marlen et al., 2016). The surprising results of van Marlen et al. (2016) even imply that attention-guiding designs generated slower transfer problem-solving. Similarly, one study of online videos in a university setting by Lamb (2015) shows that the implementation of Cognitive Theory of Multi-media Learning-informed (CTML-informed) design guidelines in lecture-based instructional videos failed to generate better learning outcomes among students (although one positive outcome was that the CTML-infused design strategy resulted in the enhanced instructional videos being considerably shorter in length). This overall unsatisfactory situation regarding matters of users’ attention leads to unresolved issues with the instructional video genres’ efficacies. Moreover, this contributes to constraints on the sought-for transferability of results, for instance, from the domain of visual instructions to the domain of technical documentation (Große, Jungmann & Drechsler, 2015). At a more general level, this situation adds to the well-known miscommunication problems with visual instructions that, in turn, result in too many users ignoring them. Tversky calls this a “rampant” problem (Daniel & Tversky, 2012).

The question then arises, how can we assess LAV-efficacies most meaningfully? This question is a pressing one since the ongoing improvement of instructional designs depends on cohesive measures of the end users’ engagement that are more responsive to contemporary screen-based learning context. Further research into how physiological engagement and psychological engagement relate to performance outcomes is necessary for providing evidence for the quality of instructional tools (cf. Clark & Mayer, 2016; Henrie et al., 2015). We need more versatile engagement measures and comparative engagement methods that are more responsive to medium specificities, as well as the user’s complete visual attention spectrum and how it plays out in real-world learning scenarios. In other words, we need engagement methods that capture both online and offline behavior. However, no mixed methods eye-tracking studies of screen-based visual instructions, discussing the end-users’ performance and satisfaction (that I know of), fully explore possible influences of offline behavior and/or relate online behavior to relevant offline contexts. I suggest that both technical communication researchers and designers of visual instructions would be better off if there was more substantial research that is more combinatorial in nature. This research approach combines a perception approach that “transcends the vagaries of design fashion” (Ware, 2013, p. xvi) with other more intersecting perspectives serving to acknowledge the critical role of stakeholders (Krippendorff & Butter, 2007, p. 6), i.e. how these stakeholders make sense of medium-specific affordances in real-world learning scenarios.

1.5 Aims and Research Questions

The aim of this thesis is to illuminate relevant efficacy-related factors of instructional LAVs, including the underlying visual behavior characteristics and cognitive ramifications that can be linked to these factors.

The objective of this thesis is to identify the designerly implications of end-users’ engagement with the instructional LAV-format and its efficacy-related factors in order to support the end-users of
Implicit in the aim and objective of the thesis are the preunderstandings inferring that videography is a design activity, that learners’ mental efforts should be distinguished from task difficulty (Sweller, Ayres & Kalyuga, 2011; O’Keefe et al., 2014) and that users’ physical behavior in the moment of interacting with live-action video instructions can be more or less conducive to learning from visual instructional material (Clark & Mayer, 2016, pp. 219-233).

Here, it must be noted, that learning does not refer to learners’ long-term learning achievements, normally referred to as learning transfer outcomes. This is important, since there is wide agreement among scholars of instructions that video-based procedural content cannot facilitate long-term learning achievements in itself, since it does not, presumably, trigger mental images and/or information mapping the same way as abstract imagery does. This thesis does not claim otherwise.

To achieve the aim and objective of this thesis, the following research questions are posed:

**RQ1:** What medium specificities of the live-action video format may be attributable to videography?

**RQ2:** How do the medium specificities attributable to videography enable users to successfully act upon instructions in live-action videos?

**RQ3:** How do the medium specificities attributable to videography hinder users from successfully acting upon instructions in live-action videos?

### 1.6 Contributions

This thesis contributes to a more comprehensive understanding of humans’ abilities to interpret the actions of others and put this knowledge into use in actual hands-on applications. It does so by illuminating how this requires the involvement of our bodies and our motor system. By relating this to medium-specific affordances, this thesis furthers important distinctions between different kinds of medium-related efficacies and affordances.

More specifically, this thesis in part bridges the knowledge gap between film production discourse and the information design field. It does so, to a large degree, by translating and extrapolating pre-existing terminologies from film production discourse into the information design field. Thus, in this study, videography is posited as a representational topic that translates between film production practice and scientific knowledge. This is akin to how other design scholars approach certain design tools that are catalysts for design thinking (Cross, 1982), such as, the pen and paper in Wikström’s research on storyboarding (Wikström, 2013).

This thesis also adds new knowledge pertaining to the application of mixed-assessment methods, including eye-tracking methods, to the learning and instruction and information design research fields. A specific contribution are the HCD-inspired (Human-Centered Design) assessment approaches discussed in this thesis that serve to delineate value deficiencies in videography design enterprises that, when transformed into design solutions, can leverage instructional LAV-communication. By applying such a HCD-inspired research approach, this thesis complements contemporary ICT research (Information and Communications Technology) centered on information design matters that predominantly analyzes the affordance perceptions of individual media users, as opposed to also inferring them from observed behaviors. It is this more ecological and holistic research approach, centered on both data from interviews/self-reports and data from real-world observations, (including ET data), that makes this thesis unique and, I hope, worthwhile.
The industrial contribution of this thesis relates to its academic contribution outlined above, but more specifically pertains to highlighting designerly ways to enhance live-action video usability, i.e. the actual design prescriptions presented in the discussions section. These design prescriptions provide designers of live-action video instructions with the basic design framework to be implemented as videographer.

1.7 Delimitations

The sharing of useful knowledge and information by means of presenting them “live” is the focus to this thesis. The central aim of the study is to demystify the purported abstractness of this communication act. Along the same lines as information design scholar Andrew Dillon (2017), this thesis does so by applying science to information design. Another scientific way of doing this would be to employ media and communication theories that provide ways of conceptualizing and delineating the relationships between mass media, producers, sponsors and audiences of such media. In particular, communication theories belonging to the long tradition of what might be labeled media effects research could prove useful in this regard. However, it is not in the scope of this thesis to approach the issue of mediated communication in this way.

Neither is it in the scope of this thesis to comment on the various kinds of medial instructional, transient objects that exist in this communication landscape. For example, hybrid multimedia formats, such as, software instruction videos, that are partly live-action (see Morain & Swartz, 2012, and van der Meij, 2007) are not really contingent on the activity of videography and, therefore, not discussed in this thesis. It is also not in the scope of this thesis to delineate the affordances and expressive aspects of audio and audial means per se. When, and if, audio is discussed in this study it is approached as one of many modalities in the multimodal composition in question, i.e. it is not analyzed in isolation. However, this is not to claim that audio cannot be studied and assessed in its own right - which it certainly can.

1.8 Research Basis and Conceptual Linkages

This thesis is based on six mixed-methods studies and one conventional eye-tracking study, all revolving around different live-action video design projects. They are mostly comparative in nature and represent a research journey (from A to G). Papers A, B, C, D and E compare different media efficacies and affordances. Papers F and G analyze one medial object, respectively.

Conceptually, some of the studies have certain aspects in common, i.e. they partly address the same issues. In this way, Papers A and B resemble one another and relate to RQ1. Papers C, D and E belong to another group of studies that share certain common denominators. These studies address issues pertaining to RQ2 and RQ3. Papers G and F have a few aspects in common and relate to RQ1, RQ2 and RQ3. See Figure 1.
The first study (Paper A) is an audience reception analysis based on data collected from a comparative blind test that served to test the notion of *verisimilitude* in live-action video content. In this paper, notions of live-action medium specificities are key and analyzed in relation to both documentary conventions and digital recording technologies. The second study (Paper B) is an audience reception analysis based on data collected from a comparative blind test that consists of three live-action fiction videos (or films) that appear to be exactly similar but are not. The analysis addresses the question of how the videos are individuated on the grounds of what recording technology was used and how this impacts on users’ cognition. As in Paper A, Paper B discusses notions of live-action medium specificities and links them to conventions and digital recording technologies. In addition, Paper B circumscribes certain effects in terms of how images perform cognitive work.

The third study (Paper C) concerns a Lego-manipulation task/design challenge and is a cross-media comparison that also includes text and pictures. This study investigates user groups’ performance and behavior characteristics, including communication between team-members, before, during and after assembly. The fourth study (Paper D) is a pilot study based on the information equivalence research approach (as in Paper C) and compares diagrams with an instructional LAV. It investigates the connection between gaze and attention as well as certain aspects of the Visual Literacy-spectrum, defined as users’ *description style*. Thus, it is also an evaluation of the usefulness of a methodological framework consisting of six ET-measures. The fifth study (Paper E) explores whether GTS (gaze time on screen) can be useful as an engagement measure. It is based on data from a mixed methods ET study. The visual instructions in question, including one live-action video (LAV), depict a solar-powered toy at different stages of assembly, as in Paper D.

The sixth study (Paper F) is centered on a design project with several stakeholders resulting in the production and design of an instructional video that communicates how to build an anti-predator fence. This design project was administered in adherence to DRM (Design Research Methodology) and informed by a Human-Centered design approach. The aim of the project was to formulate design strategies that leverage its instructional affordances. The seventh study (Paper G) focuses on continuity editing in live-action documentary film and, by analyzing certain ET-data patterns, it aims to identify how visual behavior is constrained by discontinuity and how this affects audience members’ cognitive load. This paper also provides insights into how the mediated depiction of “reality” is contingent on continuity.

*Figure 1.* A stylized depiction of how this thesis’ research questions conceptually relate to the respective studies/papers.
1.9 Outline of Thesis

After this introduction, there is a theory chapter that presents and discusses relevant theories through an information design lens. After the theory chapter, there is a method chapter describing the general methodological approach of the thesis and relating this to the specific research designs of the different studies presented in the study. Following this, there is a results chapter. In this chapter, the principal results of the different studies, including my Licentiate thesis (2013), are highlighted. After the results chapter, there is a discussion and conclusions chapter that answers the research questions and relates the respective findings to the relevant theories and synthesizes these different strains into a coherent analysis in a conclusion. This conclusion, in turn, is then extrapolated into basic, designerly prescriptions for instructional live-action video makers.
2. Theory and Previous Research

This section presents the thesis’ theoretical framework, which is discussed through an information design lens. It concerns matters of both perception and reception and how these two activities coalesce in actual design practice. The theories in question are Ecological Theory of Perception, Cognitive Load Theory, Human-Centered Design, Multimodality/Social Semiotics and Medium Theory. The interdisciplinary nature of information design is reinforced through the diverse theoretical themes and topics included in this section. Thus the thesis takes an interdisciplinary approach to design. Moreover, the thesis bears the hallmarks of information design research efforts that aim to capture complexity and create new knowledge that becomes explorable when different disciplines and specific scientific perspectives are cross-examined, overlapped and integrated. In concurrence with information design scholars, Black, Luna, Lund and Walker (2017), I suggest that this approach warrants a more grounded application-oriented and usable scientific knowledge.

2.1 The Realist Theoretical Approach

At the heart of this thesis is a scholarly attempt to evaluate a certain kind of information design, namely the instructional LAV-format. Similar to information design scholars, Tufte (2001) and Holmes (in Heller, 2006), the focus here is to identify factors that directly or indirectly enable the visual information in question to be understood by a specific audience and/or user. For information design scholars there are a great number of different theories that can be used to do this. According to Mary C. Dyson, the applied theories by information design scholars do not commonly form a natural continuum; they do not necessarily represent a theoretical unity and have no clear boundaries (2017).

However, the theories discussed in this section are indispensable to a realist research approach (cf. Latour, 1999; Lister et al., 2009). This suggests that all the theories and previous studies discussed here, when combined, become maximally inclusive of cultural, physical and technological elements. This is important, since all these are elements of the reality being studied here. In other words, when it comes to instructional communication, I believe there are no entities deserving of study that are exclusively physical or cultural. Therefore, the examination of the relationship between culture, nature and technology is hereby key. This examination entails asking questions about what technology and medium really are, their meanings and affordances.

This kind of realist theoretical approach is not uncommon in qualitative research and scientific writing concerning New Media, digital materiality and interface design (cf. Twining et al., 2017; Lister et al., 2009). Consequently, this realist research approach is well aligned with the contemporary priorities of information design, Research Through Design-inspired (RTD-inspired) approaches, Human-Centered Design-infused (HCD-infused) research efforts, Human Computer Interaction (HCI) research and Information Communication Technology (ICT) research (cf. Dillon, 2017; Löwgren, 2016).

To more clearly grasp what this realist research approach entails in the thesis, it is fruitful to think of the aforementioned research paradigms and theories as interconnected, yet circumscribed theoretical
layers. These layers, analyzed as a disparate whole, illuminate various distinct aspects of how instructional designs reverberate throughout different timespans. By doing this, we are able to consider medial events in terms of how these events are constituted and, in turn, relate them to how the human mind and body processes them/reacts to them more or less quickly. Specifically, these theories combined illuminate how and why people react or respond to a medial event in a certain way. The theories in question thus revolve around different causal relations that span over various time spans. These conceptual time horizons can be very brief moments, or extremely long ones. Thus, when combined, these theories give a more comprehensive and developed picture of what factors condition perception, comprehension and usability-related factors in the instructional, mediated setting. See Figure 2.

![Theories' Time Spans Diagram](image)

**Figure 2.** A stylized depiction of how and why humans process a medial event and react to it, more or less, quickly. At the center of the depiction is the medial event. This event is reacted upon in different ways and this may be explained by situating a specific theory in accordance with specific time spans and history time horizons.

To a large extent, research on digital audiovisual objects relates to the currency of the literature in these fields. Videos, which in the aforementioned academic fields, more often than not, are dealt with as a form of ICT, tend to develop rapidly and their uptake in society happens relatively quickly. Therefore, I concur with Twining et al. (2017), that research concerning these fields requires reference to up-to-date literature. Nevertheless, many of the factors impacting on usability aspects remain surprisingly constant over time. Humans still inhabit human bodies with extremely similar cognitive architectures, notwithstanding our socio-cultural preferences and technological fads.

Moreover, despite changes in research approaches in design, media and (educational) psychology research fields over the past twenty years or so, there are reasons to believe that research findings have not substantially changed. Hence, this thesis includes both old and new literature and theories.

### 2.2 Videography as Design Nexus

What becomes immediately apparent upon approaching instructional communication as an information design topic is the deep and broad interest in factors that shape, structure, and condition
the possibilities for communication. Here, we should primarily consider the communication activities of videography and editing.

Strangely enough, videography and editing are not very often regarded design research topics. Neither are cinematic expressions a focal point in the field of information design, in spite of the fact that contemporary films and videos usually contain many graphical, information design elements. The research carried out by information design scholar Spinillo on animated, instructional pictorial procedural sequences is a rare exception in this regard (2017). Yet, I assert, along the same lines as design scholar and ET-researcher Swenberg, that it is a constructive academic pursuit to theorize the production of moving image medial objects and moving images by approaching cinema and videography as a design topic (Swenberg, 2017; Swenberg & Eriksson, 2018).

The expressive potential of videography is vast, since, according to Kress and van Leeuwen, the meaning potential of the production and design strata is high (Kress & van Leeuwen, 2001). In the realm of videography, these strata work in tandem and double the meaning potential when they become interwoven. This interconnectedness sheds light on the dynamic tension between the constitutive nature of communication and medial objects’ instrumental, but divergent, communication possibilities. With respect to videography, this tension may be described as a negotiation between an overall design strategy that includes predetermined expressive aims and, what Kress and van Leeuwen call “the actual material articulation of the semiotic artifact” (2001, p. 6). This involves technical skills as well as skills of the eye and hand. Thus, a detailed and purposeful blue print will not, in itself, warrant purposeful and efficient LAV-communication. A good blue print for design would not be enough since there is also plenty of instructional expressive potential that needs to be managed in the production and editing phase of videography.

The Design of Visual Fragments

To exemplify this, let us begin with two basic, complementary and fundamental design definitions. Simply put, one definition posits that design is about improving systems, services and objects. The other definition posits that design is about the creation of meaning. Hence, these two intersect, since improvement often results in added, new meaning, or the reduction of it. The field of information design is a case in point. In this field, practitioners and researchers are concerned with both of these design paradigms, since information designers are predominantly concerned with the organization of visual information. This represents a designerly urge to improve visual objects by creating new meaning (cf. Coates & Ellison, 2014; Black, Luna, Lund & Walker, 2017).

Simon suggests that design essentially is about transforming something given into something preferred through intervention and, in some circumstances, invention (1996). According to Simon, designers are occupied with how things ought to be, focusing on goals and functions (1975, 1996). In this process, the designed object acquires new meaning, and design is thus also about the creation of meaning (Krippendorf, 2006). Krippendorf interrelates meaning with the notion of the transformatory powers of human senses and perception. According to these design definitions, the study of design includes the study of transformations that may be experienced via human senses. These transformations result in the redefinitions of goals, or new, needed functions and/or new meanings. This is a dichotomy of the given and the preferred. Here, we may contextualize this dichotomy in the videography context. Documentary scholar Michael Rabiger notes that videographers construct a picture from “found objects” (2009, p. 206). Hence, although videography precedes editing chronologically, it is an editing activity, since the videographer edits in his/her mind while shooting. In this way, he/she co-designs the film with, among others, the editor. This is where consequential presence takes material form, where recorded fragments of reality become experiential, and where a phenomenology is described. This is not entirely unlike how the anthropologist approaches the craft of making field notes as one way to “design” his/her research output (i.e. books or scientific articles).
This overall argument, of the videographer as a collector of sorts, highlights the importance of the designerly craft of videography and its function as the supplier of audiovisual material to the editor. Thus, to a large extent, videography is about the selection of what actions and stuff are worth showing to an audience. But this is not enough. Videography also serves in ensuring that these selections become adequately enhanced and rearranged in the editing phase. The editor, on the other hand, cannot undo the videographer’s selections and must make the visual fragments of the actions stuff either collide or become linked, in concurrence with (or in violation of) stylistic or other aesthetic paradigms (Pearlman, 2017; Swenberg, 2017).

Visual Attention

When Kress and van Leeuwen discuss the concept of the grammar of visual design, they regard editing as the creation of a forceful linear text that imposes syntagmatics (2006, p. 208). Swenberg, on the other hand, is concerned with a much more subtle structuring of visual fragments. According to him, editing is the design process whereby an editor employs, what he calls, “perceptual precision” (2017). This concept infers that there are plentiful designerly options in video making enterprises, otherwise precision would be insignificant. In fact, with regards to video editing, there is a minimum of about 25 designerly choices per second.

This calls for a very finely grained and detailed analysis of the effects of video communication in which visual attention should be critically examined and fully accounted for. This full account includes the examination of both bottom-up and top-down cognitive processes, since they work in parallel but are separate cognitive processes competing for attention (Frintrop, Rome, & Christensen, 2010; Goldstone et al., 2015; Pinto et al., 2013; Swenberg, 2017). Hence, designerly activities that serve to merge disparate fragments into a functional communication object must strive to adjust audiovisual expressions in accordance with what is intentionally meant to be perceptually top-down processed. If the designer’s intentions are unmet, this can be proof that top-down processes have been de facto canceled out by unwanted, and perhaps unknown, bottom-up perceptual inputs. In summary, then, perceptual precision is here considered a key designerly concern that also relates to the design activity of videography. The important point here is that the creation of meaning within the realm of videography may be addressed as both an issue of delivering coherent content according to a plan of intervention, and, at the same time, an issue of the videographer’s awareness of the opportunities for bodily action or bodily reactions that some visuals trigger (for better or for worse). With regards to bodily reactions, similar to Franconeri & Simmons (2005), the primary phenomenon that is of interest in this thesis is movement, in particular, human movement (Kaiser, Shiffrar & Pelphrey, 2012), and the display of it. All in all, these are the micro and macro aspects of the videographer’s design thinking (cf. Cross, 1982).

To summarize, if we treat videography as a design nexus, or as an information design nexus, we will be able to better understand how videographers may manage uncertainty and the creation of new meaning. In doing so, we realize that the role of the videographer as an information designer is diverse. There are many ways a videographer may convey audiovisual information to an audience.

2.3 Design Principles and Cognitive Load Theory

The videographer’s diverse role notwithstanding, there are some critical guiding principles that videographers involved in LAV-instructions are presumed to adhere to. Such guiding principles further an important discussion of what a given medium best lends itself to.
Cinematic Design Principles

In one way, instructional LAV-content is governed by an old set of cinematic standards and design principles that may be regarded film production’s Best Practices. In the educational psychology and technical communication research fields, this is often implicitly and rather crudely addressed in terms of, for example, good audio, good lightning, good framing, good editing, and so on (e.g. Morain & Swarts, 2012). There are thousands of cinematography manuals, expert blogs, video tutorials and books that address these “good” practices, but in a more detailed and exhaustive way. In the film production literature these best practices are concerned with entertainment value, acting, continuity, composition, angle/perspective, luminance levels, color, equipment, teams’ skill sets, work procedures, script writing, directorial input, stunts, visual FXs, grading, and so on. These best practices translate into ways of thinking about how videography materializes Production Value. “High” Production Value is the opposite of something appearing amateurish (Eriksson, 2013). This is also what all video recording gear in itself aims to achieve. Even iPhones are designed to enhance production value when used as a video recording device (Eriksson, 2013).

Assessing Instructional Medias’ Efficacies

An interest in instructional information design reflects a concern for creating useful things and the process of creating useful things. Design can be a way of understanding communication and an approach for investigating the worlds of psychology and cognitive science from the standpoint of communication. Although this thesis does not pretend to be based on communication theory, the above claim has implications for theorizing communication phenomenon that can be addressed as kinds of cognitive problems. These cognitive problems require designerly solutions and, many would argue, design principles. Since the 1990’s, the revamping of older modes of instructional media for the solution to learners’ cognitive problems has been the implicit quest of two very influential educational psychologists - Richard E. Meyer and John Sweller - the founding fathers of cognitive theory of multimedia learning (Mayer), and cognitive load theory (Sweller).

The rapid development of media technologies and the emergence of virtual classrooms has resulted in many examples of designerly approaches of questionable quality, but also many good examples. This highlights the issue of assessing audiovisual objects’ efficacies. According to Sweller, Ayres and Kalyuga, instructional quality can only be assessed if designers consider humans’ cognitive architecture: “without knowledge of human cognitive processes, instructional design is blind” (2011, p. V). One of the most straightforward ways to determine an instructional media’s quality, regardless of type, is to assess its efficacy by evaluating learning or knowledge transfer outcomes - and there are many ways to do this. All such assessment methods supposedly measure how well a specific designerly approach supports learners’ cognition. This is indeed the basic assessment approach of this thesis, as well as the approaches of Sweller and Mayer.

Multimedia Design Principles

In e-Learning and the Science of Instruction (2016), Clark and Mayer describe proven guidelines for designers of multimedia learning. “Multimedia”, in this case, is considered a static and interactive presentation mode with graphics (that in Mayerian terms may very well be pictures and photographs), with or without audio. The reason for these multimedia design guidelines being so-called proven is that they relate to a vast body of research evidence, amassed over the years, by Mayer and others. This research evidence provides clues into how human cognition is supported, or not, by application of certain design guidelines. Below, is a summary of Mayer’s instructional design guidelines. These may, in turn, be broken down into several sub-categories that are addressed as principles, all of which are thoroughly discussed in e-Learning and the Science of Instruction:
Guideline No. 1: Use words and graphics (images) rather than words alone.  
No. 2: Align words to corresponding visuals.  
No. 3: Present words as audio narration rather than on-screen text.  
No. 4: Explain visuals with words in audio or text but not both.  
No. 5: Do not add extra material.  
No. 6: Use conversational style, polite wording, human voice and virtual coaches.  
No. 7: Manage complexity by breaking down a lesson into parts.

Over the past decades various kinds of design guidelines have become plentiful. In addition to Mayer’s guidelines presented above, Colin Ware’s visualization guidelines amount to a little fewer than two hundred (2013, pp. 445-457). Sweller’s guidelines can be summarized into nine primary rules (Clark, Nguyen & Sweller, 2005), while Van der Meij and van der Meij’s software training video guidelines amount to eight (2012).

When trying to implement guidelines, such as these in LAV-instructions, a few issues become apparent. Foremost, there is the problem of choosing the right guideline and principle to implement. Since the media formats originally informing the formulation of these guidelines and design principles are mostly static, some kind of reconfiguration of the applied guideline is probably required. We can expect that this is indeed the case, since, simply put, transient media works in totally different ways from static or semi-static multimedia instructional formats (Clark & Mayer, 2016, p. 81; Lowe & Schnottz, 2008; Tversky, Morrison, & Betrancourt, 2002).

Specific design guidelines for instructional LAVs are not addressed in the design guidelines literature, although Sweller, Ayres & Kalyuga address a few broad guidelines (Sweller, Ayres & Kalyuga, 2011, pp. 222-226). In general, both Mayer and Sweller adopt a very cautionary stance to instructional videos, such as LAV-recordings. The basic problem with videos, they claim, is that they lack versatility (Clark & Mayer, 2016, p. 81; Betrancourt, 2005; Mayer et al., 2005; Tversky et al., 2003). Referring to Mayer’s comparative study of 2005 (using various visuals and sequences showing how a toilet cistern works), Ware also notes that static presentation modes generally work better than animations and that animated formats are associated with “extravagant claims” (2013, p. 337). Thus it is not surprising that in Mayer and Clark’s e-Learning and the Science of Instruction (2016), amounting to 500 pages, there are only a few pages devoted to what they label “animations” (p. 81-84). Referring to comparative media experiments with knowledge transfer tests on weather phenomenon and mechanical kinetics, Clark and Mayer note that the “passive” qualities of static illustrations with text allow for more active processing of information in comparison to animations, since the learners have to “animate” the still information in their minds, and such active processing augments learning. Moreover, they note that the lack of interactivity in animations presents an obstacle for learning, and that the transient aspect - commonly referred to as the transient effect - of animations imposes extraneous cognitive load due to the learner having to hold too much information in memory (Sweller, Ayres & Kalyuga, 2011, p. 219-229).

The Hands-on Advantage

However, Clark and Mayer recognize that there is some content particularly suited to the video format, for example, “descriptions on how to perform a motor skill” (p. 83). They refer to evidence from comparative paper-folding studies, as well as studies on tying knots, doing puzzle rings and assembly tasks (Ayres et al., 2009; Marcus et al., 2013; Watson et al., 2010). In contrast, it seems that static presentations are more effective when promoting understanding pertaining to instructions that are more explanatory, i.e. more conceptual in nature, whereas, transient media works very well to teach “hands-on procedures” (Clark & Mayer, 2016, p. 83). Ganier and deVries (2016) note that (live-action) videos provide users with quick comprehension, but that static representations are superior over time and, overall, generate higher quality outcomes. However, transient media can effectively show relationships that otherwise would not be visible, for example, distortions that invite interpretation, such as variable speed presentations, the interchanging of viewpoints, augmenting
displays with cues and “having objects leave a trace or wake.” (Hegarty, 2004, p. 345). Time-lapse photographic special effects, or the depiction of a procedure in slow motion, would be examples of this.

Many scholars recommend adding cues to instructional animations to lessen the cognitive demands of this presentation format. This indirectly explains why studies on cuing in animations are so plentiful within the field of learning and instructions, such as in the studies of de Koning et al., 2010, and Boucheix et al., 2013. This line of research is all about adding design elements to diminish the negative aspects of transience.

In summary, then, Mayer recommends, in line with Lowe and Schnottz (2008), and Höffler et al. (2007), to only use the demanding format of animation if it is really needed and can serve a useful purpose. However, it is important to note, at least in this context, that, as far as I understand, Clark and Mayer do not specifically address the format of live-action video or, to be more concrete, do not differentiate between different transient formats. To Clark and Mayer, these formats are all the same. Or, put another way, for Mayer the issue of transience is key, and this is what all animated formats have in common, regardless of genre.

Learners’ Engagement

However, both Sweller and Mayer are primarily concerned with matters that directly regard learners, not instructional media per se. The issue of students’ engagement is therefore key and “at the heart of all learning” (Clark & Mayer, 2016, p. 219). In a similar vein to Fredricks et al. (2004), Mayer distinguishes between two kinds of engagement: behavioral and psychological. He notes that behavioral engagement might add an extraneous load that distracts, and claims it is a myth that learners must be behaviorally active in order to acquire new skills. Psychological engagement, on the other hand, is a requirement for learning. Nevertheless, there is behavioral engagement that fosters psychological engagement (Clark & Mayer, 2016, p. 220-236). Namely, there is a relationship between the two kinds of engagement. Clark and Mayer suggest, therefore, that one of the challenges involved in designing useful instructional video content is to work out designerly strategies that promote learners becoming both physiologically and mentally active, noting that the rather passive observing of video lectures results in low gain scores.

According to educational psychologists, there are several aspects that obstruct learners’ engagement in training and learning settings featuring the video format. As suggested above, the most frequently discussed factor in this regard is the so-called transient effect, or transience effect (Ayres & Kalyuga, 2011; Boucheix & Forestier, 2017; Leahy & Sweller, 2011; Paas et al., 2007; Wong et al., 2012). This effect potentially reduces the positive impact that signals and guidance designs have on the allocation of attentional recourses (Koning et al., 2010; Boucheix & Lowe, 2010; Lowe & Boucheix, 2011; Boucheix et al., 2013), and how LAVs effectively may convey procedural knowledge, manual dexterity and manipulative tasks (Arguel & Jamet, 2009; Boucheix & Forestier, 2017; Ayres et al., 2009; Ganier & de Vires, 2016; Lee & Liang, 2012; Marcus et al., 2013; Wong et al., 2009).

In Cognitive Load Theory (2011), Sweller, Ayres and Kalyuga refer to the LAV-format as “video recording of real life events”, and argue that Cognitive Load Theory (CLT) can provide a better understanding of how students can become engaged when using videos, in spite of the transient effect (p. 222). First, they recognize learner control to be a decisive factor. The ability to stop or slow down the video is key. Second, they assert that segmenting, in effect, lessens the working memory overload. Finally, they contend that proper pre-training techniques may lessen the impact of the transient effect, primarily due to the fact that pre-training instructional approaches are likely to reduce a video’s overall presentation duration (p. 224-226). In summary, then, although there are ways to manage some of the problems associated with transient media, Sweller, Ayres and Kalyuga urge us to be wary of accidentally introducing videos as part of any education ICT effort.
In contrast to Mayer, educational psychologist John Sweller’s design guidelines are influenced by cognitive load theory (CLT). CLT in its current form has been expanded and anchored within a biological evolution framework (Sweller, Ayres & Kalyuga, 2011). In Sweller’s own words, the logic behind this approach in the field of learning and instructions is as follows: “If we assume the way we learn, think and solve problems is part of nature because we are part of nature, we need to know how nature learns and solves problems.” (2011, p. VI). This means that CLT takes into consideration biological primary and secondary information and regards biological evolution as a natural information processing system. This view acknowledges that communication can be understood as an exchange in a natural environment, and that this environment is an essential ingredient in the formation of messages. This formation of messages in the natural environment and in animals’ bodies takes place over great time scales and explains why some things are known without being overtly communicated and/or taught. At a concrete level, this communication system most likely takes place in creatures’ genome structures.

There are certainly many examples of this natural communication system at work. Perhaps the most striking examples are found in the realm of other animals. For example, biologically primary information is probably the reason migrating birds know where to fly, why young birds can discriminate between minute differences in singing styles (Wheatcroft & Qvarnström, 2017), and why reintroduced animals to a certain environment behave in strikingly similar ways, go to the same places, etc., to those of the same species that vanished many ages ago. In these cases, it is quite apparent that tuition is not the source of information.

When discussing evolution-primed communication systems with respect to humans, abilities such as hearing, speaking, and face-recognition come to mind. These abilities are acquired at a very young age without explicit teaching. In fact, for example, we would not know how to teach children to speak their native language or to recognize faces, and it would be a futile project to try (Paas & Sweller, 2012, p. 30). Moreover, according to Sweller, there is actually no evidence that basic decision-making skills, problem-solving skills and planning (i.e. means-end analyses) can be taught at all, due to these skills being part of our repertoire of biologically primary knowledge. In our past, we needed these thinking and problem-solving skills to search for food, defend ourselves, find shelter, give birth, attend to our young, move across difficult terrain, etc., and we could do so without needing specific instructions (p.10). Thus, biological primary knowledge cannot be taught, requires no tuition and can only be learned (Geary, 2008).

Sweller’s approach to natural information processing systems provides the rationale behind labeling LAVs as the perfect medial vehicle for exploiting humans’ cognitive architecture. Essentially, the rationale is that biologically primary knowledge is integral to LAVs, as they often feature real people who make actual real movements, are involved in basic problem solving, use human speech (often in our mother tongue) and exhibit naturally occurring cues and signals that essentially are gestures. Therefore, the teaching part in some LAV-instructions can be omitted. Immersion can be a good enough approach (Sweller, Ayres & Kalyuga, 2011, p. 5-6).

This presents another possible advantage for videographers employing the LAV-format, which is an important point. If some aspects of the overall pedagogical and instructional goal can be left without overtly attending to them, more effort can be placed in teaching what is normally the primary goal of any instruction: to drive the acquisition of biologically secondary knowledge. Biologically secondary knowledge is more recently acquired knowledge by societies for cultural reasons (Sweller, Ayres & Kalyuga, 2011, p. 3). This is also to suggest that previously acquired skills, biologically primary in nature, may leverage instructional attempts to promote acquiring biologically secondary skills (Paas & Sweller, 2012). The reason for this is that when we convey information belonging to the biologically primary knowledge realm, very few cognitive recourses are required. This frees up cognitive space for other more demanding mental activities associated with biologically secondary knowledge. Along the
same lines, Sweller suggests the activation of mirror neurons does not impact cognitive load (Sweller, Ayres & Kalyuga, 2011, p. 228), and that mirror neurons greatly facilitate the performance of movements (namely, if the movements in question are displayed by other human agents and are goal-directed). This furthers the notion that LAVs could be useful when the communication of human movements is key (Sinigaglia, 2013; Genuchten et al., 2014; Ganier & de Vries, 2016; Boucheix, 2017).

How Cognitive Load Theory Informs this Thesis

Cognitive theory of multimedia learning (CTML) and cognitive load theory (CLT) are important theories for this thesis, since this whole research project revolves around issues of instructional, technology mediated efficacies. If efficacy levels are to be established, they need to be measured, and the aforementioned theories offer ways of doing this. In addition, CTML and CLT provide a theoretical framework for judging the relevance of the results of the learners’ assessments.

2.4 Live-action Videography and Modality

The LAV-format’s purported aim is to construct audiovisual credibility. Live-action videography is the designerly activity where a videographer’s consequential presence takes material form (Eitzen, 2005). This form must withstand audiences’ and end-users’ scrutinizing of the recorded material’s reliability and authenticity (Nichols, 2010). Thus, the value-adding creation of modality markers that signify reliable evidence is at the heart of LAV endeavors. This is what Kress and van Leeuwen refer to as “designing models of reality” (2006, pp. 154-174). The question then arises, how does a videographer design a model of reality?

Designing Models of Reality

Social semiotic theory can provide valuable insights into how images’ visual design grammar forms apt expressions of the reliable and credible. Kress and van Leeuwen do this by employing the interchangeable terms “truth markers” and “modality cues” (2006).

The live-action video format embodies modality cues by, firstly, showing what is visible, and not what normally is invisible. Furthermore, LAVs show a large amount of pictorial detail, more so nowadays with high-resolution technologies. I point this out with the knowledge that not all of this information is made available for the brain. We can only pay attention to a relatively small number of stimuli at a given time (a phenomenon sometimes discussed in terms of unintentional blindness). Nevertheless, the absence of contextualization and naturalistic settings lowers modality. Things shown in a void - common in diagrams and animations - render objects generic, as opposed to specific. However, there are limits to cinematic naturalness, and overly enhanced naturalness often appears as strange. For example, if motion is too sharp, which it is when a video is recorded at 48-50 frames per second, audiences most likely feel a lack of motion blur.

LAVs also appear real because they show depth, or more specifically, the illusion of depth. The videographer shows depth by the application of depth cues (Cutting, 2005). Here, it is important to note that schematic, abstract images, such as diagrams and animations also show depth, but normally in much more contrived ways. LAVs may implicitly even show a fourth dimension, and occasionally subjects in video (and films) “break the fourth wall” by addressing the camera. Many how-to videos on YouTube employ such expressive techniques, especially those that display a guide, trainer, expert or teacher.

Moreover, LAVs appear credible since they show how time progresses. They do this although videos and films normally compress time via the technique of editing (Andy Warhol famously did not compress time in his films, and his documentaries were very long.). Furthermore, LAVs display a
close to normal color spectrum, a contrast range, i.e. brightness values. They also include shadows, as opposed to shading, which, in theory at least, represent “the fullest representation of the play of light” (Kress & van Leeuwen, 2006, p. 162).

**Degrees of Credibility**

A social semiotic analysis illuminates how credibility is constructed and communicated via medial means. Such an analysis aids in the identification of whether a proposition is presented as true, but it cannot help us point out if something is actually perceived as credible. However, it seems reasonable to believe that the more truth markers a video embodies, the more credible it is likely to be perceived (Kress & van Leeuwen, 2006, pp. 154-155). One pertinent question is whether all LAVs appear absolutely credible if they display all modality markers. According to Kress and van Leeuwen, the answer is no, since not all modality markers can be adhered to in the same frame or sequence. Moreover, according to Kress and van Leeuwen, “each realism has its naturalism”. In other words, computer generated images may also appear as credible. For example, fake moving image sequences that show (often famous) people saying and/or doing unexpected things, or simple stick-figures that move in a highly naturalistic way may look a lot more realistic than other more exact and full representations of humans.

In fact, with respect to notions of what constitutes realism, more information is sometimes less. (Anderson & Hodgins, 2005, p. 66). Accordingly, not all realisms are equal and embody the same meaning potentials. Photorealism, for now at least, is the dominant criterion for what is real and what is not. This is, what can be termed, a yardstick for the real, based on the appearance of things, and the correspondence between normal settings and the photographic setting. In summary, we can thus conclude that the issue of what is credible or not rests on cultural and historical standards. These standards further emphasize the LAV-format’s unique status in contemporary society. If we now move on to the particular issue of credibility in live-action *instructional* videos, we need to consider the above modality markers further.

**Modality Cues in How-To Videos**

According to Kress and van Leeuwen, realisms that pertain to instructional, technical and scientific visuals are of a special kind. Such visuals require another set of designerly inquiries into the workings of modality markers, based on the question of how they may be used by their users (2006, p. 164-166). In this context, the notion of mechanical and scientific objectivity is key, and how this objectivity plays out among the end-users in the fields of education, business, technology and science. Some scholars argue that photographic visual instructions disclose the photographic format’s historical status as a substitute for drawings and engravings copying nature (Daston & Galison, 2007, p. 115-190). This copying of nature rests on the, seemingly, paradoxical notion that the more a visual can generate the illusion of touch, smell and object properties, such as scale and size, the more realistic it is. This is sometimes discussed as hyper-realism and connotes the idea that scientific representations, if broken down into their meaning elements, often are based on abstractions, as opposed to naturalness. This is what information design scholars sometimes address as a mix between the literal and the metaphoric (Richards, 2017). According to Daston and Galison, this may be considered as a kind of balancing act. Objective visuals should be detailed enough to be perceived as objective, realistic and scientific, but simple enough to be easily decoded and understood. Therefore, they are also reductive, essentializing and idealizing. Thus, as noted by Daston and Galison, objectivity with regards to instructional imagery is a naïve and elusive notion, since idealization is a subjective, creative construction that, in addition, according to Daston and Galison, is informed by lofty epistemic ideals.
Coding Orientations

Another factor to be considered when discussing instructional and science visuals is the socio-cultural dimension of the education settings that frequently employ such visuals. This is a setting that, according to Kress and van Leeuwen, in itself promotes ideas of “detachment, abstraction and decontextualisation” (2006, p. 165). This argument truly touches on the social aspects of social semiotic theory. Lindstrand and Insulander (2012) consider exhibition design a focal point for such ideas and a setting in which narration and meaning making concerned with identity condition visitors’ engagement. As discussed by Kress and van Leeuwen (2006), this may be referred to as a problem of so-called coding orientations. This is akin to Daston and Galison’s discussion on objectivity in encyclopedic procedural instructions and how their decoding depends on persons’ trained eyes (2007).

Coding orientations are sets of principles that inform the ways messages are decoded by specific social groups. According to Kress and van Leeuwen, first, there is the principle of technological coding orientations that underscores some visuals’ assumed efficacies and appropriateness. For instance, color is commonly associated with low modality in science visuals. Second, there is the principle of sensory coding orientations. This principle stipulates that if pleasure and entertainment are considered success factors, then color schemes should be vivid. Third, there is the abstract coding orientations principle. This principle states that sophistication is represented through abstractions and simple designs in, so-called, high art. Fourth, there is the naturalistic coding orientation principle. This is the prevailing principle among ordinary people as members of our culture. However, as media technologies constantly develop and when new ways of thinking of style and fashion emerge, Kress and van Leeuwen claim that “naturalism is coming into crisis” (2006, p. 166). Fake LAV-based “news footage” exemplifies this crisis.

How Modality Markers Inform this Thesis

The theoretical concepts of modality markers and the activity of modality marking provide analytical tools for discerning how authenticity and credibility are constructed and communicated in instructional live-action videography. This kind of theorizing reminds us of how notions of realisms are relatable to critical social constituencies and certain historical developments. Specifically, the social semiotic analysis approach is worthwhile considering, since it indirectly makes possible a more coherent and tightly interwoven design approach for instructional live-action videographers. This is an approach whereby instructional video makers can approach meaning potentials in a more fully responsive way with regards to the users’ assumed preference for the real. This entails an understanding of the scientific, stylistic, conventions on which serious, credible, visual instructional efforts still rest.

2.5 Direct Perception and Affordances

The specific affordances associated with the LAV-format can be discussed in terms of semiotic meaning potentials, as it is in the previous paragraphs, without using the term affordance. The social semiotic theoretical perspective is limited to illuminating the representational qualities of visual representations, not actual affordances tied to human’s perceptual abilities. For this we need to consult an ecological theory of perception. So, let us here for a moment explore this theory and its two main theoretical constructs: direct perception and affordances.
James J. Gibson, the father of ecological theory of perception, asserts that humans have direct access to their environment. Thus, according to Gibson, perception is not a mental affair. He suggests that the gateway to the environment is in the form of information provided by arrays of energy that are patterned and quantifiable. This information - perceived directly - enables man to actively listen and look for the things that the environment can afford us. Therefore, Gibson coined the term “affordance” to denote the “complimentarity of animal and the environment” (1979, p. 127). To perceive an affordance, according to Gibson, is to detect an environmental property that provides opportunity for action. This property is detected by moving through the environment. Hence, affordances do not primarily exemplify causality chains - a common misunderstanding - but possibilities for actions that have been enabled by complementarities (Faraj & Azad, 2012, p. 254). So, the argument goes: the solidity of the ground affords walking. Solidity is thus an affordance. Along the same lines, scholars of communication and visual artifacts often denote *visibility* as an affordance (Evans et al., 2017).

Interestingly enough, visual instructions in the form of moving images are forever connected to the genesis of the ecological theory of perception. During World War 2, Gibson was an US Air Force lieutenant colonel, and in charge of testing instructional films for pilots. The instructional films featured simulations of other aircraft and instructions about how to land an aircraft. This filmmaking experience made him realize that films could faithfully capture essential features of reality like no other medium before. He wrote quite extensively about this experience in *Motion Picture Testing and Research: Army Air Forces Aviation Psychology Program Research No. 7* (1947). Allegedly, the realization that the film medium was indeed a very effective training tool for pilots and that they, in effect, prevented young pilots’ premature deaths, supplied the basic empirical evidence that informed the (later) formulation of the tenets of the ecological perception theory. Thus, this theory is based on the notion that human vision is a flowing optic array, oriented towards the horizon (Reed, 1988, pp. 114-80). At the time, Gibson’s idea was considered a radical idea, and a naïve one, too. Prior to Gibson's ecological theory of perception, psychologists promoted the idea that the world out there was never actually seen, which puts the activity of interpretation center stage. Gibson looked at this issue differently, with the knowledge that young aviators primarily reacted to what they saw, rather than what experience told them (since they had no experience).

### Perception as Lawfully Constrained

The perception of films and videos is lawfully constrained (Anderson, 2005). When an event occurs in the visible world, the event itself is unique and produces “a disturbance in the air that is also unique.” (Anderson & Hodgins, 2005, p.61). In purely mathematical terms, it is possible to copy this specificity and this is what animators must cope with. Yet, the human visual system has been fine-tuned, in evolutionary terms, for millions of years, and readily reacts to visual events that are copies, i.e. subtle variations of real events. For example, it appears that it is quite easy with adept software, powerful computers and designerly skills to copy natural motion in terms of its kinetics for short periods of time. But after a short while audiences might react to the motion as being contrived and too iterative. Faking works, but only for a while. Naturally occurring randomness is hard to calculate for with algorithms. At least, this is the case for now. With development of artificial intelligence - AI - this is expected to change.

Incidentally, this explains why some people consider that celluloid film grain appears more life-like than the digital kind. The silver grain in films is randomly (and chemically) organized. Perhaps this explains why James J. Gibson noted that “Moviemakers are closer to life than picture makers” when talking about celluloid film (1979, p. 182). However, the so-called uncanny valley effect, i.e. when audiences experience uneasiness when realizing that someone who ought to be a real person actually is not (Lowe & Boucheix, 2011), only occurs when there is too much information that is inconsistent. Thus, we do not care about information or react to it, when it is not there. We do not, quite obviously, perceive something directly that is missing. This explains why abstract animations can be more life-
like than those that are more complete (such as in polygonal figures). Thus, direct perception theory can be applied when analyzing abstract images, such as diagrams, and not only photographic images (Shaw & Mace, 2005).

We can thus conclude that what sets most films and videos apart from other types of visual, symbolic representations, is that they engage our perceptual system directly, just as we encounter anything else (that is visible) in the real world. This would be especially true with regards to the LAV-format. However, it is important to note that the simple and straightforward concept of direct perception does not necessarily render other explanations for what maximizes our attention most. For instance, narrative contexts augment the communication potential in films and videos, in spite of the oftentimes questionable status of their imagery (Cutting, 2005).

Research on Affordances

A considerable amount of communication research has been carried out on technology mediation about affordances and this has increased in recent years. For example, some visual management scholars build upon so-called Affordance Theory in order to explain why some visuals may improve work in practice (Beynon-Davies & Lederman, 2017). Another concrete example of affordance-driven research is the theoretical contributions of information visualization scholar Colin Ware (2013). Although he sees problems in taking Gibson’s theory literally, he acknowledges that Gibson influences much of his own thinking and that Gibson’s affordance construct is extremely useful from a design perspective (2013, pp. 17-20). Other affordance-driven research efforts do not even make reference to Gibson. For example, Machin’s research on multimodal discourses (2010) builds on Kress’ usage of the term and is inspired by Kress’ Multimodality - A Social Semiotic Approach to Contemporary Communication (2010). The diverse ways scholars apply the term affordances in communication and visualization research demonstrate the multifaceted nature of affordances.

Technical communication scholar Christa Teston explores the affordances of medical images and specifically looks at how they are mediated and “make present” the material characteristics of disease, in addition to how they facilitate (speech-based) rhetoric (2012). Teston suggests that medical professionals construct meaning by interacting with images and she identifies four ways they do this (p. 193). First, the imaging experts (at tumor board deliberations) compare different sets of images. It is the variations between images that are key. This, she states, affords understanding of how tumors shrink/increase over time, whereby the professionals know whether the cancer treatment works, or not. Second, particular material conditions can be expressed and understood by looking for and identifying images’ variations in contrast. This affords, she claims, seeing metabolic processes (among other things), whereby the professionals know if the cancer is malignant. Third, they look for variations in proximity. This reveals how tumors affect close-by areas. By seeing this, they can estimate how a patient’s life might be affected by the cancer. Four, variations in perspective afford, among other things, understanding how actively the cells are dividing, and, by seeing this, the medical professionals can know the stage of the cancer (p. 204).

Taken together, the dichotomy of seeing and knowing translates into Teston’s taxonomy of affordances. These affordances are thus both perceptual and argumentative. According to Teston, this taxonomy demonstrates that specific affordances are detected when objects (images) are viewed from various distances, when they display different color-schemes, when they are placed side by side, when they are viewed over time (at different times), and when image features blend into other categories of images. This blending is a kind of transformation that is the detection of an invariant in a changing environment (interestingly enough, computers can also detect these so-called affordances and become good at it over time). By identifying these affordances, Teston implicitly contributes to a practical contextualizing of Gibsonian concerns, such as textual gradients and invariants in an image-practice context. In so doing, she concludes that there is more to images than their “representation of an a priori reality” (p. 205).
Contemporary Affordance Definitions

Teston’s study is an example of the affordance perspective that was popularized in human-centered design and computer interaction research by Norman (1988). Norman approaches affordances as design aspects of an object. As opposed to Gibson’s findings, this psychological perspective advocates the view that affordances are never actual; they are only perceived and depend on a kind of communication between object and perceiver. Norman claims that this communication is essentially what good design is all about. Hence, to Norman the affordance of what he labels discoverability is what leads to users’ successful interaction.

There are other perspectives on direct perception and the affordance concept that cast doubts on them. For example, some explanations include a special kind of selective attention, discussed in terms of visual illusions. This is to suggest that humans can fill in missing visual gaps by detecting optical variables (de Wit et al., 2015). Some researchers emphasize observers’ varying detection capacities (Withagen & van der Kamp, 2010). In other words, the detection of an affordance is a context-sensitive matter. Other scholars suggest that there are ways to actually educate perceptual systems so that they are in “closer alignment with the requirements of formal reasoning” (Goldstone et al., 2015, p. 28). Taken to an extreme, various kinds of such so-called indirect perception propositions highlight a major concern. This concern has not been fully resolved in the field of psychology and vision research. We know that hardwired perceptual systems can be tainted by preconceptions. But if our perception of the world in part depends on experiences and wishes, then there is no unbiased evidence about reality. In a nutshell, the problem is that hallucination is extremely counterproductive in evolutionary terms (Goldstone et al., 2015).

However, there are communication scholars who advocate more stringent affordance definitions that purportedly are more fully aligned with Gibson’s original intentions. Evans et al. (2016) stipulate that lists of affordances should include conceptually developed individual affordances, not just features, objects or outcomes. Evans et al. (2016), as well as Rice et al. (2017), highlight the agency aspect of the affordance theoretical construct. This furthers the notion that an affordance belongs to the relationship between humans and their perception of the environment. For instance, a video camera’s variable shutter, big sensor, cinema lens, viewfinder etc. are not affordances. They are functions or objects, while, for example, recordability is. In accordance with this definition, what is listed as affordances by Teston, (see previous paragraph), are actually outcomes of affordances. When re-conceptualizing Teston’s taxonomy in accordance with the threshold criteria of Evans et al., other kinds of affordances seem to emerge, for example, translucence, transparence, comparability and searchability.

Moreover, visibility, editability and searchability all count as valid affordances, according to Evans et al. (also see Rice et al., 2017). According to Evans et al., affordances are never embodied with a technology or an object. For instance, signals in instructional designs are not affordances, neither is attention - the outcome of the employment of signals and cues. Attention is not an affordance since it does not denote what takes place in between the human and the perceived environment, i.e. it does not represent a relational perspective (Leonardi, 2013). Moreover, it has no variability, while an affordance always has. Attention is either on or off. And so is continuity. For the sake of clarity, here it is important to note that Gibson’s position that affordances either exist or not - i.e. that they are actual - does not contradict the notion of affordances’ variability. Hence, continuity is not an affordance, it is either experienced or not. However, there is another term and topic that is closely aligned with attention but that to a much greater degree corresponds with the threshold criteria of Evan et al. (2016). That is engagement. It is not a feature, it is not an outcome (but closely coupled with it), and it varies. In this thesis, engagement is considered an affordance rubric that essentially encompasses other affordances (see the Results section).
How Direct Perception and Affordances Inform this Thesis

Direct Perception highlights the aspects of instructional live-action video making that are relatable to humans’ bottom-up perceptual processes. These are processes that humans cannot actively control. This sheds light on the diverse role of the videographer: he/she must not only consider modality markers, narrative structures, making and overall usability aspects, but must also consider how perception is lawful, in order to ensure that bottom-up processed does not cancel out necessary psychological engagement. Moreover, the ecological theory of perception paves way for insights relatable to the aspects in the human environment that provide possibilities for action. This is exactly what live-action instructional videos aim to do - provide possibilities for action. Thus, affordance theory makes the aspects of what facilitates action in specific environments explicit as a kind of first order understanding for the videographer.

2.6 Human-Centered Design

In the following paragraphs, I will briefly outline how technical communication initiatives can be modified and approached in accordance with the demands of complex systems and stakeholders using a human-centered design (HCD) framework.

Human Acceptance of Solutions

Purposeful support tool designs should, as stated by Rouse: enhance human abilities, help overcome human limitations and foster human acceptance (2007). In this context, the first two of Rouse’s HCD guidelines are self-explanatory. Good instructions primarily leverage humans’ abilities to lead better lives, by alleviating the constraints of their cognitive architectures. The third guideline, “Human Acceptance”, infers an important preunderstanding of the research that underpins this thesis. Worthwhile instructional design projects ought not to be only about issues of usability, but should include an expansion on the view of the meaning of design: designerly efforts should affirm human dignity (Buchanan 2001a, p. 37). As proposed by the Hasso-Plattner Institute of Design at Stanford (also known as the “d-school”), this advice is reminiscent of the first phase of design thinking, “Empathize with your users”. This reminds us that useless instructions and/or support technologies do not affirm human dignity. Such instructions make us frustrated, and waste our time. The worst examples may even result in fatal accidents. This issue may also be addressed in terms of Value Sensitive Design. This is a theoretically-driven approach to the design of technology that accounts for human values in various design ventures (Weibert, Randall & Wulf, 2017).

The Designerly Dialogue

According to Cross (1982), designerly ways of knowing entail various kinds of reasoning activities, which Wikström regards as a reframing of the problem of space (2013). Following this line of thought, some design scholars argue that this reframing can only be realized if transformation and meanings are “recursively woven into interaction with others.” (Krippendorff, 2006, p. 70). This, according to Krippendorff (2006), is a “second-order understanding” that recognizes that since meanings ultimately depend on the perceptions of consumers and end-users, designers must interact with them and engage in a “conversation”, since this second-order understanding is dialogical in nature. Krippendorff’s ideas on what he labels “the semantic turn” in the design of artifacts provide a window into the proposed virtues of HCD. HCD is a philosophical, designerly stance that places emphasis on peoples’ natural needs and behavior in their natural habitats. It thus deemphasizes users’ demographics in favor of their behavior, and evaluates it in natural contexts, as opposed to artificial ones. Moreover, HCD-inspired research and design processes rely on so-called dynamic conversations with users rather than
scripted interviews. Hence, proponents of HCD consider designerly dialogue with end-users and other stakeholders key, and claim that sound design processes should aim at creating functionalities that serve all stakeholders and make sense to most (Krippendorff, 2006).

According to engineer and design scholar William B. Rouse, this dialogic design approach is also one way to indirectly recognize that designers need to embrace serendipity (2007). However, this does not mean that designers can perform their job relying only on chance. On the contrary, designers need to be “at the crossroads of serendipity” (2007, p. xvii). This is to suggest that designers need to be prepared for serendipity with a clear plan that then might need major adjustment. This, according to Rouse, is “smart” luck as opposed to “dumb” luck (p. xvii). In addition to having at least a notional plan, if designers are to stand a chance of running into smart luck, they need to embrace that a worthwhile design process is a process of “assuring that the concerns, values and perceptions of all stakeholders in a design effort are considered and balanced” (2007, p. 5). Rouse’s definition of HCD underscores the virtues of a changing business climate in which innovators are viewed as embodying the capacity to truly embrace change. Therefore, innovators can play a critical role in changing our world for the better (2007, pp. 415-16).

The HCD framework

Following this general line of reasoning and building upon Rouse’s three defining objectives in HCD endeavors - enhance human abilities, help overcome human limitations and foster human acceptance - Steen (2012) proposes that HCD approaches include:

1. Involving users to better understand their practices, needs and preferences.

2. Searching for an appropriate allocation of functions between people and technology.

3. Organizing project iterations in conducting the research, and generating and evaluating solutions.

4. Organizing multi-disciplinary team work (p. 72).

In line with Steen’s definition of objectives, it becomes clear that the HCD framework is not governed by a rigid, linear, step-by-step methodology or plan that consists of pre-defined iterations. Simply put, essentially it is not a concrete design process support approach, although it is sometimes dealt with as such, possibly due to HCD principles having been specified and standardized, as per the ISO 9241-210:2010. Rather, it is pseudo-philosophical in nature and can be viewed as a complementary design thinking support approach, or a set of designerly attitudes that serves to embrace a holistic plan. Technical Communication scholars Zachry and Spyridakis concur: “We believe that HCD is fundamentally about accounting for and reflecting shared human values in the creation of the technologies, artifacts, and systems that humanity shares in the collective pursuit of life.” (2016, p. 394).

Sociotechnical Implications

The contemporary form of HCD - which has indeed had a rather long history that is not accounted for here - represents a de-emphasis on the view that humans primarily are cognitive information processing organisms (Bannon, 2011, p. 52), a view also integral to HCI research (Human-Computer Interaction research). According to Bannon, HCD initiatives are governed by interdisciplinary approaches that highlight sociotechnical implications in academia, industry and in society at large (2011). In the field of Technical Communication, this means that technical communicators are encouraged to assume a stance that acknowledges that technical communication in education and
technical communication artifacts have socio-cultural implications that reverberate further than the rather uncontroversial and confined industrial context, and that TC scholars should not shy away from provocative research-based arguments (Zachry & Spyridakis, 2016).

Previous HCD-driven Research

In spite of the fact that HCD, when approached as a kind of theoretical framework, appears to be philosophical rather than descriptively instrumental, there are some good examples of when HCD-inspired ideas are put into practice. For example, Kluge and Termer (2017) highlight the design process of a fault-finding application for mobile devices and claim that there is a HCD methodology, and that this methodology was employed throughout their design process that proceeded in four distinct steps (p. 172 and 179):

1. In-depth interviews with end-users (with three primary questions: who will use the application?, what will they use it for?, and under what conditions will they use it?)
2. Development of fault scenarios
3. Layout and concept design development
4. Evaluation and testing in the field.

Kluge and Termer (2017) suggest that the so-called HCD methodology aided in maximizing technology acceptance, and conclude that the importance of the HCD approach during the development process was critical in the attainment of a deeper understanding of the domain within which problem-solving needed to be supported (p. 179). Along these lines, they further stress that “only if a certain level of user acceptance is achieved and if the technology leads to higher job satisfaction and is viewed as enrichment for one's work as well as a relief of cognitive load is there an added value of using new technology.” (p. 179).

Maritime technology scholars Costa et al. (2017) relate HCD approaches to activity theory, as proposed by Engeström (1999) and Latour’s actor-network theory (1999) in their study on the design of a ship maneuvering system (basically a computer display interface). The researchers use activity theory to understand technology usage in the context of human practices, and specifically employ activity theory to explain the challenges associated with the “HCD process” and principles in the industrial context (p. 56). Costa et al. describe HCD knowledge transfer and integration in terms of various actor-activities, that also include non-humans (e.g., HCD literature, project plans, screens etc.) by identifying connections between actants and how they mobilize knowledge transfer: “how they translate and negotiate terms, exert influence on one another and become allies to establish and achieve a common end-goal” (p. 56). Thus, the researchers consider the transformation of what they call HCD theory into industrial practice as a series of translations and negotiations.

Overall, the theoretical discussion of Costa et al. (2017) exemplifies the designerly urge to, on one hand, regard HCD as a philosophy, inviting the design team to seek out new solutions, and on the other hand, as a kind of design process model with clearly defined stages and objectives. What the researchers refer to as the “HCD process plan” is a plan that is based on the ISO 9241-210:2010 for human-centered design for interactive systems. These were the principles of ISO ISO 9241-210:2010:

1. The design should be based on the understanding of contexts of use (users, tasks and environments).
2. Users should be involved throughout the design process and the development stages.
3. The design should be guided and refined by user-centered/usability evaluation.
4. The design should address user experience.

5. The design team should represent multidisciplinary skills.

6. The design process should be iterative. The first iteration comprises the following stages: (a) Planning the HCD process; (b) Understanding and specifying context of use; (c) Specifying user requirements; (d) Producing design solutions to meet user requirements; (e) Evaluating the design against requirements.

In spite of the above ISO-HCD principles making the design process “less fuzzy” (p. 58), the HCD assessment of Costa et al. reads as a long list of advisory and cautionary statements with respect to how to implement what they consider rather elusive and unclear HCD principles. First, they note that the HCD approach presented an obstacle to being creative (p. 60). Secondly, the researchers raise the issue of whether HCD is really applicable in their (maritime) market, or not (p. 62). Thirdly, they note that the academic HCD jargon is not very useful in real-world design projects (p. 63). Fourth, the researchers address the issue of how stand-alone approaches employing HCD thinking have little impact, and suggest that whole corporate cultures and structures need to be permeated by HCD thinking in order for it to be able to successfully implemented (p. 63). Fifth, Costa et al. advice caution since HCD-inspired approaches may result in “the imposing of additional tasks into the industrial design cycle”. HCD can thus be regarded as “impractical” (p. 63). Finally, they conclude with the remark that HCD-infused design endeavors need more structured guidance and should be less governed by principles that require designerly input and interpretation by experienced designers: “The guidance provided to HCD novices also needs to include careful and explicit terminology that does not need to be interpreted.” (p. 64).

**How Human-Centered Design Informs this Thesis**

LAVs are complex support objects that need to be approached in concerted (and relatively longitudinal) ways that allow for thorough analysis and synthesis at the crossroads between the various needs of different stakeholders. Concretely, one constructive way of doing this is to engage in dialogical HCD-driven conversations with the stakeholders in accordance with the guidelines proposed by HCD scholars. When designing instructional LAVs, HCD-driven approaches ensure they are “good enough” for all stakeholders. HCD-inspired approaches therefore lend themselves to a more pragmatic way of thinking about designerly matters. If this pragmatism is entirely lost, instructional videos’ operational reach and usability aspects may not be fully comprehended and/or harnessed by videographers.

**2.7 Medium Theory and Human-Computer Interaction**

In Eriksson, 2013 (Licentiate Thesis), I describe how cameras in cine-production chains acquire transformative powers due to their *actor network* functionalities (Latour, 1999). These transformative powers were discussed in terms of possibilities for creativity and the potential expression of digital aesthetics. At the heart of this is the notion that digital cameras are, in part, computers and that live-action displays thus may be viewed as digital interfaces that invite human interaction. These lines of thought are relevant in this thesis too, in spite of this particular focus on digital recording technologies’ transformative powers not representing this thesis’ main theoretical strain. As far as instructional videos are concerned, most would agree that issues of perception and usability are more important than digital technologies. Nevertheless, there is no way of getting around the fact that there is a possibility, however banal this might seem, that aesthetics and digital materiality, and ways of intermingling the two, might influence usability factors, or vice versa (cf. Tractinsky et al., 2000).
Affective Responses

The potential relationship between usability factors and stylistic features is investigated by Tuch et al. (2012). Their study concerns online shopping interfaces that represent different usability levels and levels of aesthetic enhancements. Drawing upon HCI-inspired research, such as Tractinsky et al. (2000), Tuch et al. explore affective responses to display technologies in terms of the so-called “halo-effect”. This effect stipulates that the beauty of an interface might outshine all other (perhaps less attractive) features of an interface, whereby the halo-effect influences users’ evaluation of the whole system. The findings of the study of Tuch et al. indicate that Tractinsky’s notion of “what is beautiful is usable”, might be reversed to a “what is usable is beautiful” effect, under certain conditions (p. 1604). However, what exactly these conditions are remains to be established. According to the researchers, this raises the question as to what degree there are specific conditions in which perceived aesthetics and hedonic qualities by users are conditioned by experienced usability. Tuch et al. also highlight the possible implication that orderly, simple and organized designs may de facto be approached as usability features by users (p. 1604). This, they suggest, would explain why classical and simplistic design attributes correlate to (normal) usability factors.

The above assertion is concurrent with Hassenzahl and Monk (2010) and their suggestion that classical aesthetics differs from beauty. They suggest that classic aesthetics may be approached as a form of visual usability that complements interaction usability (Hassenzahl & Monk, 2010, p. 255). Moreover, Tuch et al. conclude that the impact of interface-usability on perceived aesthetics is mediated through users’ affective responses. In other words, it appears as though users’ perception of usability is partially based on actual usability factors (of the interface), whereas the evaluation of aesthetics is influenced by the user’s affective response. In turn, such responses might be caused by the experience of interacting with the interface (2012, p. 1605).

Medium and Digital Code

In this thesis, I acknowledge that LAVs are also a medium, having both a material property and being a mental intermediary. This is the opposite of being solely regarded as a form of media that essentially is an immaterial notion (and arguably more so in the age of digital media). Perhaps McLuhan’s ideas on technologies’ capacity to extend humans’ sensorium come to mind here. In other words, how the conflating of medium and technology places emphasis on the extension of human senses (McLuhan & Fiore, 1967). However, this would be a kind of deterministic idea that does not take into account media texts’ content or expression.

So, the term medium normally implies a certain de-emphasis on immateriality, story-telling, users’ motivations, the psychology of vision, taste and fashion, how visual objects are mobilized for certain ends, how they converge and revalidate with other media, and how they may be sponsored, politicized and consumed in society. A medium-driven analytical approach, on the other hand, may transcend the vagaries of instructional fashion and certain contextual factors. Instead, it will place emphasis on practical application of media technologies and medial objects’ ontological specificities. With respect to instructional LAVs, this seems an especially valid analytical approach, considering the confusing labels that are meant to denote various video genres’ medium specificities and their technological heritages (but seldom do so in a comprehensive way). As per the recommendation of Rice et al. (2017), this medium-focused research approach may also further the operationalization of the concept of affordances in other similar research efforts on medial efficacies and medium specificities, as it will make the comparison between and the understanding of results a lot easier.

The communicative ramifications of the digital have become increasingly complex to define and foresee due to the rapid advancement of digital media technologies that, in part, have made cine hardware more versatile, smaller and more powerful. This technological development cannot be
adequately addressed without first defining what constitutes materiality in the realm of digital production, and how digital material contributes to the aesthetic qualities of a particular design (Gross et al., 2014, p. 637). The definition of digital materiality will help us define what audiences and computer-users have in common with respect to how they approach matters of qualities of designs. Issues of materiality have become more prominent in recent HCI research efforts, due to the use of smaller computers and smaller, flexible computer devices generating an increased desire to understand the presumably new material characteristics of these objects. These are objects that not very long ago did not exist, or used to be of different physical dimensions and functions. This development has resulted in some HCI scholars having begun to address the implication of new media theorizing in the field of human-computer interaction.

Ideas on Materiality

The communication act of sharing becomes a lot less abstract if put in relation to the notion of materiality, i.e. the material dimensions of information. According to Gross et al. (2013), two key and basic questions that need to be addressed when examining HCI and new media and their converging trajectories are the meaning of medium and the stuff that constitutes materiality. These scholars ponder whether materiality is a physical dimension of a given interaction between a specific physical feature and the perception of its meaning, or if it is something akin to computation itself, something more metaphysical. Gross et al. take the position that the key insight to a productive understanding of materiality is working towards, and even reinventing, the conceptual role of medium (p. 638). In their view, it is the medium that signifies and embodies the relationships between physical computer gear (such as digital film cameras) and digital code. According to Gross et al., this reinventing consists of combining HCI theories with other views on matters of materiality from the art and media theory fields, in order to explore and justify the cross-pollination of concepts and ideas (p. 638). According to Gross et al., (2013, pp. 638-641), specifically, the different strains of ideas on materiality that need to be explored are:

1. Materiality as functional affordances. This ontological view places value on how efficiently material can function as evident in, for example, tangible user interfaces that represent functions with graphically expressed objects.

2. Materiality as a metaphysical component. This ontological view emphasizes the materiality of computation and information, i.e. non-perceivable workings of a computer that can only be studied and used ‘for design in composition with other materials’ (Vallgård & Sokoler, 2010).

3. Materiality as conventions and traditions. This view places value on craft-related communication in networks and assumes that digital materials are inseparable from traditions of practice, and that the expressive qualities of these digital materials benefit from refinement and reinforcement over time.

Gross et al. conclude that the materiality theory of the HCI field is broadly compatible with medium theory, and that the implication of this is a fruitful distinction between physical medium and artistic medium. This highlights how any mediated communication effort is both physically determined and shaped by conventional styles (p. 648). In concurrence with Dewdney and Ride (2014), Gross et al. suggest that the close scrutiny and analysis of interactive digital artifacts involves an understanding that it is unlikely that there will ever be one over-arching and totalizing account of how digital code as medium register can transfer meaning across the technological means of human communication. However, a useful starting point for understanding all of this will include the analysis, and critique, of the notion of medium specificity.
How HCI research on Materiality Informs this Thesis

HCI-driven theorizing can provide valuable insights into how digital code also functions as a signaling medium. Insofar as materiality theorizing does not completely eliminate one set of concerns in favor of another, HCI-driven research on digital materiality informs this thesis by highlighting the conceptual tension between conventional, stylistic expressions and so-called, functional affordances. This relationship is fraught with tension. This is the relationship between the designerly urge to, on one hand, adhere to conventional expressions and style, and on the other hand, to fully exploit the affordances of the medium by exploring the ways in which physicality is brought into tactual and embodied human practices. These two urges may not harmonize, but must be negotiated and consolidated, one way or the other, by the videographer.
3. Methods

In this chapter, I will present the methodological framework of this thesis. The chapter is divided into two main parts. First, I will discuss the methodological approach and this thesis’ overall research process in more general terms. Then, in the second part, I will specifically outline the respective studies’ methods and describe how I have carried out the research and justify why I have designed my research in a certain way.

Ultimately, the value of any research effort should be assessed by its ability to generate meaningful and relevant answers to the questions that instigated the research project in the first place. In this thesis, such meaningful and relevant answers depend on the ability to identify key assumptions about data categories, how the data in question was generated and how all of this is underpinned by ontological and epistemological positions. Hence, I will describe the research design and methodologies of the thesis using relevant terminologies to underscore this research project’s unique qualities, theoretical underpinnings, scope and limitations.

3.1 Overall Research Approach

This thesis includes both methods that generate numerical and non-numerical data. For example, focus group interviews, reception tests, questionnaires, self-reports, learning tests, YouTube analytics, eye-tracking and video-recorded observations. These methods are used to gather relevant data that may indicate something valuable regarding the interpretations, assessments, perceptions, preferences and sensational experiences of live-action video users. It might be advantageous to discuss the methods and data categories of this thesis in terms of how they fit into an overall research approach that encompasses a research design. This design implies a goal-directed research process over an extended period of time, linking research questions with the methods, synthesis, analysis and conclusions, in a logical (but not necessarily linear) way.

The particular analysis present in this thesis is based on different categories of data. The critical question is not whether the thesis includes a mixture of non-numerical data and numerical data, but how this data is dealt with. Both numerical and non-numerical data is viewed similarly; thus all data is regarded as a kind of representation, which needs to be interpreted. Its meaning is context-dependent and subjective, and in this thesis, data is used primarily for inferential purposes. For example, if a learner is associated with few build errors, I regard him/her as having a high ability to act upon a particular visual instruction. Following this, I regard the build error measure as a performance measure. Therefore, the overall research approach is qualitative in nature. However, as recommended by Twining et al. (2017), in order to avoid any confusion, I will not use the term qualitative with regard to the respective methods. I will use the terms numerical and non-numerical when referring to the type of data that the different methods generate.

The methods discussed in this chapter are considered not only appropriate but also congruent and feasible with the other aspects of the thesis’ overall research approach. In the following paragraphs the methods are described in detail. In agreement with Spencer et al. (2003), this specification includes
detailed descriptions of the methods and instruments and, in addition, notes on limitations and research design implications.

**Research Through Design**

The research approach described above is influenced by several research through design-inspired projects (RTD). This means that the numerical and non-numerical data collected are the result of projects that involve the actual production and design of various videos. Thus, I am the researcher, designer/producer and co-producer. Typically, RTD-inspired projects involve the development of prototypes that may be viewed as a finished product that plays a central role in the research and knowledge generating process. This prototype-approach forces the designer/researcher to navigate around the real-world hindrances between the product and its usership and, in doing so, making them explicit to other academics (Giaccardi et al., 2013). Moreover, this also translates into a professional stance with regards to design endeavors in general. This is to acknowledge that complex design activities require both explicit knowledge and tacit knowledge and that designers with ample design-related experiences could make better sense of both these types of knowledge (Löwgren, 2016).

**Being an Insider: Ethical Considerations**

In this thesis, I address issues related to user agency by engaging in observation that generates both quantitative and qualitative data categories. In design research literature on efficacy-related affairs, and in qualitative research in general, much attention is paid to the role of the researcher as observer, as well as the complexities associated with potentially opposing perspectives between the researcher and researched. Although the value of reflexive self-awareness among researchers sometimes is contested (excessive self-analysis at the expense of the research aims should be avoided), most scholars agree that the practice of continuous reflexive awareness in qualitative research on human subjects/animals is paramount. However, there is not one one-size-fits-all solution to the problem of asymmetric relationships in qualitative research (Holian & Coghlan, 2013).

In this thesis, the problem of the inherent imbalance in the relation between the observer and the observed has an added dimension, since the researcher embodies the dual roles of being both an observing researcher and a designer of the objects of study that the observed enact upon. This calls for further nuancing of how knowledge is co-produced in this thesis, with respect to the research situations in which the researcher can be considered both an insider - having experienced the stimuli extensively beforehand - and an outsider.

This particular dynamic has two interrelated dimensions that may be addressed in terms of research challenges related to insider inquiries (Coghlan & Brannick, 2010). First, this highlights the issue of identification, which here prompts me to make a declarative statement. In this thesis, I identify both as a cineaste of the trade and as a scholar/researcher who carries out research on video instructions’ efficacies. The advantage of this dual-role approach is that by maintaining creative control in this way, I can concentrate on the ramifications of intentionality in a straightforward way, without having to resort to guesswork. In other words, by actively engaging in the production of live-action videos, I can better judge the rationale behind erroneous design choices and, therefore, also judge whether a proposed alternative approach would generate better results. However, this also leads to a potential disadvantage, the problem of knowingly, or unknowingly, misinterpreting the data patterns linked with the stimuli, and/or, more generally, losing sight of the big picture that perhaps is saying something other than expected. This is the second challenge of insider inquiry, namely the issue of a researcher’s preunderstandings and how one should, as Holian & Coghlan put it, “hold both closeness to the data and have distance to it” (2013, p. 401). In this thesis, this research problem is mitigated, primarily, by a research approach in which transparency is key. This means that my own designerly rationales are discussed in a transparent way. Thus, when I address causal relations with respect to stimuli and the
subject of who is being stimulated, I strive to take into consideration other design professionals and theorists’ ideas with regards to a certain design strategy’s validity, overall efficacies, and validity of data categories. Specifically, this means that in the articles serving as the basis for this thesis, I have proceeded as follows:

1. Described the prevailing theories
2. Described the methods
3. Described specific methodological constraints
4. Justified assumed relations between learners’ attention and specific areas (AoIs/Areas of Interests) of certain images/video sequences
5. Explained why certain data categories are relatable to these specific areas of certain images/video sequences in a detailed and exhaustive way.

3.2 Sampling and Sample Sizes

All the methods presented in this thesis involve scientific sampling in order to obtain an overall perspective over a large group of data. Considering this thesis’ overall qualitative research approach and its epistemological and ontological view, it is important to note that I do not consider the samples to be representative of the entire population. In other words, the sample sizes used are smaller than is required for extremely broad generalizations, a common problem in qualitative studies (Twining et al., 2017, p. 7). Nevertheless, the findings of this study may be considered valid/applicable across contexts.

Firstly, I extend its findings’ relevance beyond the boundaries of video production through the development of theory. Secondly, I relate a certain research setting to other similar research settings. By doing so, this thesis’ findings will resonate and may thereby help to illustrate or suggest explanations for phenomena in other similar settings. Therefore, on the whole, the sample of this thesis is considered to be broad enough to capture the many facets of the phenomenon studied, and limitations to the sample are clearly justified.

Being open and clear about one’s research methodology is a prerequisite for any researcher who wants to be transparent, trustworthy and credible. Building on Spencer et al. (2003), below are some key qualitative research guidelines as summarized by Twining et al. (2017, p. A5). These guidelines are elaborated upon and justified in this thesis, and are more specifically commented upon in the respective papers/studies:

1. On what grounds the sample was selected
2. How the sample was selected (e.g. purposive, convenience and typical cases, etc.)
3. How participants were approached and possible implications of this
4. Gaps in coverage and non-participation implications for the study
5. Sample size and profile of the sample.

3.3 The Qualitative Research Approach

This thesis’ qualitative research approach reflects the aim and objectives of the study and also frames
the research questions, essentially making them a little more explorative in nature. Some scholars (especially in the field of business studies) suggest that such explorative research questions align very well with research governed by so-called abductive reasoning (Bryman & Bell, 2015). Abductive research approaches are equal to the seeking out of the best explanation, among many plausible explanations, to certain “puzzling” phenomenon extrapolated from rather incomplete observations (i.e. small sample sizes). Here, such puzzles may be, for example, why young audiences prefer LAVs made on iPhones instead of on high-end cameras (Paper B), or why relatively low levels of visual attention during study-time among how-to video users do not significantly affect their performance levels, although the assumption is that it should (Paper E).

This explorative and qualitative research approach also reflects a paradigmatic scholarly stance with respect to how knowledge is constructed and expressed in this thesis. This stance is based on the notion that some topics and scientific investigations lend themselves to a hermeneutic and interpretivist research approach, or a more relativist ontological one. This thesis is one example of such an interpretivist/relativist ontological approach. This indirectly leads to one broad agreement in the associated research fields that qualitative research approaches should exemplify an alignment between the ontological (nature of reality) and epistemological (nature of knowledge) positions, the research aims and what methods are used (Twining et al., 2017).

From the perspective of a researcher engaged in a qualitative research approach, data gathering and analysis is inevitably a rather subjective affair. In order to enhance the credibility of such a research approach it is considered important to explicitly take the research's influence on the data gathering process and analysis into consideration. The way to do this is by being reflexive in one way or the other. Twining et al. (2017, p. A7) suggest that a researcher’s trustworthiness can be enhanced through analytic processes, such as the following:

1. Triangulation of data, i.e. using data from different participants, different settings, at different times

2. Triangulation of methods, i.e. using multiple methods to collect data

3. Give participants the opportunity to comment on transcripts and emerging findings

4. Triangulation of investigators, i.e. having two or more researchers involved in the data collection and/or analysis

5. Triangulation of theories, i.e. interpreting the data using two or more theoretical frameworks.

In this thesis, and in its associated studies (Papers A, B, C, D, E, F and G) all these credibility enhancement points are considered to various degrees and with different emphasizes. However, on a more general level, this thesis’ qualitative methodological approach is leveraged by a thorough and concerted analysis. In agreement with Avenier & Thomas (2015), this involves not only showing the relevant data (i.e. paraphrasing it) but, more importantly, interpreting it. Furthermore, the analysis part of this thesis’ research approach strives to move from descriptive information to patterns and abstractions. It is also explanatory, in the sense that I will use extracts from the data to elucidate how I have made sense of it. Finally, concurring with Spencer et al. (2003), I have tried to make the analyses convincing through explicitly seeking out and testing rival explanations and counter examples, if they exist.

3.4 Choice of Methods and Methodological Reflections

As previously touched upon, this thesis not only exemplifies a qualitative research approach, but also an explorative approach. This also means that the research process does not exclusively start from a predetermined starting point and then continues through a fixed sequence of increments, but it also involves interaction among the different design components. This kind of interaction requires several
different methods. This highlights the mixed methods approach of the thesis and why I have chosen certain design components and methods, as well as the mixed methods approach’s potential shortcomings.

3.4.1 Questionnaires and Self-reports

Film theorist and philosopher David Davies recommends that base-line assessments and comparisons of medial objects in audience reception studies need to ask the ontological question of “when do we have two instances of the same film, and when do we have instances of two different films?” (2009, p. 217). The simplest and most straightforward way to do this is to employ questionnaires. This method was used in the Verisimilitude Reception Test (Paper A), and Image Quality Reception Test (Paper B). The questionnaire method was also employed in the Lego Truck Assembly Study (Paper C). In all of these studies, the numerical data was statistically analyzed (in basic ways).

More informal screenings of video material in the field by users also play a role in this thesis overall research approach. A variant of the questionnaire method is the self-report method that generally requires less interference from researchers, and therefore may be used in such informal situations. It was employed in the Anti-Predator Fence Construction Study (Paper F).

The interpretation of the questionnaire data mainly provides clues to what aspects of the live-action medium are paid attention to and appreciated by users, and aspects that relate to interaction between team members. In addition, in this thesis, data from self-reports illuminates in what ways these aspects either augment or constrain instructional live-action videos’ usability, what I mainly discuss in terms of perceptual affordances.

The Verisimilitude Reception Study

In the study that compares different video sequences communicating degrees of LAV-verisimilitudes (Paper A), the data sets were collected from 75 audience members. Data sets were used as numerical data. The audience members were randomly approached, and were predominantly 18-25 years of age. The audience members were media-production, Bachelor-degree students at Dalarna University (at Mediehuset). Specific data on other various demographic aspects (including gender) was not collected. The stimuli were presented on screens of various sizes (including big movie theatre screens). The audience members were seated approximately 3 - 10 meters from the screen (the longer distance when they watched the videos on a large movie screen). Each screening of five short videos lasted for about 8 minutes. The screening was of the “blind-test” kind, i.e. hardly any information about the test/the videos was communicated before the screening. Directly after the screening, the audience members answered two questions (see Appendix for original Questionnaire in Swedish):

1. How truthful do you think this video is? Answer: 1–10 (10 is maximum truthful)
2. Do you think this video could be used as evidence? (Answer Yes or No)

The Image Quality Reception Study

In the study that evaluates various versions of the same LAV-fiction film (Paper B), both numerical and non-numerical data was gathered by means of a questionnaire. There were 61 audience members in total, all between 15-22 years of age. Some of the participants were media production Bachelor-degree students (Dalarna University, Mediehuset); while others were studying media in high school
Specific data on various demographic aspects (including gender) was not collected, except information on the educational program the informants were enrolled in. The audience members were seated approximately 3 - 10 meters from the screen. The screen sizes were rather large (like a movie screen in a small theatre). The questionnaire featured 5 questions, including both open-ended questions and questions that only generated numerical data. It was administered directly after the reception test/blind-test. BA-degree students, Johnny Isaksson and Anders Israels, administered the questionnaires for course credits (bq2014, DU-Course Code). These were the questions (see Appendix for original Questionnaire in Swedish):

1. In what ways do you think these films are different from one another?
2. Which version looked the best?
3. Which version looked like a movie you would watch in a theatre?
4. Which version sounded the best?
5. If there were any instances in which you felt that the audiovisual quality distracted you from following the storyline, what were they and when did they occur?

The Lego Truck Assembly Study

The Lego Truck Assembly Study of this thesis involved 15 participants in 5 groups (Paper C). The participants were randomly assigned to different stimuli groups: text only, text plus drawings, pictures only, video with narration and no-stimuli. The participants were students from the undergraduate Information Design program of Mälardalen University (except one student who was a PhD candidate in production engineering). The questionnaire included five open-ended questions, as well as several statement-based questions that featured a five-point Likert scale, from strongly disagree to strongly agree. (See Appendix for the original questionnaire). The questionnaire was also used to collect some basic demographic information about the participants. Demographic information was not collected, except for name, age and educational. The questions that were formulated in English revolved around issues that pertained to:

- The effect of medium on interaction and discussion
- The effect of medium on the understanding of the instructions
- The selection and application of possible solutions
- The effect of medium on performance time
- The effect of using instructions vs. no instructions.

The Anti-Predator Fence Construction Study

In the Anti-Predator Fence Construction Study (Paper F), one of the methods used was the self-report method. The self-report featured three open-ended questions. The answers to the self-report were subsequently discussed in a follow-up discussion (on the phone) with the respective managers of the two-fence construction teams. The self-reports were not coded and/or statistically analyzed. The following are the questions of the self-reports (see Appendix for original questions in Swedish):
1. To what extent the information presented in the video helped them to construct the fence

2. To what extent the content of the video was easily understood and applied

3. To what extent the video supported their knowledge and information needs during construction.

**The Questionnaire and Self-report: Methodological reflections**

Questionnaires can provide a good window into audiences’ first-hand reactions. They also provide insights into value statements that allow for a coherent synthesis of data categories that, in turn, are easily compared and quantifiable. However, they must include relevant and easily understood questions. If audience members/informants do not understand the questions, no reliable answers will be obtained. One way of partly overcoming this problem is to include questions that (if answered) may indicate levels of confusion/low levels of engagement. Question Number Four in the Image Quality Reception Test that asks, “Which version sounded the best?” is one example of such a “poser”. If answered, these data sources/answered questionnaires may then be discarded or analyzed for alternative purposes.

Self-reports, on the other hand, are more privy to possible sources of contingency and non-representation, since they, in effect, are not answered directly after the moment of interaction with the stimuli in question. However, this is not to say that this method is not valid. Similar to the questionnaire method, self-reports facilitate the quick and cheap collection of relevant data that may be both non-numerical and numerical. Follow-up discussions between researchers and informants help to ensure that questions are not misunderstood and allow for more probing, in-depth follow-up questions.

**3.4.2 Video Observation and Learner Tests**

Video recordings enable fine-grained analysis of communication, actions and social organization (Heath, Hindmarsh & Luff, 2011). The interpretation of the video data provides clues to what aspects of the live-action medium are paid attention to by users, and in what ways these aspects either augment or constrain instructional live-action videos’ usability. The data from video observations is non-numerical, except when it is used to time user actions, or for counting build errors. Video data in this thesis has provided me with the opportunity to examine the ways participants interact with various material resources (mostly videos and diagrams) to accomplish actions. These recordings may provide concrete data measures and/or more indirect data that, for example, may provide clues to issues that pertain to a certain study’s experiment design. For example, some video recordings (of the pilot-study in Paper D) revealed that one of the stimuli was not so user-friendly and therefore needed to be slightly altered, which it was (this updated stimuli was then used in the study in Paper E). The interactions that were video recorded are what I call learner tests. These learner tests are assembly activities in teams (Paper C), or by individuals (Papers D and E). Some video data of the numerical kind was statistically analyzed (analysis of correlations and ANOVA).

**The Lego Truck Assembly Study**

The Lego Truck Assembly experiment was performed at Mälardalen University’s user experience lab/space (see previous Questionnaire-section on other information with regards to sample size and demographic data etc.). In this studio, there were three mounted cameras and two cameras on tripods that were used to document interaction and behavioral patterns among design team members. The duration of the experiment was about one hour, consisting of five stages (see Paper C). The video
The video observation method generated the data categories/measures *number of stimuli reviews during assembly, duration of assembly sessions* and *off-line behavior* (including conversation/talk).

**The Solar-Powered Toy Assembly Studies**

The other two studies that included video recordings of assembly sessions (i.e. learner tests) were the studies investigating the relations between visual attention and learners’ performance (Papers D and E).

There were 12 participants of the Solar-Powered Toy Assembly pilot-study (Paper D), with an average age of 24 years, of which 7 were female and 5 male. The participants were randomly approached (on campus) and assigned to three stimuli groups: a structural diagram group (N=4), an action diagram group (N=4), and a video group (N=4). Participants were all ignorant of the purpose of the assembly sessions, but they were informed afterwards in writing, and confirmed the use of their generated data. All participants were students recruited from Dalarna University film and TV education program. Data on the participants’ educational background and mother tongue were also captured in order to be able to (later) discriminate between possible background factors influencing the viewing data. Participation was voluntarily and rewarded with a cinema ticket (€15).

There were 57 participants in the Solar-Powered Toy Assembly Study (Paper E), of which 25 were male, and 32 female. The average age of the participants was 26 years. All participants were students recruited from Dalarna University. The group consisted mainly of BA-students, with a few MA-students. About half of the participants consisted of a mixed group of engineering, technology and economics students, while the rest were media students (TV/Film/Graphic, Design/Commercials, Production/Music Production). A little more demographic data than educational background was also collected (such as participants’ native language). The participants were randomly approached (on campus) and assigned to three stimuli groups: a structural diagram group (N=17), an action diagram group (N=20), and a video group (N=20). All were considered novices with regard to the assembly task. This assumption was made due to the novel nature of the solar-powered toy to be assembled and not formally assessed. None of the students were at the time enrolled in the researchers/teachers’ classes and some were approached by a third party when this was considered ethically correct. They all received a movie theater gift certificate (€15 value) for their participation, regardless of whether their data was used or not. All participants gave informed consent (one pre-experiment, and one-post experiment) and their participation was voluntary.

The video observation method generated the data categories/measures *number of stimuli reviews during assembly, duration of assembly session, build error, off-line behavior* (including conversation/talk) and, indirectly, via means of the eye-tracker’s video camera the *GTS-measure* (see Paper E).

Predominantly HD-quality video, recorded on a Panasonic HVX 201, was generated, with a few exceptions (such as the video data generated from the Mac-Laptop, and the eye-tracker that is of overall lower quality than HD-quality). The camera was mounted on a tripod when data was recorded during the learner tests.

**Video Observations and Learner Tests: Methodological Reflections**

Video data allows for repeated viewing. Repeated viewing is of utmost importance if the validity of video-related measures is to be assured. It simply takes time to discern details. This highlights the ontological dimension of the (LAV-) medium as a detail-rich documenting medium that prompts researchers to take materiality and ecological situations seriously (Boucheix & Forestier, 2017; Heath,
Hindmarsh & Luff, 2011). Indeed, this “richness” is the main reason why video data is gathered by researchers in the first place. It allows for close scrutiny and researchers’ analytic appreciation of the “taken-for-granted complexities” of diverse kinds of interactions that involve humans (Heath, Hindmarsh & Luff, 2011, p. 13).

However, this kind of richness also presents some challenges, one of which relates to (the affordance of) visibility. If the video material is to function in optimal ways as data source, some technical expertise is required to ensure that the material may audiovisually reveal what it is intended to reveal. Another challenge has to do with project planning, in particular gaining access and agreement to record. In this thesis, these particular problems are handled through avoiding closed access settings and by ensuring informed consent from informants.

3.4.3 Eye-tracking

The Eye-tracking method was chosen because it provides unbiased and very detailed information that concerns visual decoding behavior and the distribution of visual attention regarding a stimuli’s AoIs (Areas of Interests) and/or a particular screen. ET-data is numerical data. In this thesis, the primary data categories of interest that the ET-method generated are gaze time on screen scores (GTS scores) and participants’ pupil dilation (that indicates cognitive load). The numerical data is based on the calculation of the focus of users’ gazes that is compared to the position of a near-infrared light (reflected in the eye) to the position of the pupil, as well as the size of the pupil. This calculation occurs many times a second. This information, combined with the position of the user’s head, may be extrapolated to determine the point the user’s eyes are focused on the AoI, and its corresponding screen coordinates. All relevant data was statistically analyzed in the respective studies (Papers D, E and G).

The Solar-Powered Toy Assembly Studies

The Solar-Powered Toy Assembly Studies investigate the relations between visual attention and learners’ performance (Papers D and E). All participants had normal, or corrected to normal, vision (see previous section on video observations for other information with regards to sample size, demographics etc.). The eye-tracking experiments were designed and overseen by myself and fellow researcher Thorbjörn Swenberg. First, the participants were asked to settle down for a moment, in order for all those participating to “assume a similar state of mind” (Holmqvist et al., 2011, p.115). Then, the participants were informed that they were to look at a visual assembly instruction on a computer screen, and afterwards they were expected to assemble the object depicted in the instruction. The participants viewed one of the visual instructions (1-3) each, one participant at a time, comfortably seated in front of the screen and the speakers. The stimuli were one structural diagram, one action diagram and one instructional live-action video. The stimuli sequence of the experiment also consisted of a picture from a story-book, a blank (grey) screen. The blank screen was used for re-telling of the story-book image.

The stimuli were run with SMI Experiment Centre 3.4.119 software. This is a table-mounted eye-tracker. We used a 9-point calibration with 4-point validation. The eye movements were recorded with a SMI RED250 stationary eye-tracker, sampling eye data at 120Hz, with iViewX 2.8.26. The LAV was generated at 25 frames per second and screened on a computer screen, Dell P2211, driven by a NVIDIA GeForce GT440 video card, in 1680x1050 px resolution and mp4 codec (SMI default). Light emittance was measured to 90 cd/m² (at brightness level 65%, and contrast level 75%) for a 255-255-255 white screen. The viewing position was 60-80 cms from the screen. The sound was calibrated as balanced around an 82dB threshold for pink noise in the listening position. To avoid disturbance to the
participant when running the experiment, the lab setting was configured with a wall-screen in-between the participant and the researcher. Eye movement data was analyzed with SMI BeGaze 3.0.181 software. No stimuli was considered harmful in any way. The participants were all ignorant of the purpose of the study when viewing, but were informed afterwards in writing, and they confirmed the use of their eye-tracking data. The eye-tracking research projects were checked by the researchers for ethical aspects, according to the local Bill-of-Self-Audit (Dalarna University Research Ethics Committee, 2008), and passed all stipulated criteria. Both procedure and stimuli were considered ethical in adherence to Dalarna University research standards.

The Cognitive Load Study

There were 33 participants in this study (Paper G), selected out of a group of 50 potential participants. 17 potential participants did not contribute to the data sets that were analyzed, due to (but not exclusively so) inferior calibration and low GTS-scores. 3 participants were graduate students, while the others were undergraduates, all recruited from Dalarna University. The participants were randomly approached on campus, and their average age was 26 years. All participants had normal or corrected to normal eye sight, and were unaware of the purpose of the experiment. 17 were female, and all but one of the participants (32) had no previous filmmaking/editing experience. A little other demographic data was also collected (such as mother tongue). Participation was voluntarily and rewarded with a movie theatre gift certificate (€10 value). The stimuli in question consisted of documentary live-action video (see Paper G). See previous section on the solar- powered toy studies for information on the apparatus and other relevant information.

Eye-tracking - Methodological reflections

Eye-tracking as scientific method raises the same information privacy and data protection concerns that recorded video data raises. When employing eye-tracking as a data generating technique, it is therefore of utmost importance that a working arrangement and trust with all stakeholders is established and that they agree to the research.

The eye-tracking method helps me to answer all of the research questions of this thesis, since the analysis and interpretation of certain gaze data patterns provide clues to what the live-action video medium specificities are, and what the possible relations are between material conditions, attention and learning outcomes. However, with respect to its assumed efficacy and reach as a research tool, it is important to point out that I do not consider this a primary method, but a complimentary one. Eye-tracking may not replace other methods that are better at providing a more complete window into how attention plays out over time, and how seeing is affected by mental processes. Visual focus is one matter, cognitive focus and/or covert attention is a whole other matter.

Moreover, as several prominent eye-tracking scholars have noted, including Holmqvist (2011), eye-tracking is ill-suited for more explorative and/or abductive research approaches. For example, as far as this thesis is concerned, it is possible that a more concise and consistent research hypothesis with respect to the impact of online time (what I call GTS) would have resulted in a less cumbersome and longitudinal research project. In other words, eye-tracking is well-suited for investigations that are hypothesis-driven and well-anchored in previous theoretical contributions, just as it was in the study that assesses audience members’ cognitive load (Paper G).

3.4.4 YouTube Analytics

YouTube analytics is used in this thesis since it generates other types of engagement metrics than, for example, interviews, recorded video sessions and eye-tracking. Here, it is deemed a highly relevant
method considering its bearing on the digital, internet-based, learning implications. This is along the same lines as how Skågeby (2015) approaches user-discussions in his assessment of Internet-based sociotechnical cultures and our everyday technologies. This method is also reflective of the indisputable fact that the video in question (Paper F) is used and interacted with via YouTube (this is also partly true for the “Miral video sequence” discussed in paper A, see next paragraph). Primarily, YouTube analytics contain numerical data, but its graphs are non-numerical and mainly visually represented. Most YouTube analytics-data may be regarded hidden data, since it can only be accessed by a producer of content, i.e. when a video is uploaded and made accessible for audiences via a personal YouTube account (“Min kanal”, in Swedish). It provides data, both numerical and non-numerical, that concerns demographics, screen preferences, user-feedback and, most importantly, retention rates. Retention rates indicate for how long viewers actually watch a given video and when/where on the video’s timeline so called “attention” drops (see Paper F). This video timeline, and retention graph, was accessed via my own YouTube channel. In this thesis, these retention rates are analyzed in conjunction with self-reports and focus group data. The purpose of this analysis is to illuminate LAV-affordances.

**The Verisimilitude Reception Study**

In the study that compares different video sequences communicating degrees of LAV-verisimilitudes, a few relevant comments from viewers of the “Miral video sequence” were analyzed (see Paper A). Comments/user-feedback may be accessed as YouTube analytics data, or, if made public, via the standard YouTube content pages.

**The Anti-Predator Fence Construction Study**

The data from YouTube is based on 8270 views (Paper F). The YouTube analytics data categories (i.e. the measures) of interest were *number of views, average viewing time, viewers’ gender, type of screening technology used* and *average viewing time per type of screening technology*. In addition to this, user-feedback (from the video’s comments-field) and the YouTube-relative attention graph were analyzed. The relative attention graph provides information with respect to at what instance during the video’s 27 min. timeline viewers chose to stop video/close window on the viewing device and/or play video.

**YouTube-analytics - Methodological reflections**

In this thesis, the YouTube analytics data provides insights into how medium specificities constrain and/or make possible instructional communication possibilities. However, usability and a viewership’s preferences may be assessed in multiple ways. Hence, here, I consider YouTube analytics data as a complimentary method. For example, it is not certain that the number of views really reflects “popularity” and/or redeeming aspects that have to do with usability. One limitation to this method is the SEO-factors (Search Engine Optimization) that, most likely, factor into any viewership on YouTube (which is run/owned by Google). Moreover, it is not disclosed on YouTube how exactly so-called “relative attention” is computed (as opposed to YouTube’s *absolute attention graph* that is a lot more self-evident in terms of how it is computed). This further sheds light on YouTube analytics data as a relatively obtuse form of data that includes many unknowns. For researchers, such unknowns may become less problematic if more/more detailed data is purchased from YouTube that then may be cross-compared. However, the limited scope (and budget) of the Anti-Predator Fence Construction Study project did not justify such an expense. However, some scholars would argue, along the same lines as Morain and Swarts (2012), that to a certain degree, the sheer quantity of data and data
categories that may be tapped into via YouTube somewhat compensates for the non-disclosed factors that most likely factor into the viewership.

3.4.5 The Focus Group Interview

The focus group method is an expedient and effective method. It allows for probing questions, which (hopefully) will not generate mere knee-jerk responses with several interviewees in the almost same instance. It generates non-numerical data (in this case, the researcher’s notes). The focus group method was deemed a valid and constructive method, since it allowed for the stakeholders to influence one another and interact with one another during the interview (Paper F). By doing so, this interview-form also touches on shared virtues and perceptions. This is key, since the object studied (the video of Paper F) is meant to function in collective scenarios with several stakeholders working together towards a common goal. This method is thus more responsive to the dynamics of an assumed real-life scenario and the idea that certain medial objects are actively shared than one on-one interviews. In this sense, it is also responsive to the basic notion that usability hinge on stakeholders/users shared values, concerns and needs.

The Anti-Predator Fence Construction Study

The focus group interview method was employed with regards to the Anti-Predator Fence Construction Video Project (Paper F). 5 stakeholders who were considered users of the video in question commented upon the video in a semi-structured fashion (i.e. some key questions were formulated beforehand). The interview was in the form of a probing, open-ended, group interview, lasting for 1 hour 45 mins. All of the participants/stakeholders are keepers of livestock in a rural part of Sweden (in an area with wolves and bears), and with prior experience of electrical fence construction. Three individuals are male, and two are female, with an average age of 48 years. The focus-group session was held in a screening venue (at Mediehuset) at Dalarna University with a relatively big screen. During the interview the complete video (Video Prototype 1) was screened once, and some sequences of interest were rescreened and discussed at the stakeholders’ requests. During the interview/screening I took notes. Three primary questions that revolved around the video’s cognitive design were addressed:

1. Which sequences are perceived as presenting facts and actions in an accurate way (if any), which ones are not (if any)?

2. Is the structure of the video understandable, do the events/actions portrayed appear complete, can they be applied?

3. Is the content pertinent; does it have an instructional purpose, if not, why?

The answers of the participants/stakeholders were subsequently summarized and extrapolated into a checklist of the video’s 6 sequences to see which actions the participants had (fully) grasped, in order to identify possible ways to improve the video. No parts of the interview were statistically analyzed.

The Focus Group Interview: Methodological Reflections

The focus group interview method is a good way for researchers to capture potential disagreements and to exchange critical viewpoints. However, since the focus group interview (quite obviously) takes place in a group setting, there is a risk that some participants will get considerably less talking time than those who tend to be more dominant. The researcher might mitigate this potential shortcoming, but when intervening, he/she may appear as having a bias. In brief, then, I think this method is valid when a particular analytical point and/or hypothesis is to be validated, but less purposeful for more
general, probing assessments that may result in being too unstructured and imbalanced. However, a good moderator (i.e. the researcher) may mitigate disproportional contributions/speaking time. In other words, the role of the moderator is key.
4. Results and Analyses

The aim of this thesis is to illuminate factors that either augment or obstruct efficient communication of instructions via the medial means of live-action videos. In order to do this, this thesis focuses on the interpretations, assessments, perceptions, preferences and sensational experiences of live-action video users. Essentially, this approach means that this thesis is a qualitative research effort. Consequently, this results chapter does not go to great lengths to differentiate between this thesis’ results per se and the analysis thereof. Instead, here, I will contextualize the relevant findings by putting them in relation to the research questions and relevant theoretical frameworks. In other words, in this chapter there is an emphasis on understanding this thesis’ results, not merely to delineate and categorize them. Thus, in what follows, results and analysis are addressed in combination.

4.1 Videography-Generated Medium Specificities

This section views digital recording technologies as something that should be considered from the perspective of their effects. For instance, here the cameras’ capacity to render conventions and traditions perceivable is of interest, thus highlighting the semiotic potential of the designerly activity of digital videography. This can also be addressed as a matter of digital aesthetic qualities and how different material formations in audiovisual ontologies are appreciated and experienced by audiences.

Luminance, Color and Synchronous Audio

One of the rather unsurprising findings of this thesis indicates that the values and color of brightness increase the visibility of LAV’s space configurations and other video contents. Naturally, without luminance and a color spectrum, LAV-communication would not be possible (see Papers A and B). However, these image aspects are not uniquely tied to the activity of videography, but are integral to visual media in general. Synchronous audio, on the other hand, is a medium specificity of (most) Live-action Videos. This material aspect is expected to match movement and other visual indexical content of the video image. Thereby, it enhances the powers of the affordance of comparability (see Paper F).

Another medium specificity that truly makes videography-generated imagery unique is the kind of digital code that enables digital videography, and, more generally, makes mediated transient communication, with all of its time and velocity implications, technologically feasible.

Digital Code

Videography-related digital file formats and codec encode and transmit information, in particular ways (see Paper B). These are technical characteristics of the LAV-medium and facilitate the critical depiction of, among many other visible phenomenon, movement. In doing so, file formats and codec embody and signify the relationship between cameras, their image sensors and other aspects that are mainly metaphysical (for example, algorithms). One visible output of this complex relationship is digital noise.
Digital Noise

The live-action video medium’s materiality has a metaphysical, but sometimes perceivable, component. This is what is mainly discussed in this thesis in terms of digital noise. Such noise is inseparable from conventions and traditions as it converges with, and reinforces, other image aspects. For example, digital noise reinforces the viewers’ perception of flatness and depth in video imagery (see Paper B). Depth, in turn, signifies certain stylistic paradigms. In this way, the LAV-medium functions as a conduit for traditions of practice, as well as for metaphysical components that become perceivable when they converge with other ontological dimensions of the live-action image. One such tradition of practice is the cinematic tradition of displaying shallow, or great depth, of field.

The High and Low Production Aesthetic Paradigms

Shallow/great depth of field and digital noise interact with other image aspects that are camera-specific. These other image aspects are, primarily, framing, perspective (such as, Point of View), and how the image frame moves, or does not move (i.e. moving camera, or static camera). These aspects are what together essentially communicate the LAV-imagery’s space configurations, or what might be called compositions, and make them verifiable and believable (see Paper F). This might be discussed in terms of a kind of realism or indexical linkages, or degrees of designerly and artistic control, in contrast to improvisation. A lot of control, both technological and artistic, equals high production value, and vice versa. However, this sought-for control is sometimes mitigated when a videographer is unable to foresee the experiential and perceptual reactions from audience members, due to videography-related visual formations that audiences interpret as an intentional style, for example, the High Production Value style (see Paper A).

Materiality is inseparable from traditions of practice (see Papers A and B). This means that the expressive qualities of the digital adhere to conventions, for example, the conventional look of fiction films and that of high-budget films professionally made. This is what I discuss in terms of High Production Value. This might be considered an indicator of levels of cinematic hedonic embellishments that contribute to a film or video’s appeal and acceptance. However, it seems that the two production value conventions (both low and high) have become unstable. At least, this appears to be true with regards to the High Production Value paradigm. It may have begun to play out its dominant role in live-action filmmaking (see Paper B). In comparison, in LAV-production efforts of a non-fiction kind, the established conventions appear to stay more stable (see Paper A). On a more general note, this is an indication that not all LAV-genres function and realize their semiotic potential in the same way, in spite of all being videography-driven.

Perceiving the Metaphysical

As briefly touched upon in the previous paragraphs, one primary, but basic, finding of this thesis indicates that viewers of LAV-content react to material properties of information (see Papers A and B). On one hand, these material dimensions may be simply described as visible elements that make up a complete whole. When addressing this whole - the complete, multimodal accumulation of cinematic, expressional elements conformed into a coherent audiovisual narrative – the affective responses of most viewers would be considered quite ordinary and expected. On the other hand, if some critical material dimensions are analyzed in isolation and subsequently put in relation to certain media technologies, we discover that some visual phenomenon are appreciated and reacted to in rather unpredictable ways that may, or may not, reflect directorial input.
Digital LAV’s-materiality has a metaphysical component that is part of its ontology. This emphasizes the perceivable workings of a computer, or, more precisely, the perceivable workings of the sensor and its interaction with the camera lens, its position, digital codec and file format. This also places value on the idea that the digital can only be studied in composition with other materials, such as screens or camera lenses. It is on the screen that the metaphysical ceases to be purely metaphysical and, for example, becomes perceivable digital noise that sometimes “tickles” the viewer’s mind to a positive effect.

When material aspects and ontological features converge they sometimes dispel the illusion of what reality should look like. For example, noise might make background details emerge into the foreground, and then, recede again - a phenomenon I discuss in terms of images being syncretistic (Paper B). However, this matter of fluctuation most likely happens without audiences knowing or understanding that this is happening. This is perceived by some, without their being aware of it, and may be best described as a kind of cognitive stimulation. This stimulation is one explanation for why iPhones (at least, the older models) might be perfectly fit for generating certain kinds of imagery, even imagery that is expected to look high production value (see Paper B). Perhaps visual syncretism of this kind compensates for other less redeeming qualities in iPhone live-action video making.

4.2 Leveraging LAV-instructions’ Efficacies

This thesis revolves around the evaluations, perceptions, expectations and experiences of users. The following paragraphs will illuminate these user-related dimensions and explore the material underpinnings and interconnections of these aspects in order to identify the factors that make efficient instructional LAV-communication possible.

The Realism Factor

One key finding of this thesis indicates that the medium specificities that facilitate the realistic and authentic depiction of humans doing things with their bodies make instructional LAVs useful. These specificities might be discussed in terms of specific ontological image and audio elements. Elements such as these augment the affordances of immediacy and comparability (See Paper F). This is another way of saying that the typical LAV-ontology makes possible its desired, instructional, expressive capacity by exploiting the procedural motor advantage. Without the realistic depiction of, for example, hands moving, this advantage cannot be exploited. Another aspect of this is that LAVs can do so, in part, since they inherently embody information that is self-explanatory and that can be grasped by means of immersion only.

The Biological Information Implications

Instructional LAVs that explicitly show procedures that involve humans and human movement are effective support tools (if they are high in visibility and audible, that is). Static counterparts may be used, but are no perfect substitute if quick results are desired, albeit the structural diagram type might be an exception. This is consistent with previous empirical evidence with regards to motor skills communication efficacies and the display of hand-movements in live-action videos. This procedural motor advantage is in large part due to a video designer’s employment of two instructional strategies that may run in tandem and render instructional communication more efficient. One such instructional strategy explicitly communicates secondary biological information, while the other lets users experience biological primary information by immersion only, without explicit guidance (see Paper F). Purposeful and efficient instructional strategies involve designs in which biologically primary and
secondary information are not treated the same way, and/or are confused with one another. This highlights a kind of balancing act that a designer of LAV-instructions must actively consider. This implies distinguishing between biologically primary and secondary information and how these two information-type paradigms can be consolidated and combined to make purposeful, efficient, instructional LAVs.

The Eye-Movement Behavior Factor

Transience, and/or recorded movement are kinds of general medium specific attributes that are attributable to videography. This highlights the finding of this thesis that indicates that live-action video medium specificities make possible their desired, instructional, expressive capacity in situations in which users have a relatively focused decoding style (see Papers D and E). Thus, the key insight is that displayed movement and transient information require relatively focused decoding styles in order for the displayed instructions to be adequately understood. However, it seems that the display of movement, and the way it functions as an attention grabber is enough to leverage purposeful focused viewing styles. It is a possibility that the video stimuli, in itself, trigger relatively high GTS-scores among the video users, since the video users have higher GTS-scores, on average, in comparison to the static media stimuli groups (see Paper E). In part, this exemplifies the LAV-format’s unique status in comparison with other kinds of visual instructions.

In other words, learners with high Gaze Time on Screen-scores (GTS), on average, perform slightly better than those with lower GTS-scores. There is a low (but significant) correlation between assembly-build error and on-target eye-movement behavior (see Paper E). However, to what degree added levels of focused eye-movement behavior on the part of viewers would really warrant superior learning outcomes, remains to be established. It seems that eye-movement behavior that includes a little offline behavior (around 10% of total viewing time) is not particularly detrimental to comprehension, regardless of type of stimuli. Consequently, it appears that it is probably impossible for learners to compensate for otherwise poor instructional design by staying more focused during study time. Indirectly, this finding calls into question the assumed positive impact of learners’ visual attention on comprehension, and prompts future researchers to further investigate to what degree visual attention is necessary for learning from visual instructions.

The Cognitive Load Factor

Recorded video content that moves, i.e. transient content, makes video editing possible. Recorded movement contains editing points that may, or not, be exploited. Editing, i.e. designerly selections, in turn, makes instructions more concise, less meandering and thus more useful for users. However, LAV-content can only be successfully edited if it is high in formability or malleability. Digital malleability affords continuity editing that permits users’ non-interrupted and flowing bottom-up perceptual processes (see Paper G). Such perceptual processes do not obstruct top-down processes, whereby a user may, to a fuller extent, tap into his/her cognitive and perceptual powers when acting upon instructions. In other words, continuity editing warrants low levels of cognitive load on the part of the learner.

The Perceptual Affordances

Another primary finding of this thesis indicates that materiality may be understood as representing functional affordances in certain learning situations. These are referred to as perceptual affordances. I consider these actual, not only implied or imagined. However, their powers depend on how an agent
perceives them, to what degree they are “engaged”, and how users apply them to facilitate action (see Paper F). This basic understanding provides a good starting point for a further investigation into how live-action imagery leverages instructional practices and why it might do so.

The videography-related affordances of immediacy, comparability, verifiability, recordability and visibility are considered key perceptual affordances in instructional scenarios that employ the LAV-format as a visual support system (see Paper F). These affordances, together, make possible immediate, reliable and verifiable comparisons. For users, the comparison between the video’s content and what is in front of them in a learning situation is primary (see Paper C). Also intra-comparisons are important. For example, the comparison between how something sounds and how it looks should correspond. However, such comparisons are only made possible if the medium’s visibility is high and the audio is clearly audible. Without high visibility, no other affordances may be exploited (or analyzed). This probably explains why many LAV-users prefer large screens (larger than smartphone screens) when interacting with LAV-content (see Paper F). Larger screens afford greater visibility, and greater visibility makes the discerning of a great amount of detail possible.

4.3 Obstructing LAV-instructions’ Efficacies

Although transience and recorded movement are LAV-specific attributes that make instructional LAVs useful, paradoxically, they may also hinder their usefulness. These attributes may cause confusion, distraction and intra-communication barriers.

The Interactivity Problem

One finding of this thesis indicates that the LAV-format influences the way participants refer to and conceptually frame the object’s (a Lego truck) parts in front of them while assembling it. More specifically, the live-action video instructions influenced the support of dialogue between assembly team-members negatively (Paper C). What we might call collaborative interactivity potential, for lack of a better term, is thus definitely not something that may be ascribed to the use of instructional, procedural LAVs. It seems that this problem has to do with learners not managing to remember all the steps in a procedure. Therefore, they feel urged to play and watch the video on many different instances while assembling. They also stop and play the video over and over again. Thus, LAV instructions do not seem to function optimally in collaborative, instructional, so-called “design-challenge” settings. Possibly, interactivity features could alleviate some of its instructive shortcomings in such settings, since frequent play/stop actions appear to be a success factor.

From the above, I extrapolate that transience presents an obstacle to learning from instructional live-action videos. On the whole, this supports the findings of other research efforts into the problem of transient information in instructional videos, as well as findings regarding the relevance of interactivity features in videos, so-called macro- and micro-scaffolding techniques, as well as segmenting techniques. Such techniques may, in effect, lessen the transience effect.

Thus, live-action video medium specificities constrain or resist the video’s desired, instructional expressive capacity by lacking in interactivity. This occurs both in the technological sense (LAVs have no built-in buttons or swipe functions, and so on), and in the sense that they negatively condition potentially constructive discussion in design teams. Hence, live-action video instructions are not a substitute for face-to-face communication. The video medium represents a detailed ontology, and as such it is indeed “rich”, but it may not match the degree of communication richness required by the task.
The Audio Edit Effect

In some circumstances, the medium specificity of synchronous audio appears to obstruct the usefulness of LAV-instructions by indirectly lessening the communication potential of the procedural motor advantage (see Paper F). Synchronous audios may do so, indirectly, since audial communication means complement and summarize complex movement, and thereby invite the editor to edit in accordance with what is communicated audially. Such edits contain much talk and little uncommented movement. The resulting videos are deceptively concise and super-packed with audial information. However, the audio, in spite of the fact that speech, in itself, may be regarded as biological primary information, cannot compensate for the lack of human movement and gestures. Especially speech that by means of editing has become extremely “densified”, and thereby unnecessarily complex, cannot compensate for the showing of unambiguous and self-explanatory human movement. Thus, speech may be regarded as both primary and secondary information. However, more often than not, in how-to videos speech should be regarded in the latter information category.

The Bad Weather Effect

Another finding of this thesis indicates that the affordance of recordability sometimes hinders users from successfully acting upon instructions in live-action videos by its propensity to cause distraction, or even worse, it may cause confusion. Depending on the type of distraction, the affordance of recordability might also make users less motivated to actively engage with the LAV-format as an information source (see Paper F). Bad weather could be regarded as a kind of such distraction. This problem is due to the videographer perhaps being incapable of avoiding things and phenomenon that distract or go against the overall pedagogical goals of the instruction when recording video. Such things and phenomenon are, for example, unwarranted gestures, odd facial expressions, glitches, distracting details in the background, bad weather, and alarm sounding noises, etc. In brief, reality is difficult to control, therefore LAV-recordings are seldom picture-perfect.

The Cognitive Load Effect Revisited

As previously noted, continuity is not a mere aesthetic concern, but also relatable to the cognitive processing of the video users. The opposite of continuity - discontinuity - causes disrupted cognitive processing. One finding of this thesis indicates that altered film sequences constrain viewers and increase their cognitive load (Paper G). The designerly activity of editing exemplifies a craft-related capacity that serves to achieve effortless watching, thereby enabling the constructed film reality to mesh with the human mind. However, the problem is that the editor might not always be able to achieve this, due to poor camera handling in production.

This elucidates the symbiotic relationship between the designerly activities of editing and videography, and the role that medium plays in users’ and producers’ evaluation and appreciation of a given live-action video. This is analogous to other kinds of artistic medium that are more physical in nature, and which must afford the ability to be shaped into something else without breaking, such as, clay, wood, plaster, textile, gold, sheet-metal, and so on. In other words, many medial objects need to be malleable, including digital ones.

4.4 Implications of Findings

The findings and the analyses thereof presented in the previous sections, to a large extent, align well
with previous research. Above all, these findings concur with previous research that pertains to motor skills communication efficacies, relatable to the display of hand-movements in live-action videos (Boucheix & Forestier, 2017; Berney & Betrancourt, 2016; Boucheix & Schneider, 2009; Castro-Alonso et al., 2015; Ganier & de Vries, 2016; Höffler & Leutner, 2007; Lowe & Schnottz, 2008; Marcus et al., 2013; Tversky et al., 2002).

Moreover, the findings of this thesis are concurrent with other research evidence that indicates that transient, screen-based communication is a challenging format in many instructional scenarios (Sweller, Ayres & Kalyuga, 2011; Boucheix & Forestier, 2017; Leahy & Sweller, 2011; Paas et al., 2007; Wong et al., 2012).

However, the analysis of the findings of this thesis also introduces some understandings which may be regarded as novel insights. Considering the scope, aim and objective of the thesis, I summarize, below, two such insights that I consider key.

**The Visual Attention Factor**

One novel insight of this thesis is that the LAV-format, in spite of being transient in nature, is not extremely sensitive to unfocused decoding strategies. This places emphasis on how procedural narratives are created, and to what degree they invite and allow for unfocused viewing behavior. Creating a kind of *fabula* out of separate narrative building blocks, without being overly visually attentive, may be considered a cognitively constructive mapping activity, not only for action diagram-users (cf. Daniel & Tversky, 2012), but also for LAV-users, notwithstanding that there are certainly limits to being visually inattentive. Perhaps, this is equal to being *cognitively* focused. In any case, along the same lines as Cutting, this places value on narrative contexts in filmmaking design ventures (2005), as well as on the visual instruction users’ decoding behavior, what Lowe discusses in terms of information design users “graphicacy” abilities (2017).

Visual attention is automatically triggered by motion in video content (cf. Boucheix & Lowe, 2010; Boucheix et al., 2013; De Koning et al., 2010; Kriz & Hegarty, 2007, 2010; Ozcelik et al., 2009, 2010; Wang & Antonenko, 2017; Wang, Tsai & Tsai, 2016). However, to what degree video content that displays movement, in itself, is enough to trigger good enough focused behavior remains to be fully understood. Nevertheless, with the knowledge that looming objects in naturalistic contexts “capture” attention (Franconeri & Simmons, 2005), and with the knowledge that motion displayed by human agents causes neural responses by our socially tuned brains (Kaiser, Shiffrar & Pelphrey, 2012), it is possible that this is a more potent factor than the influence of so-called cognitive styles on visual decoding and eye-movement behavior (cf. Höffler et al., 2017).

**The Malleability Factor**

Another new insight of this thesis infers that *continuity* is a medium-specificity that is relatable to videography. With the knowledge that continuity normally is considered a perceptual sensation, created by an editor in the editing phase, the analyses featured in this results chapter clearly show that continuity editing is absolutely not a sole post-production concern. Continuity is first created when a videographer shoots LAV-content. Thus, videography is also an editing activity and the original one, since the videographer edits in his/her head while shooting. This highlights the videographer as a (hopefully) multi-competent crafts-person and designer, and not a mere delivery-person of raw and unrefined digital material for others to exploit.

*Malleability* (or *formability*) is an affordance that videographers exploit in order to establish editing points while shooting. Normally, they do this to accomplish the compression of time in the (later)
editing phase, without compromising continuity. This is important, since longer videos drain precious working memory recourses much more than shorter videos, due to the transience effect (Sweller, Ayres & Kalyuga, 2011; Boucheix & Forestier, 2017; Leahy & Sweller, 2011; Paas et al., 2007; Wong et al., 2012). This explains why most how-to videos are short in length.

However, malleability is not only an enabling affordance, but also one that is challenging to control. The illumination of the negative side of malleability helps us understand how live-action video medium specificities constrain or resist its desired, instructional expressive capacity by being extremely dependent on the videographer’s ability to exploit digital malleability. Some video material is low in malleability due to poor videography strategies and bad camera handling. Such video material tends not to embody many obvious or natural editing points. Instead, such video content is, typically, either dead static without any alterations to the composition, or, worse, constantly “on the move”, zooming in and out.

In summary, therefore, this is to suggest that the videographer’s capacity to manage and control digital malleability is a key concern in live-action video making. However, this capacity should not be taken for granted.
5. Discussion and Conclusion

The Discussion and Conclusion section of this thesis will draw on the presented results in the previous chapter thus formulating a sound and relevant basis for certain key information design prescriptions for live-action videographers to convey instructions. In drawing up this formulation, I will also highlight this thesis’ unique contribution, both with regards to industry and academia. At the end of this chapter, I will discuss the limitations the thesis’ implications for future studies. But first, I will summarize the results of the thesis and link them to the research questions to concisely highlight the critical themes that subsequently will be discussed in this chapter.

5.1 The Results - An Overview

The LAV-format and instructions presented in this study present certain specific, designerly challenges. These challenges may be linked, on one hand, to this format’s particular way of recording and displaying the stuff in front of the lens. This might be considered a technological dimension in which matters of “the digital” are key (cf. Gross et al., 2013). On the other hand, these challenges are relatable to truly human-centered factors, such as the videographer’s directorial capacity to control this kind of technological mediation in order to support the perception (including auditory perception) and cognition of the user. Addressed as a problem that concerns information design, this is the problem of enabling “the user to find a way through the maze of information” (Bonsiepe, 1999, p. 59). All of these aforementioned and interconnected issues may be regarded as first-order understanding. The videographer must attend to these matters, well aware of the fact that users are also cultural beings, and not solely biological processing machines with certain socio-culturally inclined preferences. These aspects, on the other hand, may be regarded as second-order understanding. See Figure 3 for a stylized depiction of this thesis’ main results expressed as themes grouped around a specific research question that address both first- and second-order understandings.
As the above mapping of this thesis’ results implies, the aim of the study is, in a small way, to remedy an unsatisfactory situation in academia, in essence, a kind of knowledge gap that contributes to a general confusion regarding how and why live-action videos need to be designerly approached in certain ways in the *production stratum* (Kress & van Leeuwen, 2001).

In brief, then, viewed as design nexus, live-action videography is associated with medium specificities that need to be managed to achieve instructional efficacy and communication congruence. In other words, simply expressed, a videographer may be unable to ensure that the key components of an instructional message can support each other by just functioning as a person who merely presses the record button at certain instances, while leaving everything else to chance. In the realm of instructional live-action videography, there are many medium-specific aspects and factors that either obstruct or augment instructional communication that need to be attended to, in one way or another, by the videographer.

In my Licentiate Thesis (2013), I set forth a research agenda to disentangle the factors that make possible or obstruct video-mediated communication of instructions. This disentanglement is grounded in everyday reality, making central the cognition of the user and the material qualities of the image. In this final section of this thesis, I can assert that this is also what I have aimed to do in this study, and I conclude that this research approach has generated four basic insights, as follows: firstly, management of the complexity of instructional efforts, secondly, instructional communication strategies regarding the type of information, thirdly, users’ decoding/screen-interaction behavior and, lastly, the LAV-affordances.

*The Management of Complexity Through HCD*

This thesis shows that value deficiencies drive transformation and that complex support efforts need to consult users in order to identify possible validity threats to the instructional design, i.e. factors that diminish the affordances of the LAV formats (cf. Rouse, 2007; Krippendorff, 2006). Design-process management frameworks, in combination with HCD-inspired approaches, offer constructive ways of doing this. In brief, in the instructional LAV-context, what need to be identified with the help of stakeholders are procedures and actions that require different levels of explicit instructions, or none at all, and possible designerly improvement schemes. If this is done in a concerted and structured way.
fashion, users’ confusion, irritation and disengagement can be avoided. The disadvantage of this approach is that it may render impossible improvised, “run and gun”, quick turnaround, video production.

*Show, Do Not Tell*

The second insight of this thesis is that users’ cognitive processing and visual decoding depend on the power of cinema to show actual human behavior and action. If only telling by means of text, added graphics or audio narration would suffice, this power would go unused. A purposeful and useful live-action video instruction should, therefore, include humans and human movement, focusing on doing, rather than explaining. This is to acknowledge that good, live-action instructional videos use biological primary information as a means of freeing up cognitive recourses for the demanding parts of any instruction, which exist in the realm of biological secondary information (Sweller, Ayres & Kalyuga, 2011). In other words, bottom-up perceptual processes and biologically primary information provide not only the foundation upon which medial affordances can be assessed but, in addition, they make possible sound socio-culturally informed instructional practices.

*The Focused Behavior Factor*

The third basic insight of this thesis is that focused decoding styles have little to do with LAV-instructions’ efficacies. Some unfocused behavior is not detrimental to LAV-communication. Nevertheless, on the whole, designers of LAV-instructions are better off if they employ designerly techniques that help users to focus on relevant information (in the AoIs). Segmenting is one way of indirectly achieving this. Segmenting allows for closure (i.e. cognitive focus), and the onset of motion that normally follows the actual segmenting-point greatly activates attention (Franconeri & Simmons, 2005) This is consistent with previous research findings that pertain to enhancement methods in multimedia lessons (cf. Sung & Mayer, 2013). In other words, transience is probably the main concern here (and the limitations of working memory recourses), not focused behavior, since videos that contain ample human movement, probably, in themselves, trigger good enough visual focus. In concurrence with Clark and Mayer (2016), this particular finding implies that psychological engagement is more important than physiological engagement. However, these two kinds of engagement styles interact, and interrupted physiological engagement has the capacity to obstruct psychological engagement and thereby indirectly impede learning (cf. Clark and Mayer, 2016, pp. 219-236). This is along the same lines as Swenberg (2017), who asserts that the acknowledgement of the interplay between inputs that compete for the privilege to attention is key in film production design ventures.

One designerly implication of this is that videography per se, as overall design strategy in instructional live-action videomaking endeavors, is not a sufficient design strategy in all conditions. Designerly alterations and features, such as scaffolding techniques, or segmenting via software and computers may leverage LAV- efficacies (cf. Cojean & Jamet, 2017). This is also to recognize that the narrative functionality of instructional LAVs can, and should, be adjusted. This could also be achieved by means of slowing down certain segments. However, regarding slow-motion techniques, there appears to be a tradeoff, since this would possibly undermine the affordance of reliability. Perhaps mixing “crisp” slow-motion sequences (i.e. with no motion blur) with real-time sequences belonging to the same original content may be a possible solution. In that case, the user would have it, so to speak, both ways. Such videos would then, in theory at least, accommodate diverse cognitive styles and decoding preferences, both focused and unfocused (cf. Höfler et al., 2017;Holsanova, 2001). Furthermore, the aforementioned kind of mixing would afford the user the possibility of more frequently being able to revisit the depicted objects’ AoIs. This would require shooting video at high-frame rates, in addition to high resolution.
In any case, the important point, here, is that designers of instructional LAVs should aspire to identify the base level at which the LAV format works for most users, diverse decoding styles notwithstanding. In this respect, medium specific features and affordances are key. Consequently, for example, narrative functionality should not be adjusted at the expense of reliability (cf. Cutting, 2005). Designerly techniques must harmonize with the LAV-affordances that afford instructional efficacy, otherwise, there is distinct possibility that users with both focused and overview decoding styles will miss out.

The Interaction Factor

As suggested above, this thesis’ primarily links the efficacy aspects of LAV-instructions with users’ behavioral engagement. In this study, I address two kinds of behavioral engagement regarding instructional scenarios featuring screens. The first kind of engagement concerns eye-movement behavior (see the previous paragraph), while the second engagement concerns human, interactive behavior in a group setting. The open-ended, probing discussions of the user that serve to delineate possible solutions in a design process and/or assembly process appear unsupported by the instructional LAV-format. The designerly implication of this is to employ LAVs only in the correct instructional circumstances, or to use another type of visual instruction, such as diagrams (Tversky, Morrison & Betrancourt, 2002). This insight is not new, and is consistent with earlier research findings in the field of educational psychology. This basic finding is also reflective of the way I have used affordances, which, in itself, is a theoretical construct that connotes physical activity (Gibson, 1979), as opposed to psychological activity (Norman, 1988). However, what is new, here, is that this thesis adds to the understanding of how this engagement plays out in instructional situations, and, in turn, how the engagement is conditioned by the ontology of LAV-instructions.

The Perceptual Affordance Factors

In order to explain how the LAV-instructional format can be managed and controlled for its betterment, we need to further highlight its key affordances and identify the factors that lessen and/or strengthen their powers. As this thesis demonstrates, this is the logical consequence of, to paraphrase visualization scholar Colin Ware, the acknowledgement that Gibson’s affordance construct is extremely useful from a design perspective (2013, pp. 17-20).

Visibility

In live-action videography, visibility can be optimized by using high-resolution recording gear. Visibility provides the user with critical information, without which content is nothing, since this affordance impacts the potential of all other affordances. Great visibility allows for the focused viewing behavior of the user, and, most likely, renders familiarization/overview gaze patterns less necessary and prominent. Video images’ coarseness, together with perceived dimensional flatness, compromise visibility, but might also cause unexpected cognitive responses, for better or for worse, it seems. Another argument for not opting for the more crude recording gear is the prospect of negating “halo-effects”, i.e. high production value. However, high usability might also be “beautiful” since what is useful can also be attractive (Tuch et al., 2012). With respect to creating a high-production value look, there is one possible caveat though, the prospect of changing convention adherence. To what degree this is a relevant problem in live-action instructional video making remains to be investigated, but a sensible approach would be not to emphasize the depth cue of accommodation too much (i.e. extremely shallow depth of field), and instead try to engage other depth cues in combination with other more conventional de-cluttering techniques on set. In brief, then, the problem here is that
while affordances and expressive potentials might be quite readily circumscribed from a theoretical standpoint, their cognitive ramifications and socio-cultural implications are not. Thus, there are other causes than purely purposive causes in the realm of digital, live-action video making.

Recordability

Recordability also highlights technological aspects of live-action videography, and further underscores the need to actively manage aspects of LAV-production. Perhaps, needless to point out, this affordance is of utmost relevance in instructional live-action ventures, since it affords the reproduction of unique, authentic instructional situations that, for one reason or another, need to be communicated to a greater audience, via the relatively simple means of directorial input, a camera, uploading capacities and screens (cf. Morain & Swarts, 2012). The disadvantage of this affordance is that if the specificities of the recording situation cannot be controlled, the captured material will most likely cease to function as an instruction, and become at worst, a parody, or, what is more often the case, a bad instruction that distracts, or fails to motivate. Recordability also affords the depiction of time, i.e. how time makes possible the communication of movement and certain LAV-specific space configurations (certain space configurations can only appear over time).

Within the context of visual instructions, the depiction of time plays a significant role, since it is this dimension that sets the motion picture formats apart from static instructional formats, and the reason why it should be used as an instructional means in the first place (cf. Ware, 2013, pp. 337-343). However, utmost care must be taken not to compromise the affordance of reliability when tinkering with the video formats velocity dimensions. Hence, a piece of valid designerly advice for video makers is to handle manipulations of time with great care, and to be wary that otherwise sound multimedia design guidelines that pertain to segmenting and the breaking down of lessons do not, in themselves, indicate anything about how to actually achieve this in motion pictures. Perceptual precision, and methods that pertain to how to achieve continuity, on the other hand, are key in this regard (Swenberg, 2017). This may come as no surprise, since it is no secret that “Hollywood style” editing and cinematography techniques mesh perfectly with the human visual system, as long as the narrative subordinates the de facto manipulations (Cutting, 2005).

Reliability and Immediacy

Reliability highlights the most sought-for quality in instructional LAV-content - its realism, verisimilitude, evidential and indexical dimensions. According to Nichols, this is what “bears witness to the way the world is” (2010, p. XV). This is a highly regarded quality of LAV-instructions, since reliability affords certain real consequences that other, more contrived and constructed, formats do not afford so readily. This seems especially true in instructional situations showing human movement. This is also to suggest that reliability is a prerequisite for the affordance of immediacy, i.e. the ability of users to recognize video content immediately without having to interpret it, tapping into humans’ ability for direct perception, as well as, in turn, humans’ ability to learn from biological, primary information without tuition. Immediacy and reliability are thus two sides of the same coin.

One might say that the affordance of immediacy is the urgent aspect of usability (cf. Rouse, 2007), but here we are talking about a time frame of about 300 milliseconds. Reliability is also relatable to users’ activation of mirror neurons. This is important in instructional applications, since immersive communication strategies are augmented by the activation of mirror neurons, at least in instructional LAVs that display humans doing human things. This is, as it were, comparability at work in the mind, highlighting the fact that action execution and observation are intimately related processes. One practical implication of this is that LAV-instructions should facilitate users’ (almost) simultaneous action and watching in order for users to fully exploit the mirror neuron activation advantage. In
summary, therefore, the basic implication here is that if users cannot imitate a displayed procedure in a video, the video has low reliability.

Comparability

The affordance of comparability elucidates key aspects of live-action instructional content, since it affords straightforward comparisons between the natural world and the world reproduced (Paas & Sweller, 2012). Here, again, mirror-neurons play a certain role. At a more general level, this has profound implications for learning, since compare and contrast is our foremost natural form of thought (Silver, Strong & Perini, 2007), and there is ample compelling evidence that compares and contrasts learning strategies leading to eye-opening gains in student achievement (Marzano, 2007). Comparability affords multimodal presentation techniques, whereby users’ comparisons between modalities enhance its verisimilitude. The instantaneous comparison between audio and its visual referent is key. Hence, one good piece of advice for videographers is not to disregard the expressive capacity of audio, and to safeguard that audio levels/quality is at least on a par, or better, with the visuals, since bad audio spoils the visuals. Unwarranted, sharp sound transients alarm and distract us, and may compromise comparability (Eidsvik, 2005). It is also likely that cognitive load due to the transience effect lessens the effect of comparability.

Malleability

One of the hallmarks of LAVs is that they represent a low degree of malleability. However, the malleability that exists affords LAV-content to be transformed into a comprehensive and coherent audiovisual, procedural instruction which can be edited. It also allows for the video content to be adapted for certain uses, for example, to be more functional for users with overview and/or low gaze time on screen (GTS) decoding styles. Hence, this affordance should be the concern of the designer/videographer, since it is probably easier to accommodate diverse viewing/decoding styles than force users to adapt their behavior - although this remains to be tested. Above all, without the affordance of malleability, the designerly activity of editing is not feasible, fix-it-in-post activities notwithstanding.

This leads to a kind of a designerly balancing act. On one hand, the videographer must honor the completeness of procedures and actions, and their built-in narratives, while on the other hand, most real-world procedures that need to be communicated are not picture-perfect, and too lengthy to fit technology mediated communication efforts, thus needing to be edited. Furthermore, shortened video durations lessen the prospect of the learner feeling bored and becoming inattentive, which is, perhaps, the biggest problem with real-life procedural learning efforts in general. Therefore, the videographer must ensure that the recorded content is editable, that it contains natural editing points, concise beginnings and ends. An editor cannot make perfect already imperfect LAV-content. This is also to point out that malleability affords continuity, or more concretely, continuity editing (cf. Swenberg, 2017). This is critical in live-action video instructional practices, since discontinuity causes cognitive load to increase, and thereby can be assumed to leave little cognitive capacity for the processing of the important stuff - the instructional message. Users’ increased cognitive load obstructs purposeful instructional communication. Thus, malleability affords live-action instructional video communication efficacy.

5.2 Instructional Live-Action Video Design Prescriptions

In concurrence with information design scholar, Andrew Dillon, this thesis recognizes that “reading”
information artifacts is both a physical and cognitive process (2017). However, as this study demonstrates, if the physical aspects of such a reading are neglected, users’ cognition will become constrained (cf. Sweller, Ayres & Kalyuga, 2011). Therefore, any LAV-guidelines that may assist designer thinking must begin to address certain basic first-order understandings that center on physical engagement (Clark & Mayer, 206, p. 219-233).

In this paragraph, I will present a summary of the principles and design prescriptions discussed in the preceding paragraphs centered on perceptual LAV-affordances. First, I will present these principles and prescriptions in the form of a list with 18 recommendations for videographers and producers of LAV-instructions. Then, in order to facilitate a quick and uncomplicated overview (that may be more user-friendly than a list), inspired by instructional technology scholar, Hans Van der Meij’s “minimalist design principles and heuristics” (2007, p. 295), I present a table consisting of two columns, a principle column and a design prescriptions column. See Table 1.

1. Use a minimum of 2K resolution recording gear. Avoid highly compressed video formats.


3. Acknowledge best-practice video production recommendations (such as correct exposure, audible audio, lighting, sharp focus, avoid too many wide shots etc.).

4. Do not over-emphasize cinema stylistic techniques, such as extreme shallow depth of field.

5. Encourage users to use relatively big screens. Opt for many close-ups if smartphones are expected to be used as screening device.

6. Record unambiguous representations of procedures without information that distracts. Re-record if content is not perfect.

7. If the illusion of a 3D space is key, which it often is in live-action videos, consider camera moves that emphasize this illusion (i.e. employ depth cues that depend on the perspective of the camera changing over time).

8. Some video file formats/video codec generate noise in certain conditions. Consider how this will affect the viewer, as it will interact with other image factors.

9. Think of recorded human movement and simple procedures as the base layer that may be grasped via immersion only. Everything else requires explicit instructional designs (highlights, signals, explanatory audio etc.).

10. Show humans and human movement/actions.

11. Talking and human speech cannot compensate for lack of unambiguous visuals with human action and movement.

12. Human movement and actions involved in showing simple procedures may be comprehended by all humans. All other additions to the video content make the video suitable for a certain target audience.

13. Ensure audio is synchronous, clearly audible and matches visible content.

14. Encourage users to act upon the instruction while it is showing.
15. Instructions should be brief, concise and properly paced.

16. Employ skilled videographers, since they are more likely to make video content editable by actively creating editing points during the recording phase. This warrants continuity.

17. If you are the videographer, segment video while shooting. Vary compositions and/or static and tracking shots. Record complete actions. Let objects enter and exit video frame.

18. Edit video content (but it is normally futile to do so if prescriptions #16 and 17 above are not adhered to). A skilled editor is more likely to enhance continuity.

Table 1. Principles and Design Prescriptions for instructional LAV-videographers and designers of LAV-instructions.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Design Prescriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increase Visibility</strong></td>
<td>1. Use high resolution recording gear</td>
</tr>
<tr>
<td></td>
<td>2. De-clutter on set before recording</td>
</tr>
<tr>
<td></td>
<td>3. Acknowledge best-practice video production recommendations</td>
</tr>
<tr>
<td></td>
<td>4. Do not over-emphasize cinema stylistic techniques</td>
</tr>
<tr>
<td></td>
<td>5. Encourage users to use relatively big screens</td>
</tr>
<tr>
<td><strong>Exploit/Control Recordability</strong></td>
<td>6. Re-record video content if it contains distractions</td>
</tr>
<tr>
<td></td>
<td>7. Move the camera to emphasize depth, but do not shake it</td>
</tr>
<tr>
<td></td>
<td>8. Avoid highly compressed video formats</td>
</tr>
<tr>
<td></td>
<td>9. Use recorded video as base layer of instructional design</td>
</tr>
<tr>
<td><strong>Exploit Reliability and Immediacy</strong></td>
<td>10. Show humans and human action</td>
</tr>
<tr>
<td></td>
<td>11. Avoid too much talking in recorded video</td>
</tr>
<tr>
<td></td>
<td>12. Use signals and other additions to recorded video as content that make video specific for a certain audience</td>
</tr>
<tr>
<td><strong>Enhance Comparability</strong></td>
<td>13. Ensure audio is audible and synchronous</td>
</tr>
<tr>
<td></td>
<td>14. Encourage users to simultaneously watch and act upon video instructions</td>
</tr>
<tr>
<td></td>
<td>15. Keep instructions brief but properly paced</td>
</tr>
<tr>
<td><strong>Exploit Malleability</strong></td>
<td>16. Employ skilled videographers</td>
</tr>
<tr>
<td></td>
<td>17. Segment video and create editing points during recording</td>
</tr>
</tbody>
</table>
5.3 Concluding Remarks

The key question raised in this thesis is not what a technological medium is, but rather what it does to the user. Some of these effects are obvious and appear purposeful. This reminds us of the classic Marxist view on technology, which emphasizes machines’ social transformative powers as designed tools of the ruling class for economic benefit (of the ruling class). Unlike Marx, I do not ascribe all technology-effects as necessarily purposeful. Neither are all of these effects very obvious. Digital cameras with built-in computers and other computer-related electronics - chips, sensors and electronic circuits - are not “ineffecual and self-contained”, and thus work as conduits that serve as a part of the communication regulatory system that de facto both constrain and enable the communication act of sharing. (Eriksson, 2013, p. 27). At the heart of this regulatory system is the medium of digital code and its cycles of encoding, decoding and transmittance. This regulatory system is hard to fully account for and its effects are, in some ways, rather mysterious and abstract (cf. Skågeby, 2015). However, as this thesis hopefully shows, by interrelating instructional artifacts with users’ experiences, many of the video medium’s effects on users may be adequately defined and understood. This understanding, I suggest, enables designers of LAV-instructions to control and manage key aspects of digital videography in order to make the potential effects on users more purposeful. This leads to the core message of this thesis. Videography-related design process improvement schemes and design strategies, based on informed judgment rather than guesswork, make videography-driven design purposeful. In other words, purposeful videography-driven design is made possible, as information design scholar Dillon recommends (2017), through the application of science to information design. This is indeed the basic approach of this thesis. The guidelines put forward in this thesis are one concrete output of such a scientific application. These guidelines - simplistic as they may seem - are also one end-result of a concerted and extensive research journey that illustrates that there are valuable insights to be made by intersecting theories of neighboring disciplines centered on learning, perception, design, user studies and film production, respectively. However, my way of approaching instructional video design by combining separate theories does not necessarily contradict earlier scientific contributions into mediated instructional efforts, but allows for an analysis that accounts for the complexity of screen-based instructional communication.

The guidelines presented in this thesis indirectly call for the co-producer’s critical consideration for the videographer’s professional status and skill-sets, which are perhaps sometimes taken for granted. Moreover, this also calls for careful planning of the video content’s overall aim and purpose. For no videographer may save an inherently bad instructional communication strategy, although research shows that, in general, expert and highly motivated audiences are less sensitive to haphazard and improvised video and multimedia content. This is what Mayer discusses in terms of *boundary conditions* (2005). However, what more general pedagogical aims and purposes should entail, this thesis says very little about. Fortunately, other prominent scholars in the field of educational psychology provide many valuable insights regarding this question.

We have seen dramatic shifts over the last few decades in how instructional communication takes place, thanks to new digital technologies that bridge temporal and geospatial boundaries. With the increased pervasiveness of live-action video instructions, especially on YouTube, this thesis shows that it is critical to empirically evaluate the relationship of the user with these medial objects, in ways that account for the attributes and affordances in real-world instructional, mediated spaces. This thesis illuminates the powerful roles materiality and LAV-related affordances play in media use, and how these condition users’ engagement, in particular visual attention. By doing this, I extend existing information design research into the efficacies of instructional media in general, and transient media in particular.
5.4 Limitations and Future Research

To define material dimensions of a certain medium is not a particularly challenging scholarly feat (although it needs to be done). Especially, this is the case if such definitions already are explicitly addressed in neighboring research fields. To define affordances, on the other hand, is a different story. No measurements or data categories, no matter how refined, sophisticated or intelligible, or whether they are perceptual, reported, inferred or observed, may encompass and make all potential affordances that are relatable to live-action instructional videos identifiable. In fact, it is very unlikely that there ever will be one final, exhaustive set of affordances that are relatable to instructional videos. Indeed, affordances tend to be sublime and are so instinctual and routinized (Burlamaqui & Dong, 2015) that, depending on who perceives them, can easily become “invisible” (Ortmann & Kuhn, 2010). As noted by Stendal, Thapa & Lanamaki (2016), it is difficult to identify all affordances, to categorize and label them; moreover, they claim, it is not always clear who has the mandate to do so. Thus, on epistemological grounds, some scholars reject the notion that affordances can be evaluated at all, and/or made distinct from other environment-user relations. Some scholars argue that the idea that some aspects connoting relations can be studied directly is a suspicious one (Bygstad, Munkvold & Volkoff, 2016). This raises the issue of whether certain dimensionalities of camera movement, actors’ acting, the picture frame, negative spaces, aspects of immersion and phenomenology, and so on, may be examples of elements and phenomenon that are difficult to study in a more direct fashion. It is possible that such elements and phenomenon can only be understood through qualitative, extensive, investigations into their associated issues and events that reverberate far away from the instructional situation per se. Undoubtedly, instructional images actualize and enter into divergent associations in broader contexts than the mere, rather brief, instructional situation.

Yet, research efforts employing the concept of affordances to assess relational and interactivity aspects in environments that include various kinds of screen, communication and recording technologies are plentiful, and have increased over the last decade or so (Evans et al., 2017). This thesis is one example of such a research effort. In this study, I have used both more inductive and deductive approaches, and complimentary dual ones, including the assessment of different data patterns, in order to make determinations as to what facilitates certain actions in certain, circumscribed contexts. In addition, I have followed up on theoretical leads presented in the existing literature. The theoretical leads in question are one way of extending this thesis’ reach. This, I claim, is a fruitful way to explore the differential possibilities for action that are present in technology mediation contexts.

However, I also recognize that there is no singular way of doing this, and that other well-motivated research efforts might also prove to be constructive in this regard. The plethora of research approaches in the research fields in question are indeed indicative of this. One way of conducting research into the affordances of instructional videos is to focus on users’ decision-making in the moment of interaction. Although instructional LAVs represent different system complexities than airplane cockpits, one could use SA (Situation awareness) theory to do this. Possibly Endsley’s “error taxonomy”, based on the SA-model that identifies the underlying factors resulting in SA-errors in perception, comprehension and projection would prove useful in this regard (Endsley, 2015). Another way of doing this would be to approach the knowledge construction process in live-action video supported instructional efforts in concurrence with Vivian Sobchack’s phenomenology-infused film theory (1992). Drawing on Sobchack’s semiotic phenomenology, it may be possible to illuminate the signifying nature of users’ embodied visions in order to account for critical factors that augment cinematic intelligibility when users interact with LAV-instructions.

Since there are no other studies I am aware of that assess actual, perceivable, affordances of live-action instructional videos per se, this thesis should be viewed as a kind of base-line research effort. It is thus a small, but, I hope, sure-footed attempt to begin to explore the communication possibilities of this format. Future research should obtain additional samples associated with other user groups to cross-validate and reassess the affordance constructs presented in this thesis and their associated designerly implications. Given the relatively small number of affordances discussed here, and the
relatively small sample sizes, it should be relatively easy to replicate/investigate these within other contexts, especially those that center on specific media uses in specific organizational contexts.

I concur with Rice et al. (2017) that in research on affordance possibilities with respect to specific media types, there are many possible sources of contingency and non-representation, such as types of work, subgenre/hybrid media types, visual literacy capacities and national/organizational cultures. However, in this specific thesis, I believe I may have avoided some implications of generalizability, since the thesis represents a base-line research effort that revolves around affordances and how they relate to categories of human behavior - engagement - that are understood to be universal and which, in addition, exist in the relevant literature.

Nevertheless, behind the relatively anonymous data categories presented in this study there are real people, with real-life concerns, needs and preferences. Future research efforts into the live-action instructional format’s efficacies must begin to address the diverse and plentiful application arenas in which visual instructional communication efficacies are at stake. At the heart of such research endeavors is the prevailing idea that the relevant question is not whether technologies have features, whether signals and alerts exist or not, but whether individuals can increase or decrease engagement when interacting with mediated instructions, and what the knowledge outcomes are. Certainly, these relations are not as straightforward as they first may seem. By creating a coherent and uniform way of gaining more knowledge into the possibilities, both perceived and inferred, across a large group of LAV-users, in different places in organizational structures, we can develop more reliable representations of the breadth of affordances in a media technologically-enhanced environment in which bodily actions and engagement are key. This knowledge might prove useful, perhaps, in the not-so-distant future when LAV-instructions might be replaced by AR-applications. However, the ground is now laid to proceed to other more far-reaching research projects that investigate how “reality” might be enhanced with instructional, audiovisual materiality.

References


Ganier, F., & de Vries, P. (2016). Are instructions in video format always better than photographs when learning manual techniques? the case of learning how to do sutures. Learning and Instruction, 44, 87-96. doi:10.1016/j.learninstruc.2016.03.004


Watson, G., Butterfield, J., Curran, R., & Craig, C. (2010). Do dynamic work instructions provide an advantage over static instructions in a small scale assembly task? *Learning and Instruction, 20*(1), 84-93. doi:10.1016/j.learninstruc.2009.05.001


Wikström, A. (2013). Mälardalens högskola, innovation och produktrealisering, Akademin för innovation, design och teknik. *Storyboarding: Framing and reframing opportunities in the front-front end of innovation*


Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction, 22*(6), 449-457. doi:10.1016/j.learninstruc.2012.05.004


Questionnaires and Additional Information
**The Verisimilitude Reception Study (Paper A)**

Original questionnaire:

1. Hur trovärdig är denna video? Svara: 1–10 (10 är maximalt trovärdig)

2. Tror du att denna video kan användas som bevismaterial? (Svara ’Ja’ eller ’Nej’)

**The Image Quality Reception Study (Paper B)**

Original questionnaire:

1. Vad tror du skiljer filmerna åt?

2. Vilken version tycker du såg bäst ut?

3. Vilken version såg mest ut som en "biofilm"?

4. Vilken version tycker du hade bäst ljud?

5. Upplevde du någon gång att filmernas bildkvalitet var distraherande eller störde berättelsen? Om så, när och varför?
The Lego Truck Assembly Study (Paper C)

Original questionnaire:

QUESTIONNAIRE

Demographic Information

Name:……………………………………Age:…………………………………………………..

Education:…………………………………………………………………………………..

Rate your overall experience on a five-point Likert scale (1 being not very much to 5 being extremely skilled):

Lego building (1 2 3 4 5)………………………………………………………………………

Instructional medium that you have used in the experiment (1 2 3 4 5)……………………

Using instructions to learn how to perform tasks (1 2 3 4 5)………………………………

Usability Test

Mark the following answers using a five-point Likert scale (1-strongly disagree, 2-disagree, 3-some what agree, 4-agree, and 5-strongly agree)

<table>
<thead>
<tr>
<th>Statements</th>
<th>Scale (1 2 3 4 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructions were easy to understand</td>
<td></td>
</tr>
<tr>
<td>The organization of information in the instructions was easy to follow</td>
<td></td>
</tr>
<tr>
<td>The terminology used in the instructions was easy to understand</td>
<td></td>
</tr>
<tr>
<td>The instructions were easy to use as a reference as I was performing the tasks</td>
<td></td>
</tr>
<tr>
<td>The instructions helped me to select possible solutions quickly and apply them in the given task</td>
<td></td>
</tr>
<tr>
<td>I felt comfortable using the instructions to perform the tasks</td>
<td></td>
</tr>
<tr>
<td>The instructions were time-consuming to use</td>
<td></td>
</tr>
<tr>
<td>The instructions were frustrating to use</td>
<td></td>
</tr>
<tr>
<td>I could easily remember information from the instructions as I completed the tasks</td>
<td></td>
</tr>
</tbody>
</table>
I could find exactly what I was looking for in the instructions when I needed to do so
I could quickly locate the information I needed in the instructions when performing the task
My overall reaction to the design of instructions is very satisfactory

Open-Ended Questions

1. What did you like best about the instructions?

………………………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………………………

2. In what ways was the design of the instruction helpful?

……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
6. During assembly, how did you use or interact with the instruction, why?

……………………………………………………………………………………………………………………………………………………………
...
……………………………………………………………………………………………………………………………………………………………
...

7. Overall, which type of instructions did you prefer using for this type of specific tasks?

……………………………………………………………………………………………………………………………………………………………
...
……………………………………………………………………………………………………………………………………………………………
...

8. If you have any other comments, then please write it here.

……………………………………………………………………………………………………………………………………………………………
...
……………………………………………………………………………………………………………………………………………………………
...
The Anti-Predator Fence Construction Study (Paper F)

Original questions in Swedish:

1. Till vilken grad var informationen som finns i videon användbar och till nytta när ni byggde stängslet?
2. Till vilken grad var videons innehåll lätt att förstå och lätt att applicera?
3. Till vilken grad fyllde videoern specifika kunskapsbehov under byggnationen av stängslet?

YouTube-addresses to videos or/and versions of videos that are discussed in the thesis:

https://www.youtube.com/watch?v=igns0Tm0rYo
http://youtu.be/kt93U7dSo50
https://www.youtube.com/watch?v=SbLMHS5Vm74&t=16s
https://www.youtube.com/watch?v=eqPMK4map28
https://www.youtube.com/watch?v=NyfNmgq9bQY
https://www.youtube.com/watch?v=1Ec-xGBRiio
Videography as Design Nexus is a doctoral thesis in Innovation and Design on how to improve the usabilities of recorded video instructions via design-erly means. This thesis highlights the videographer’s craft, how he/she handles technological constraints, records procedural instructions and “edits in the head” while recording. In so doing the videographer must consider perceptual affordances and live-action video medium specificities with the video-user as both a biological and cultural being in mind.

In this study, several live-action videos, including story-driven films and typical procedural video instructions serve as Research Through Design (RTD) prototypes that permit critical inquiries with regards to their assumed communication efficacies. These efficacies are explored and assessed in tests that feature actual users and audiences in more or less ecologically valid assembly trial set-ups, as well as in blind tests that compare different variants of video styles. A number of these tests also include eye-tracking as a method in order to identify how live-action video instructions’ usability may be conditioned by users’ basic visual decoding strategies. Moreover, YouTube data is gathered and analyzed as a complementary way of circumscribing a video’s usability aspects in the online, screen-based, learning context.

The conclusion of this research undertaking is that medium specificities of live-action videos influence how such videos are appreciated and experienced by users in the moment of interacting with them. For example, this study shows how degrees of perceived verisimilarities, in part, depend on camera-handling techniques and choice of recording gear. Other findings imply that live-action video instructions are not so sensitive to unfocused viewing behavior, and that the display of discontinuity in recorded videos results in users experiencing increased cognitive load.

Thus, this study addresses both first- and second-order understandings of instructional video communication. It does so by demonstrating how both sublime image quality factors and the video instructions’ overall communication congruence impact on the ability of the video medium to support procedural knowledge and purposeful information processing of users. We also gain some concrete recommendations with regards to how to manage actual live-action video production constraints as videographer or designer of recorded video instructions, for the betterment of the video-user.

Per Erik Eriksson is a Ph.D. candidate in Innovation and Design at the School of Innovation, Design and Engineering, Mälardalen University, Sweden. He holds a Ph.Lic. in Innovation and Design at Mälardalen University, an MA in Communication at Stanford University, USA and a BFA at Massachusetts College of Art and Design, USA. Eriksson is a former TV producer, editor and documentary filmmaker. Since 2006, he holds a position as Lecturer in Moving Image Production at Dalarna University, Sweden. He is currently Director of the Short Story Audiovisual Design Program at Dalarna University. Eriksson is also part of the ISTUD research profile, as well as the Audiovisual Research group at Dalarna University, with a research focus on audiovisual knowledge in design processes.