



Universität Regensburg

Empirically Measuring Saliency of Objects for Use in Pedestrian Navigation

An der Fakultät für Sprach-, Literatur- und Kulturwissenschaften der
Universität Regensburg zur Erlangung des akademischen Grades eines
Doctor philosophiae (Dr. phil.) eingereichte Inaugural-Dissertation

von: Markus Kattenbeck
geboren in: Regensburg

Erstgutachter: Prof. Dr.-Ing. Bernd Ludwig
Zweitgutachter: Prof. Dr. techn. Sabine Timpf (Augsburg)

Eingereicht am: 4. Mai 2016
Tag der Disputation: 22. Juli 2016

*Begegnet uns jemand, der uns Dank
schuldigt ist, gleich fällt es uns ein. Wie oft
können wir jemandem begegnen, dem wir
Dank schuldig sind, ohne daran zu denken!*

Johann Wolfgang von Goethe,
Maximen und Reflexionen, 5

Ziemlich genau 69 Monate ist es her, dass ich
als Doktorand und wissenschaftlicher Mitarbeiter
an den Lehrstuhl für Informationswissenschaft
zurückkehren konnte.

Ziemlich genau 15 Wochen sind vergangen, seitdem mein persönlich
und wissenschaftlich verehrter Mentor und Lehrer,
mein Doktorvater, Professor Dr. rer. soc. Rainer Hammwöhner,
verstorben ist.

Vor genau 6,5 Wochen habe ich meine Dissertation eingereicht.

Nun schreibe ich also am Abend des 22. Juni 2016, des ersten Tages,
an dem meine Dissertation inklusive der Gutachten ausliegt, dieses Vor-
wort.

Ganz besonders bedanken möchte ich mich bei den beiden Perso-
nen, die innerhalb so kurzer Zeit meine Arbeit bewertet haben. Prof. Dr.
Bernd Ludwig war, trotz mancher Scharmützel, die wir uns am Anfang
meiner Arbeit an dieser Dissertation um empirische Herangehenswei-
sen geliefert haben, ursprünglich bereit, mein Zweitgutachter zu sein. Er
war es auch, der nach dem plötzlichen Tod von Prof. Dr. Hammwöhner
ohne zu zögern eingewilligt hat, meine Dissertation als Erstgutachter
zu begutachten. Er stand mir mit Rat und Tat in den Endspurtwochen
meiner Schrifffassung zur Seite und tut dies bis heute. Danke, Bernd.
Prof. Dr. Sabine Timpf danke ich dafür, dass sie nach dem Tod von Prof.
Dr. Hammwöhner—noch bevor wir uns persönlich kannten—zugesagt
hat, meine Arbeit als Zweitgutachterin zu bewerten. Ich hoffe, dass sich

auch nach der Beendigung meines Promotionsverfahrens die Möglichkeit bietet, mit der Augsburger Geoinformatik zu kooperieren.

119 Versuchspersonen waren bereit, mit mir bei Wind und Wetter durch das Altstadtgebiet von Regensburg zu gehen. Ohne diese Leidenschaft und das damit zum Ausdruck gebrachte Interesse an meinem Forschungsvorhaben wäre dieser Text undenkbar.

Dank gebührt auch allen Kommilitoninnen und Kommilitonen im Doktorandenseminar unseres Hauses – dafür, dass sie meine oft kritischen, niemals aber böse gemeinten Fragen zu empirischen Sachverhalten so geduldig ertragen haben. Ebenso danke ich Prof. Dr. Christian Wolff, der wesentlichen Anteil an meiner Rückkehr an den Lehrstuhl hat, indem er als „prominenter Fürsprecher“, wie Prof. Dr. Hammwöhner es formulierte, für mich fungierte. Prof. Dr. Daniel Drascek, Prof. Dr. Christof Kuhbandner, Prof. Dr. Heidrun Stöger, Dr. Erika Lindig, Dr. Karin Lahoda sowie Dr. Ludwig Kreuzpointner danke ich herzlich für die Erlaubnis, in ihren Lehrveranstaltungen Werbung für mein Experiment machen zu dürfen. Letzterer hatte immer auch ein offenes Ohr für mich und das ein oder andere Experiment- und/oder Fragebogenproblem, wofür ich sehr dankbar bin. Robert Jackermeier hat die Grundlage für die App geschaffen, auf der ich mein Erhebungsinstrument aufgebaut habe; Dr. Richard Schaller hat den Code zum versenden der Experimentdaten zur Verfügung gestellt. Vielen Dank Euch beiden.

Ich hatte das große Glück, die ganzen Jahre hinweg mit den angenehmsten Kolleginnen und Kollegen und Freunden zusammenzuarbeiten, die man sich denken und wünschen kann. Ihnen allen schulde ich großen Dank. Manuel Müller in seiner stets ungemein reflektierten und zugleich zurückhaltenden Art, den ich immer um seine formidablen Mathematikkenntnisse beneiden werde; Christina, für gemeinsame, gut sichtbare Veröffentlichungen; Florian Meier, mit dem ich relativ selten, wenn dann aber umso intensiver über statistische Dinge sprechen konnte – was ich immer sehr inspirierend fand; Markus Fuchs, mit dem ich manchmal stundenlang über die Lehre gesprochen habe und davon immer profitieren konnte; Dr. Stefan Bienk, zu dessen Fähigkeit, Formeln wie Fließtext lesen zu können, ich immer aufschauen werde; Dr. Hanna

Knäusl, mit der ich noch studiert habe und mit der ich einige Monate ein Büro teilte; Renate Vogl, die mir immer in allen möglichen administrativen Dingen den Rücken frei hielt und mich des Öfteren mit Nervennahrung in Form von Gummibärchen versorgte; PD Dr. Jürgen Reischer, dessen Anmerkungen mich manches Mal zum Überdenken der verwendeten Begriffe angeregt haben; Barbara Ströhl, mit der ich seit fast zwei Jahren ein Büro teilen darf, die meinen Hang zur Unordentlichkeit mit dem ihr eigenen Gleichmut ertragen hat und mich auch schon vor unserer gemeinsamen Bürozeit oft in meinem Vorgehen/Vorhaben bestärkt hat; Dr. Manuel Burghardt, der durch sein Verständnis für meine Art zu denken und zu arbeiten ganz oft den richtigen Ton gefunden hat, um mich ein Stück von meinem eigenen Streben nach Perfektion abzubringen; der bereit war, mit mir das Doktorandenseminar über Jahre hinweg und vor allem auch nach seiner Promotion zu moderieren; Christoph Pfeiffer, der mit seinen fundierten Textsatz-Kenntnissen so wesentlich zur Ansehnlichkeit dieses Textes beigetragen hat; Christine Fraunhofer, die durch zweimaliges Korrekturlesen alle formalen Fehler getilgt hat (falls doch noch welche vorhanden sind, so sind diese absichtlich vom Verfasser eingefügt); beiden danke ich, dass sie in den letzten Wochen vor meiner Abgabe stets ein offenes Ohr für mich hatten; schließlich Dr. David Elsweiler, mit dem ich fast drei Jahre ein Büro teilte, dessen untrügliches Gespür für wissenschaftliche Fragestellungen, deren empirische Ausarbeitung und die zugehörige Dissemination für mich ein unerschöpflicher Quell der Inspiration waren. Von der Tatsache, dass er bereit war, Rohentwürfe dieses Textes zu lesen und mit Anmerkungen zu versehen, ganz zu schweigen. Ta muchly, mate, I owe you so much.

Den größten Dank aber schulde ich Herrn Prof. Dr. Rainer Hammwöhner, mit dem ich viele Stunden lang über meine Dissertation und andere Dinge des menschlichen (Zusammen-)lebens sprechen durfte. Er war es, der in seiner unnachahmlichen Art, wie er sich ausdrückte, „jedes Haar mehrmals zu spalten“, mein Denken im Allgemeinen und den wissenschaftlichen Fortschritt meiner Arbeit im Besonderen geprägt hat, wie niemand sonst. Sosehr ich es bedauere, dass er den heutigen Tag nicht mehr miterleben kann, sosehr bin ich davon überzeugt, dass sein Geist und seine Art, die Dinge anzugehen, in mir und meiner Arbeit fortleben. Jedes noch so kleine Detail mehrmals zu durchdenken ist etwas, das ich

für's Leben, nicht nur für die Wissenschaft von ihm gelernt habe. Ich bin stolz darauf, dass ich sein Schüler sein durfte.

Auch wenn sie es vehement von sich weisen würden, so haben meine Eltern großen Anteil an dieser, meiner Dissertation. Sie gaben mir von Kindesbeinen an die Freiheit, mich mit dem Nachdenken auseinanderzusetzen. Glaubt mir, Mama und Papa, ohne diese Freiheit wäre ich heute nicht der, der ich bin.

Schließlich danke ich meiner über alles geliebten Frau Anita. Wir sind fast genau so lange verheiratet, wie es gedauert hat, bis ich nun mein Promotionsverfahren erfolgreich abschließen konnte. In diesen Jahren haben wir einiges zusammen erlebt und durchgemacht.

Du, Anita, hast immer zu mir gehalten und mich in Allem und mit
Allem unterstützt.

Ohne Dich, Anita, wäre diese Arbeit undenkbar.

Dir, Anita, widme ich diese Arbeit.

ABSTRACT

BACKGROUND

Humans usually refer to landmarks when they give route directions to pedestrians. One of the reasons why current mobile pedestrian navigation systems do not yet mimic this mode of communication is the lack of available data sources. The usefulness of a crowd-sourced data acquisition approach to overcome this problem has long been mooted. However, to date no empirically sound way of measuring the *saliency* of objects by means of surveys exists.

GOAL

Given this background, this doctoral work has three goals:

1. To achieve a sound way of measuring *saliency* and its subdimensions, i.e. *visibility in advance*, *cognitive saliency*, *prototypicality*, *structural saliency*, and *visual saliency* based on taking dimensions revealed in earlier studies systematically and simultaneously into account.
2. To find subgroups of visual features among the large number of visual attributes known from the literature.
3. To find the most important subdimensions of *saliency* by means of estimating two different structural equation models. *Model I* is based on assumptions of independence among subdimensions, whereas *model II* reflects hypotheses of mediation.

Taken as a whole, achieving these goals will foster both, the advancement of theories of *saliency* and landmark acquisition methods.

METHODOLOGY

A large scale, in-situ experiment was implemented, trying to overcome weaknesses of earlier attempts made to estimate *saliency*. An

appropriate sample size of buildings and non-buildings was calculated a priori ($n_{obj} = 360$). Objects were randomly selected based on their geographical coordinates and randomly grouped into $n_r = 55$ routes. Participants were required to rate objects by means of a survey. The questions were derived from empirical evidence found in earlier studies. Each route was walked by two different participants ($n = 112$), id est (i.e.) two ratings per object were collected for data analysis.

FINDINGS

Model I and *model II* were analyzed using PLS Path Modeling and consistent PLS Path Modeling, respectively. The measurement models proposed showed a good fit, although some weaknesses were identified for *prototypicality* and *cognitive salience*. Geometrical aspects as well as features like (visual) age turned out to have a stronger impact on *visual salience* than color. *Model I* did not yield reasonable structural model results based on consistent Partial Least Squares Path Modeling. *Model II*, however, showed that *visual salience* had a very high impact on *visibility in advance* which, in turn, heavily influenced *structural salience*. An analysis of the predictive capabilities of *model II* revealed important, but rather small effects.

VALUE OF WORK

This doctoral work adds to *salience* models as well as to its empirical, survey-based, in-situ measurement. The results of the mediation analysis as well as the predictive capabilities of *model II* suggest that important subdimensions of *salience* are missing in current theories. *Emotional salience* and *familiarity* are identified as two candidate constructs. The structural relationships found during the analysis of *model II* provide, in combination with the measurement model results, a sound basis to choose important features for surveys which are usable to gain crowd-sourced *salience* ratings. Furthermore, several important aspects for future studies are identified. These include heterogeneity analyses for different subgroups of users of pedestrian navigation systems as well as local environments different to the historic one used in this study.

CONTENTS

1	INTRODUCTION	1
1.1	Research Agenda	5
1.2	On the Pages Yet to Come	6
1.2.1	What Landmarks Are and What the Meaning of Saliency Is	6
1.2.2	Deriving a Measurement Model for Saliency . . .	7
1.2.3	A Random Sample of Objects and Further Prepa- rations	7
1.2.4	Assessing Different Structural Equation Models for Overall Saliency	7
1.2.5	Relating the Empirical Results Achieved to Prior Research	8
1.2.6	What Do the Results Mean—and What Remains to Be Done?	8
1.3	Why Saliency Estimation Research	8
1.4	Related Publications	11
1.4.1	With a Strong Focus on Topics Covered in This Thesis	11
1.4.2	In the Broader Field of Landmark-Based Pedes- trian Navigation Systems	12
2	SALIENCY AS A CONCEPT	13
2.1	Introduction	13
2.2	On Objects Used to Orient Oneself	13
2.2.1	The Origin	13
2.2.2	Current Usage—and Why Points of Interest and LMs Are Not Synonyms	15
2.3	A Review of Appropriate Literature on Saliency	18
2.3.1	In the Information Science Community	18
2.3.2	In Computer Vision	20
2.3.3	In the GIScience-Community	23
2.4	Towards a Preliminary Definition of Saliency	30
2.5	Chapter Summary	32

Contents

3	HOW CAN SALIENCE BE MEASURED?	35
3.1	Introduction	35
3.2	A Review of Methods Proposed in the Literature	35
3.2.1	Automatic Assessment	36
3.2.2	Lab-Based Experiments	39
3.2.3	In-Situ-Experiments	41
3.2.4	Criticisms and Their Impact on Empirical Design	41
3.3	Introducing SEM, PLS-PM.	44
3.3.1	Why Use Structural Equation Models—Pros and Cons and a Decision	45
3.3.2	How PLS Path Modeling Works	49
3.3.3	The Measurement Model	59
3.3.4	About the Phrasing of Questions	69
3.4	Chapter Summary	72
4	EMPIRICAL METHOD	75
4.1	Introduction	75
4.2	Objects, Participants, Routes	75
4.2.1	Considering a Suitable Number of Objects	75
4.2.2	A Set of Randomly Selected Objects	76
4.2.3	Routes	80
4.2.4	Number and Acquisition of Participants	81
4.3	Apparatus	83
4.3.1	Application	83
4.3.2	Measuring Sense of Direction	87
4.3.3	Demographic Data	89
4.4	Procedure	91
4.4.1	Assigning Routes Randomly	91
4.4.2	Conducting Experiments	91
4.5	Chapter Summary	92
5	EMPIRICAL RESULTS	95
5.1	Introduction	95
5.2	Descriptive Statistics	96
5.2.1	Participants and Trials	96
5.2.2	Objects	97
5.2.3	Routes	101
5.2.4	Influence of Weather Conditions on Ratings	101

5.3	A First Structural Model	106
5.4	Estimating SEMs Using PLS-PM	107
5.4.1	General Steps and Parameters	107
5.4.2	Assessing the Reflective Measurement Model . .	109
5.4.3	Assessing the Formative Measurement Model . .	117
5.4.4	Assessing the Structural Model	128
5.5	Adapting the Model	132
5.5.1	Adapted VISF	133
5.5.2	Constructs Do Have an Influence on Each Other	137
5.6	Chapter Summary	152
6	DISCUSSION	155
6.1	Introduction	155
6.2	Ongoing Methodological Discussion.	155
6.2.1	Formative Measurement in General	155
6.2.2	Partial Least Squares Path Modeling (PLS-PM) and Formative Measurement	161
6.2.3	PLS-PM in General	163
6.2.4	Specific Consequences for My Analysis	167
6.3	The Measurement Models	168
6.3.1	The Reflectively Measured Part	169
6.3.2	Interpreting the Results for Visual Salience Mea- surement	172
6.4	The Structural Model Results	174
6.4.1	PLSc results	174
6.4.2	Discussing the Results	176
6.5	General Limitations	181
6.6	Chapter Summary	185
7	CONCLUSIONS AND FUTURE WORK	187
7.1	Introduction	187
7.2	Contributions	187
7.2.1	With Respect to Landmark Identification	187
7.2.2	With Respect to Theory about Salience	188
7.2.3	With Respect to Collaborative Landmark Mining	189
7.2.4	Regarding Experimental Design in GIScience . .	191
7.3	Future Work	192

Contents

7.3.1	Find New Structural Models Using Bayesian Networks	192
7.3.2	Heterogeneity Analysis	193
7.3.3	Emotional Salience and Familiarity Experiments	194
7.3.4	Use Different Local Environments	195
7.3.5	The Model in Different Mental Contexts	195
7.4	Chapter Summary	196
A	APPENDIX	197
A.1	Surveys	197
A.1.1	Assessing Sense of Direction by Using a German Language Scale	197
A.1.2	Questions Not Used in the Measurement Model	199
A.1.3	Demographic Data	200
A.2	Trials	201
A.2.1	Trial Statistics Per Day	201
A.2.2	Route Statistics	203
A.3	Participants	205
A.3.1	Explanations Given	205
A.3.2	Research Diary	208
B	LIST OF ABBREVIATIONS	213
	List of Figures	215
	List of Tables	217
	References	219

I INTRODUCTION

Human navigation is an intrinsically complex task, involving a diverse range of spatial cues, computational mechanisms and spatial representations (confer (cf.) Wolbers & Hegarty, 2010). “Fundamentally, wayfinding is the process of finding your way to a destination in a familiar or unfamiliar setting using cues given by the environment” (Farr et al., 2012, page (p.) 2). Despite its complexity, humans are able to navigate successfully on a day-to-day basis. Up until now the research disciplines involved have been unable to give a concise insight into navigational abilities. However, the importance of landmarks (LMs) for human navigation is undoubted across disciplines. LMs have been known as fundamental to spatial environment learning in human spatial cognition for 40 years (Siegel & White, 1975). They are essential to the representations of spatial configurations (cf. exemplum gerens (e.g.) Couclelis et al., 1987). They play a central part in everyday route communication (cf. Daniel & Denis, 1998), and support human navigation in large scale spaces, i.e. in cases where the destination may not be seen from the starting point (cf. e.g. Tom & Denis, 2003; Vinson, 1999). Prototypical systems using LMs revealed their usefulness in supporting human wayfinding of pedestrians and drivers, alike (cf. Richter, 2013, p. 83). Despite the importance of LMs for human wayfinding their use is not widespread in commercial pedestrian navigation systems (PNSs) (cf. Richter, 2013, p. 84). However, users of these systems expect LMs to be present in route instructions (cf. e.g. May et al. (2003), who report on a requirements study concerning this phenomenon). I will describe fundamental challenges for the widespread use of LMs in PNSs, which have been unresolved by an example:

Imagine a person P_1 wants to find her way through the Old Town of *Regensburg*, a city located in Germany at the most northerly point of the river *Donau*. More specifically, P_1 wants to walk from the *Reichstagsmuseum*, which is located in the Old Town Hall, to one of the most

1 INTRODUCTION

important churches in the city, the so-called Old Chapel (*Stiftskirche Unserer Lieben Frau zur Alten Kapelle*), which is located at a square called *Alter Kornmarkt*. Figure 1 shows both, starting point *A* and destination *B*, respectively.



Figure 1.: An OpenStreetMap (cf. Haklay & Weber, 2008) map which depicts the example route. The route was added using QGIS (cf. QGIS Development Team, 2014) and annotations were made with GIMP (cf. The GIMP Team, 1997-2013).

Imagine further, P_1 uses a PNS S installed on a smartphone to get directions.¹ Suppose S issues the following description on-line², i.e. each statement S_n is not presented to P_1 before the distance mentioned in S_{n-1} has been walked.

- S_1 : On *Rathausplatz* head towards southeast to *Kohlenmarkt*. DISTANCE: 16m
- S_2 : Stay in *Kohlenmarkt*. DISTANCE: 58m.

1 Suppose a textual representation of the route description by S . Furthermore, imagine P_1 to behave perfectly cooperatively, i.e. P_1 is able to understand and willing to follow the description.

2 The description is taken from Google Maps (Google Inc., 2016). Route directions were translated to English.

S_3 : Turn right to stay in *Kohlenmarkt*. DISTANCE: 49m

S_4 : Turn left into *Kramgasse*. DISTANCE: 130m

S_5 : Turn right onto *Domplatz*. DISTANCE: 170m

S_6 : Continue into *Domstraße*. DISTANCE: 59m

S_7 : Turn right onto *Alter Kornmarkt*. DISTANCE: 45m

Although being presented sequentially during navigation this description uses only distances and cardinal directions. These can be derived easily from official data sources, e.g. provided by geoinformation agencies. This means, S does not use salient objects (SOs) to orient P_1 and, therefore, neglects the LM-based mode of instructions preferred by users (cf. e.g. Rehrl, Häusler, & Leitinger, 2010). Severe confusion will likely result: It is hard to make sense of instructions S_2 and S_3 due to the vagueness of the expression “stay in the square”. In case of automobile navigation these instructions may prove useful, because cars must follow the course of the road. Pedestrians, however, need not stick to pathways et cetera (etc.).

In contrast to S one would expect a PNS L using LMs provides route directions to a person P_2^3 in a very different manner.

L_1 : Leave the Old Town Hall and cross the square towards *Café Princess*.

L_2 : Currently, you are at the *Rathausplatz*. However, please note that the *Rathausplatz* adjoins to another square called *Kohlenmarkt*. Please walk to the *Kohlenmarkt*.

L_3 : Cross the *Kohlenmarkt*, thereby heading to the burger restaurant called *Hans im Glück*.

L_4 : Turn right in front of the entrance and walk alongside the building.

L_5 : The next building hosts a shop called *Schwesternliebe*. Turn left immediately into *Kramgasse* once you have passed this building.

³ The preconditions mentioned above remain.

- L_6 : Orientate yourself towards the cathedral's towers. Follow the *Kramgasse* to the cathedral.
- L_7 : Leaving *Kramgasse* you are entering the *Domplatz*. Please head towards cathedral's western entrance.
- L_8 : Reaching the cathedral you will recognize a monument depicting the Bavarian king *König Ludwig I.* riding a horse to your right hand side.
- L_9 : Walk to the monument and turn left in front of it. You will see a large yellow building called *Weltenburger am Dom*, located next to the monument. Take the passage.
- L_{10} : Once leaving the passage you have entered the square called *Alter Kornmarkt*, already. Walk alongside the large yellow building and you'll reach the Old Chapel's Main Entrance within a few steps.

Although these are turn-by-turn directions, L_1 - L_{10} are fundamentally different to the description given by S . Instead of relying on cardinal directions and distances, L uses salient objects (SOs) in order to identify decision points (DPs). This means, objects are used to guide users which draw an observer's attention in a local environment, i.e. which are salient. I will provide a rationale why landmarks and salient objects are used as synonyms throughout this thesis in section 2.2 and give a thorough discussion of *salience* in different disciplines in section 2.3. This example further uses SOs not only to identify DPs but to confirm the right track as well. L_5 is a good example, where the shop is used to acknowledge that P_2 is on the right track. However, several issues must be resolved in order to provide directions of this kind on a scalable basis.

AVAILABILITY SOs are beyond the scope of official data sources. Whereas buildings may be part of these sources, the features rendering them salient (e.g. the color of façades) are not part of these. Beyond building-data, various other types of objects may become useful in route instructions. However, data about these objects is not readily available and collecting such data is time-consuming (cf. Richter, 2013).

SELECTION Even if a sufficiently large database of LMs and their corresponding *salience* were available, in order to provide route instructions in a PNS on a smartphone, strategies to select the most appropriate LM must be developed. These may be highly dependent on both, the current route and the person navigating. For example, L_6 does not make sense to users, who do not know what a cathedral looks like.

EXPLANATION The explanations to be given in LM-based route instructions are rather different compared to providing distances. Two possibilities come to mind: textual representations and visual ones. Whereas metric distances may easily be given using text and arrows (e.g. S_3 may be accompanied by an arrow pointing to the right), suitable expressions to describe LMs may be difficult to derive. Some entities may be referred to using well-known concepts (e.g. the monument referred to in L_8) whereas others may be both, difficult to describe and visualize: The cathedral's western portal used in L_7 is highly visually salient but may be hard to visualize without a three-dimensional (3D)-model of the Cathedral.

Research literature exists relating to all of these aspects. Nevertheless, the problems remain unsolved. In particular, there are no comprehensive information about LMs and their *salience*.

1.1 RESEARCH AGENDA

From a long-range perspective it is desirable to guide users of PNSs based on objects drawing an observer's attention. However, the amount of data needed to fulfill this goal is vast. Motivating users to engage in providing *salience* estimations, i.e. using a crowd-sourced like approach, seems a promising approach to address this problem (cf. Richter, 2013). Currently, attempts to empirically sound, survey-based in-situ measurement of the *salience* of objects do not exist. The knowledge on LM identification, however, has generally been growing for years. Theories about *salience* were established (cf. e.g. Raubal & Winter, 2002; Sorrows & Hirtle, 1999) and several empirical studies (cf. e.g. Ishikawa & Nakamura, 2012; Nothegger et al., 2004) were published reporting on dimensions

revealed to measure *saliency* in small-scale in-situ studies. What is still missing (cf. Duckham, Winter, & Robinson, 2010; Richter, 2013) is a large scale experiment, bridging theories and empirical evidence. It is of benefit, however, to empirically reveal which of the known dimensions are most important and how the different sub-dimensions of *saliency* relate to each other. This kind of knowledge can foster the development of theory on *saliency* as well as its empirical measurement. Consequently, the three aims of this thesis are:

1. To provide a critical overview of prior empirical and theoretical work regarding *saliency*.
2. To derive a measurement model for each of the proposed sub-dimensions of *saliency*.
3. To analyze two different structural models for *overall saliency* based on data of a large scale in-situ experiment. This analysis includes the assessment
 - a) whether sensible subgroups of visual dimensions can be derived and to determine how their importance differs.
 - b) whether the influence of visual aspects is mediated by other latent variables (LVs).

1.2 ON THE PAGES YET TO COME

1.2.1 *What Landmarks Are and What the Meaning of Saliency Is*

As this text is dedicated to the measurement of *saliency* of landmarks, Chapter 2 presents a review of literature on the major concepts of this thesis. It starts with some historical remarks on LMs as a concept and provides a rationale for an up-to-date synonym understanding of landmarks (LMs) and salient objects (SOs). The review of literature on the concept of *saliency* starts with its usage in important subfields of Information Science and the area of Computer Vision. Both domains provide major insights into the theories on *saliency* used within Geographical Information Science (GIScience). This leads to the operational definition of *saliency* employed throughout this doctoral work.

1.2.2 *Deriving a Measurement Model for Salience*

Based on these considerations, Chapter 3 is dedicated to methodology in terms of data collection and analysis. Its aims are fourfold:

1. Provide a rationale for the design of a large-scale, empirical in-situ study based on weaknesses of prior studies.
2. Justify the decision to use Structural Equation Models (SEMs) as a data modeling technique and Partial Least Squares Path Modeling (PLS-PM) as a mode of data analysis.
3. Explain how PLS-PM works.
4. Derive a measurement model for the LVs introduced in Chapter 2 from prior theoretical and empirical work.

1.2.3 *A Random Sample of Objects and Further Preparations*

Complementing the measurement model development, Chapter 4 focuses on the data acquisition method used in this thesis. This includes sample size considerations with respect to the number of objects and participants. The way objects were randomly selected and how they were grouped into routes will be described, too. The presentation of the questions to participants and how demographic data and sense of direction (SoD) was assessed will be detailed, afterwards. Finally, the procedure taken to conduct experiments will be described.

1.2.4 *Assessing Different Structural Equation Models for Overall Salience*

The experimental procedure leads to the empirical results presented in Chapter 5. It starts with descriptive statistics about participants (including their SoD), objects and routes. This section finishes with an analysis of the influence snowy weather had on ratings. Using PLS-PM as a mode of analysis the measurement models for LVs are analyzed in detail. Two different structural models connecting those LVs will be examined, subsequently. *Model I* makes extensive use of assumptions

of independence whereas *model II* draws on mediation hypotheses. Furthermore, an adaptation to the measurement model of *visual salience* will be proposed by building subgroups of measured variables (MVs).

1.2.5 *Relating the Empirical Results Achieved to Prior Research*

Discussing these results in Chapter 6 is done from a methodological perspective as well as in relation to prior empirical and theoretical results. Several recent methodological advancements in theory about formative measurement and PLS-PM have to be discussed with respect to their meaning to the results of this doctoral work. In particular, a reassessment based on Consistent Partial Least Squares Path Modeling (PLSc) will shed light on both structural models. Based on this assessment, the findings will be discussed with respect to the results of prior studies, including proposals for theoretical advancement. The limitations of the current work build the last section of the discussion. They focus on the environment used, limitations in terms of statistical methods, and restrictions applying to sampling of participants.

1.2.6 *What Do the Results Mean—and What Remains to Be Done?*

Finally, conclusions will be made regarding the contributions of this doctoral work. It adds to both, theory about and measurement of *salience*. Furthermore, the findings can be used to implement a crowd-sourcing based approach to LM identification. Some implications for experiments involving LMs are given, too. These findings lead to future work, which includes the use of different local environments and user groups. Assessment of *emotional salience* as well as a cross-validation of the results using qualitative methods will lead to further insights into both, *salience* theory and measurement.

1.3 WHY SALIENCE ESTIMATION RESEARCH IS RELEVANT TO INFORMATION SCIENCE (IN REGENSBURG)

In 1992 Michael F. Goodchild coined the term *Geographical Information Science (GIScience)* to describe a shift of focus in geography which reflected

the debate about the role of science in geography (cf. Goodchild, 1992). 11 years later, Mark (2003) cites the By Laws of the University Consortium for Geographic Information Science⁴, a definition Goodchild (2010) stresses to be particularly important to the development of the field:

“[GIScience is dedicated to] [t]he development and use of theories, methods, technology, and data for understanding geographic processes, relationships and patterns” (University Consortium for Geographic Information Science, 2016).

LMS are thought to belong to the foundations of GIScience and have been an aid to the advancement of the field over the last 20 years and are supposed to remain influential in the next 20 years (cf. Onsrud & Kuhn, 2016). LMS are considered an important part of GIScience and its aim to deliver systems and services adapted to users (cf. Richter & Winter, 2014). As I will show that LMS and *saliency* are closely related (cf. Chapter 2), some rationale is only needed, why an information scientist from *Regensburg* focuses on *saliency* in pedestrian navigation. First of all, the concept of *saliency* is not new to the broader field of Information Science as well. Indeed, it has a rather long tradition in Information Retrieval (IR) research. Chapter 2 will consequently provide an insight into commonalities between the IR understanding of *saliency* and its GIScience counterpart. Second, generally speaking, research in mobile PNSs has important connections to major concerns of Information Behaviour Research. I have provided a rationale elsewhere (cf. Kattenbeck et al., 2015) in German language and will repeat some of the reasons here. Among others, Information Behaviour research is concerned with three major aspects:

INFORMATION SEEKING

i.e. seeking information suitable to fulfill a particular information need (cf. e.g. Wilson, 1999).

INFORMATION AVOIDING

i.e. actively ignore information in order to successfully fulfill the current information need (cf. e.g. Godbold, 2006).

⁴ These remain still unchained, which is the reason why I cite the current source.

1 INTRODUCTION

CONTEXT

i.e. all aspects influencing information behaviour such as personal and situational aspects or time constraints (cf. e.g. Ingwersen & Järvelin, 2005).

Most of the published work is about Information Seeking (cf. e.g. Case & Spink, 2012). It is relevant to research about pedestrian navigation, too. The relatedness of navigation in both, real-world and electronic spaces is stressed from the early days of *saliency* theories in GIScience-related research, too (cf. Sorrows & Hirtle, 1999). In fact, navigating in unknown environments results in actively seeking directions as a piece of information. Metaphors of navigation have been prevalent in Information Seeking research for many years. As early as 1993, O'Day and Jeffries (1993) describe search session as wayfinding behaviour. 19 years later West and Leskovec (2012) depict the usage of links in Wikipedia as navigation in information spaces. Recent publications in Information Behaviour research stress the importance of Context and Information Avoiding (cf. e.g. Sin, 2011). Information avoiding is relevant to PNS scenarios, e.g. in studying what pieces of information about the full route are actively ignored. Similarly, context is of major importance in *saliency* estimations for pedestrian navigation. Current theories of the *saliency* of objects stress the influence local surroundings have in rendering a particular feature salient (cf. Caduff & Timpf, 2008). This holds particularly true for *visual saliency*. A red-colored façade of a building may become salient in a neighborhood with only white-washed houses. However, if there are two or more red houses in the local surroundings, this house may become far less salient. Overall, focusing on *saliency* in this doctoral work means to find an empirical approach to a widely known concept applied to the field of pedestrian navigation, which is relatively new to Information Science. In that sense, this thesis adds to the foundations of an information system answering a particular information need, i.e. to enable route instructions based on SOs. In order to be able to enhance PNSs by means of SOs a scalable method to gain this data is necessary. The survey-based approach presented in this doctoral work is a major possibility to acquire this data. Gaining an insight into the capability of the proposed model to predict *overall saliency* is an important step to enable collaborative acquisition of data about LMs.

The *salience* values calculated from these ratings can be used to select the most useful object in future route instructions. The data analysis method used (cf. section 5.4) will provide a detailed account of the importance of particular features to measure *salience* and reveal the impact different dimensions of *salience* have on each other. These insights can be used to ask the most important questions about an object's *salience*. This means, the results from this study can be used to choose the most important aspects to be asked in a real-world crowd-sourcing scenario. Users of PNSs may not be willing to contribute *salience* data, if it is a time consuming task. Therefore, it is desirable to know which of the questions are most important to predict *overall salience*.

1.4 RELATED PUBLICATIONS

1.4.1 *With a Strong Focus on Topics Covered in This Thesis*

In the course of completing the research described in this thesis, I published four articles/papers covering specific aspects presented in this thesis:

Kattenbeck (2012) gives an early-stage outline of the thesis, which had a slightly different focus at that time.

Kattenbeck et al. (2013) details the way data provided by the Bavarian Agency for Surveying and Geographic Information was incorporated in the database (DB) several PNS prototypes were based on.

Kattenbeck (2015b) provides the results of a pre-study about the way participants understand questions.

Kattenbeck et al. (2015) explains why research in PNSs is research in Information Behaviour.

Kattenbeck (2015a) is a milestone publication giving a detailed account of the SEM based on independence assumptions (referred to as *model I* throughout this thesis) and the results of its analysis using non-consistent PLS-PM. These results as well as the literature review given there are used in this doctoral work and will be incorporated in a larger and more detailed context of analysis.

1.4.2 *In the Broader Field of Landmark-Based Pedestrian Navigation Systems*

Beyond these publications I co-authored three further papers in the broader field of research into PNSs

Ludwig et al. (2013) provides an insight into recognition of the façade's of buildings.

Bienk, Kattenbeck, Ludwig, Müller, and Ohm (2013) details an experiment about the ease of self-localization depending on 3D-ego-perspective vs. 2D-allocentric views.

Ohm, Bienk, Kattenbeck, Ludwig, and Müller (2016) extends the findings of Bienk et al. (2013) by adding an indoor study on landmark presentation.

*The existing scientific concepts cover
always only a very limited part of reality,
and the other part that has not yet been
understood is infinite.*

HEISENBERG (1958)

2 SALIENCE AS A CONCEPT

2.1 INTRODUCTION

Following the line of research just described, this chapter is dedicated to the two most important concepts of this thesis: *landmarks* and *salience*. A short history of the concept of LMs provides insight into current usage of the term as a synonym to salient object (SO). The second part of the chapter provides an insight into the meaning of *salience* in different disciplines. Starting with Information Science (IS) the main aspect of frequency-dependency will become obvious. The review of *salience* in research on visual attention will further stress the importance of context. Finally, examining GIScience literature reveals the understanding of the concept as based on the local environment, the observer and the observed.

2.2 ON OBJECTS USED TO ORIENT ONESELF—MORE THAN A NOTE ON LANDMARK TERMINOLOGY

The term *salience* is closely connected to the term *landmark* (LM). Consequently, it is advisable to provide a detailed account of the understanding of LMs before the meaning of *salience* is reviewed.

2.2.1 *The Origin*

LMs as a term and concept are intrinsically tied to Lynch (1960). Aiming at a city's architecture, which is appropriate to the way people think about space, Lynch examined the elements Cognitive Maps (CMs) consist of and the spatial relations between these. He used map drawing as a main mode of investigation of CMs, thereby identifying five fundamental parts:

2 SALIENCE AS A CONCEPT

WAYS are used to move through space. They are the more recognizable the broader they are. If they function as borders, however, they are highly recognizable as well.

BORDERS are line-like elements of a city's structure which are not used to move along. Often they distinguish areas from each other as well as they keep them related closely.

AREAS are two-dimensional (2D) parts of a city. Travelers may enter as well as leave these. Any feature set defining an area relies on topical closeness. Examples of facets are structure, shapes, details, symbols, type of buildings, modes of use, inhabitants, traffic, condition of buildings, topography etc.

FOCAL POINTS force observers to make decisions whilst approaching a destination because many ways or features concur at an accessible area.

LANDMARKS are visual points of reference. The more familiar people are with a city the increasingly dense the mesh of LMs is, which they use to orientate themselves. A significant percentage of people characterize an object's suitability as LM along the following lines: Landmarks are simple, yet distinct in shape, they are identifiable from a distance and tend to contrast with their background.

According to Lynch, a combination of these elements forms *The Image of the City* (Lynch, 1960) a person has. Its main characteristics are (cf. Lynch, 1960):

1. The better known to a person, the larger and the more detailed people imagine a region.
2. Boundaries, streets and other line like features of a region (e.g. rivers) are straightened compared to reality.
3. The dominant shape of junctions is orthogonal.
4. Sub-variations of geographic directions are neglected, i.e. north, south, east, and west are used instead of south-west, north-east etc.

5. The way elements are combined and their spatial relations are volatile.

It is important to note, that Lynch's understanding of LMs shows important aspects of nowadays understanding of SOs. He notes that the visual dimensions are very important, he stresses the influence the local context has and claims identifiability from afar to be a major aspect. Yet, he refers to buildings only—a notion no longer prevalent in nowadays usage of the term.

2.2.2 *Current Usage—and Why Points of Interest and LMs Are Not Synonyms*

LMs as a concept have been prevalent in research on CMs from its early days. They play a significant role in the well-known theory of cognitive mapping proposed by Siegel and White (1975):

LANDMARK KNOWLEDGE is a state of knowledge, where people are able to recognize places they have previously been.

ROUTE KNOWLEDGE is given, when a person is able to connect LMs with a route network.

SURVEY KNOWLEDGE is reached if persons are able to include new points of reference in their existing route network connecting LMs.

Although Presson and Montello (1988) question the strict order of sequence proposed by Siegel and White (1975), the importance of LMs in the acquisition of knowledge about an environment has rarely been called into question since Siegel and White (1975). Strong empirical evidence suggests the importance of LMs for human orientation (cf. e.g. Sadeghian and Kantardzic (2008, p. 253) or Sorrows and Hirtle (1999, p. 41)). The term *landmark* was, however, picked up by a variety of different research areas as a useful metaphor during the 1990s. One important example is the use of LMs in early hypertext design research (cf. Glenn & Chignell, 1992). Landmark pages were those pages, where many links referred to. In the same line of research Sorrows and Hirtle (1999) draw analogies between navigation in real and electronic spaces. This is known to be the first publication which sets out to build models

for *salience* in research on navigation. The research on navigation and how LMs can and do add to improve navigation yields two consequences. First of all, the important distinction between two types of LMs regarding the route context has been stressed: Global landmarks are distant objects, which are visible and recognizable, although they are not part of the route (cf. Lovelace, Hegarty, & Montello, 1999). However, empirical evidence indicates that the availability of LMs along the route and at DPs is of crucial importance, e.g. reflected in decreased recognition times (cf. e.g. Janzen & C., 2010). These are referred to as *local* from a route's perspective (cf. Lovelace et al., 1999) and this thesis solely deals with this type. Overall, local LMs may play different roles in route instructions:

1. They may identify starting points and destinations (cf. Richter, 2007; Richter & Duckham, 2008).
2. They are used to ensure that navigators do not feel lost, (cf. Lovelace et al., 1999; Michon & Denis, 2001; Richter, 2007; Richter & Duckham, 2008).
3. They are employed to locate so called DPs, i.e. those points, where a change of direction may be necessary (cf. Lovelace et al., 1999; Michon & Denis, 2001; Richter, 2007; Richter & Duckham, 2008).

The focus on local LMs goes along with a weakening of the understanding of the term itself. In current research on LMs in navigation any object "that aids the user in navigating and understanding the space [. . .]" (Sorrows & Hirtle, 1999, p. 41) may be referred to as LM. The general applicability of the term includes entities in the broadest sense and of different sizes, given that they are suitable to identify a geographic location. This is reflected by the prepositions and verbs used in the corresponding route instructions (cf. table 1).

Having said this one final remark is necessary with respect to my understanding of LMs. Despite the fact that the terms LM and point of interest (POI) are often treated as synonyms, I do not agree. In my view, any POI may serve as a point of reference if and only if it has distinctive characteristics given its local environment. Several features of POIs have been revealed, which may contribute to its *salience*. The following description provides the factors named by Duckham et al.

	Prepositions	Verbs
point-like	across from at after away from before behind beside in front of near next to opposite past to to the left of to the right of to the side of toward	hit pass reach
line-like	along alongside	follow
area-like	around across in inside (of) into out of outside (of) through within without	cross enter leave

Table 1.: A list of verbs and prepositions to be used for different extents of LMs as given by Furlan et al. (2007, p. 12).

(2010, p. 38) and explains how these may contribute to the degree as to which a POI is like a LM.

PHYSICAL SIZE The larger the more suitable as a LM.

PROMINENCE The POI is the more suitable as a LM the more distinct its architecture is or the more eye-catching signs attached to it are.

DIFFERENCE FROM SURROUNDINGS The more unusual compared to other nearby POIs the better its suitability as a LM.

NIGHT-TIME VS DAY-TIME SALIENCE The less the visibility of a POI depends on natural light, the higher its suitability as LM is.

PROXIMITY TO ROAD The closer a POI is located at the road the higher its suitability as a LM is.

These characteristics show that POIs may be SOs, but they need not be. Using POIs as LMs by default, would not address the lack of data about SOs, consequently. In contrast, using POIs as LMs would add a bias to the available data, e.g. due to overrepresentation of shops in this kind of data (cf. Richter, 2013, p. 90). Moreover, “[t]he selection of points of interest does not consider the appearance to the human senses, and more generally, does not aim to support human orientation and wayfinding” (cf. Richter & Winter, 2014, p. 17).

In summary, the terms landmark and *salient object* will be used as synonyms throughout this text and local LMs will be the only ones dealt

with. Natural objects (e.g. trees, parks) as well as artificial ones (e.g. houses, fountains, park benches, toilets) may become salient and be referred to as LMs. In that sense any object, whether natural or artificial, may be referred to as LM—given two preconditions. First, they are outstanding given the environment they are located in (cf. Golledge, 1999), i.e. one or more of its features must contribute to the object's *salience*. Second, the object per se and its features must be sufficiently persistent to be successfully used in navigation instructions.

2.3 A REVIEW OF APPROPRIATE LITERATURE ON SALIENCE

Treating the terms landmark and salient object as synonyms throughout this doctoral work yields the need to discuss the meaning of *salience* in greater detail. The aims of this section are, consequently, threefold. First, I try to show that the notion of *salience* is not new to Information Science. I'm going to base my argument on a particular field, i.e. automatic summarizing. The reason to choose this research area is its importance in the larger field of IR research. A second field of research, which needs attention in order to understand *salience* is computer vision: Conklin and McDonald (1982) mention a strong correlation between the salient features of a picture and the fact that these features are mentioned in a description of the picture. Finally, the last part of this section is dedicated to important models of *salience* in GIScience. Presenting these developments in theory and research will be based on their timeline of development.

2.3.1 *In the Information Science Community*

Automatic Summarization As an Example

Within the IS-community, *salience* has played a major role in the broader area of IR. That being said, *salience* is particularly important in the field of automatic text summarization⁵, i.e. the “reductive transformation of source text to summary text through content condensation by selec-

⁵ A state-of-the-art overview of text summarizing can be found in Lloret and Palomar (2012).

tion and/or generalization on what is important in the source” (Jones, 1999, p. 2). *Saliency* is used to denote these important aspects of a source. “Saliency is a measure of the relative prominence of objects in discourse: objects with high saliency are the focus of attention; those with low saliency are at the periphery” (Boguraev & Kennedy, 1997, p. 5). Particularly within extractive approaches to automatic summarization, local is distinguished from global importance (cf. Boguraev & Kennedy, 1997). Regardless its range of importance, however, *saliency* is most often treated as a statistical problem of frequency counts (cf. e.g. Mani & Maybury, 1999)—a notion prevalent in the field of IR as a whole (cf. Park & Li, 2006)⁶. *Saliency* may be attributed to both, terms as well as sentences—and may be reinforced mutually: If a term or its anaphorae occur(s) in many highly salient sentences, the term is regarded to be salient. Conversely, if a sentence consists of many highly salient terms its saliency score increases (cf. Zha, 2002). Let me give two examples: In the broader field of IR, Lioma, Blanco, and Moens (2009) use *saliency* as a means of query expansion. In this case social tags are used to expand user queries. Tags are selected by means of a frequency based measure to determine a tag’s *saliency*. A second example is taken from the field of automatic summarization. Dunietz and Gillick (2014) use the notion of *saliency* to detect entities. Given a single document, they try to assign a *saliency* score reflecting their respective prominence to each of the entities mentioned in a given document. The approach employs binary logistic regression by using the location in the text where an entity is mentioned as one of its input parameters. A further example which relies even more on the structure of the document than on the content is provided by Sándor and de Waard (2012). They exploit rhetorical *saliency* within a highly specialized domain of discourse, namely biomedical articles. They propose two kinds of manifestations of *rhetorical saliency*, both of which are equally important in terms of identifying gained knowledge. One is given by the overall structure of the paper, i.e. findings are more likely to occur in the results section than in the introduction. The second is based on metadiscourse, i.e. phrases indicating results rather than assumptions.

⁶ In fact, this is the reason, why automatic summarization is suitable to be used as a *pars pro toto* for the concept of *saliency* in IR.

Regardless the specifics of any of the attempts, the term *salience* is used as a synonym to importance in automatic summarization. Overall, three distinct features can be derived:

1. *Salience* is often based on term frequency counts, i.e. the more often a term occurs in a document the more salient it is.
2. *Salience* is attributed not inherent, i.e. no entity, object, word, sentence etc. is *eo ipso* salient.
3. *Salience* depends on context, i.e. the surroundings of a sentence or an entity are of major importance to yield *salience*.

As we shall see, the importance of frequency is different in GIScience-literature, where *salience* is given based on rarity of a feature in a given neighborhood. However, 2 and 3 are most important to the concept of *visual salience* as used throughout this thesis. In fact they are in computer vision research, as well.

2.3.2 *In Computer Vision*

Although computer vision is not part of the narrower field of IS-research, the way *visual salience* is used in this field is worthy of examination. As mentioned in the introduction to this section, as early as 1982, Conklin and McDonald use *salience* in pictures to determine the facts and the ordering of facts in an textual description of the picture. Empirical results indicate a strong correlation between an object's salience in a picture and the likeliness of appearance in a text produced by a person to describe the picture (cf. Conklin & McDonald, 1982). In general, *visual salience* computations are closely related to visual attention research. Models of *salience* in computer vision research aim to mimic human fixations, i.e. these models try to predict fixations of humans viewing an image for the first time (cf. e.g. Zhang et al. (2008) or Goferman et al. (2012, p. 1915)).

What Does Visual Salience Mean?

The way *salience* values are calculated for images is strongly related to the way the study of attention in neuroscience has evolved. Itti (2007)

describes the framework of bottom-up *saliency* calculation in human brains using a three-step-model:

1. Visual neurons, which are sensitive for low-level visual features, compute so-called feature maps in parallel.
2. The feature maps are subjected to a process of highlighting, i.e. any location different from its local surroundings is highlighted.
3. All resulting highlighted areas are combined to build a single saliency map per image.

Accordingly, the term *saliency* is used to denote local changes in basic visual features, which stresses the importance of local context (cf. Q. Zhao & Koch, 2013). This means high *saliency* values at a given location yield a high probability of a fixation at this area (cf. Itti & Koch, 2000; Parkhurst et al., 2002; Zhang et al., 2008). In their highly renowned (cf. Valuch et al., 2013) work, Itti, Koch, and Niebur (1998) develop a model for *saliency* computation⁷ based on local contrasts of color, intensity, and orientation. However, the list of features to be used as model input, are still subject to discussion (cf. Q. Zhao & Koch, 2013), despite the fact that the importance of *saliency* for cognitive processes has been known for many years (cf. Conklin & McDonald, 1982). Itti (2007) points to a review by Wolfe and Horowitz (2004), which presents research outcomes regarding the prevalent features of *visual saliency* in computer vision:

UNDOUBTFUL colour / motion / orientation / size (including length and spatial frequency)

PROBABLE luminance onset / luminance polarity / vernier offset / stereoscopic tilt and depth / pictorial depth cues / shape / line termination / closure / topological status / curvature

POSSIBLE lighting direction (shading) / glossiness / expansion / number / aspect ratio

DOUBTFUL novelty / letter identity / alphanumeric category

⁷ nota bene (NB:) This model is able to predict fixations in free viewing tasks but not in visual search tasks (cf. Niu, Todd, Kyan, & Anderson, 2012).

PROBABLE NON-ATTRIBUTES intersection / optic flow / color change /
three-dimensional volumes / faces / your name / semantic category

The ongoing discussion results from the fact that the models are based on numerous assumptions and parameters, with values ascribed in an ad-hoc manner. With the further raise of eyetracking technology data driven models may be derived using highly specialized machine learning algorithms. Additionally to the vast amount of research on low-level features adding to *visual salience* the development of computational models of *salience* was fostered recently by including global high-level features. For example, Goferman et al. (2012) remain using low-level features (e.g. colour), but include high-level features (e.g. face recognition) as well. Additionally, the frequency of features is taken into account globally, i.e. less frequent features are acknowledged more than often occurring ones. Finally, their algorithm takes rules of visual organization of images into account (e.g. center-surroundings).

Summarizing the Principal Insights

Overall, the most important insights about the nature of *visual salience* in computer vision, which are to be derived from this discussion, are:

1. A given visual stimulus is not salient per se but *salience* is ascribed to a visual stimulus given a set of other stimuli. "It is important to remember that salience is the consequence of an interaction of a stimulus with other stimuli, as well as with a visual system" (Itti, 2007, no pagination given). Context (i.e. local feature change) is of major importance in this attribution of *visual salience*.
2. The experienced *visual salience* is supposed to be comparable across individuals. This is a consequence of the "fairly low-level and stereotypical computations in the early stages of visual processing" (Itti, 2007, no pagination given) the saliency maps result from. This is an important fact reinforcing the priority of *visual salience*.

Taken as a whole, these aspects are very similar to those found for IS-research. Both fields share the view that *salience* is attributed and

depends on local context. These aspects will play an important role in the understanding of the concepts throughout this doctoral work.

2.3.3 *In the GIScience-Community*

The Early Days—and an Influential Model

Beyond Lynch (1960) the earliest empirical attempt to gain insight into the factors, which contribute to a building's *saliency*, dates back to Appleyard (1969). Participants were asked to provide a list of buildings located in their home town they remembered best. A subsequent qualitative analysis revealed form, visibility, and meaning as key features which make buildings memorable. In fact, these are three kinds of features still used in the predominant model of *saliency* in GIScience research. Beyond those early studies⁸ regarding the features contributing to *saliency* it was only around the turn of the century that the concept was formalized. Five papers, published between 1999 and 2005 build the nucleus of the work done. The initial seed was sowed by Sorrows and Hirtle (1999, pp. 45–46), who identify features of LM in “Real and Electronic Spaces” (Sorrows & Hirtle, 1999, p. 37). They distinguish four dimensions contributing to *saliency*: visual, structural, cognitive aspects, and prototypicality, thereby drawing heavily on Rosch et al. (1976) to derive the latter. Based on the work by Sorrows and Hirtle (1999), Raubal and Winter (2002) introduce a formal model providing measures for each of the three measures.

VISUAL ATTRACTION “Our formal model of landmark saliency includes four measures regarding visual attraction: Facade area, shape, color, and visibility” (Raubal & Winter, 2002, p. 249).

SEMANTIC MEASURES “[...] comprise cultural and historical importance of an object, and explicit marks” (Raubal & Winter, 2002, p. 249).

⁸ Refer to Sadeghian and Kantardzic (2008) or Duckham et al. (2010) for similar yet differently focused reviews of literature on *saliency*.

STRUCTURAL ATTRACTION “A landmark is structurally attractive if it plays a major role or has a prominent location in the structure of the spatial environment” (Raubal & Winter, 2002, p. 250).

The measures provided for each of these dimensions are subject to hypothesis tests in order to ensure a significant difference given the local surroundings of the feature. A significant difference yields a value of $s = 1$ for each of the features used for a given dimension and $s = 0$ in all other cases. Consequently, the local environment is most important in this model and it shares this importance with the non formal account in Sorrows and Hirtle (1999). All significance values belonging to a dimension are summed up and multiplied by a weight factor which may be used to account for personal preferences or modality issues. A linear combination of these sums is used as a global *salience* value. Raubal and Winter (2002, p. 250) neglect the aspect of prototypicality due to the demanding task of large-scale human subject testing necessary to acquire a sufficiently large dataset. Furthermore, they substitute the term *cognitive* by *semantic*. The authors, however, do not provide a detailed explanation for this substitution. The model presented by Raubal and Winter (2002) proved useful in Nothegger et al. (2004), where *visual salience* and *semantic salience* values of façades were derived. These values were useful for distinguishing between different geographical objects. Two major extensions were made to enhance the model by Raubal and Winter (2002): Winter (2003) adds the notion of visibility in advance as contributing to a LM’s *salience*, thereby stressing the importance of the route. This importance is further accounted for by Klippel and Winter (2005), who give a very detailed account of *structural salience*, and, in doing so, change the meaning proposed by Raubal and Winter (2002). Klippel and Winter define: “Objects are called structurally salient if their location is cognitively or linguistically easy to conceptualize in route directions” (Klippel & Winter, 2005, p. 347). As fig. 2 shows, their taxonomy of spatial relations between LMs and a given route pays special attention to DPs. Klippel and Winter (2005, p. 350) provide a detailed account of spatial configurations where LMs may occur. In a first step they distinguish between off-route and on-route LMs. The latter may either be located at a DP or non-DP. Any DP may either make a direction change necessary or not. In both cases, LMs located at DPs may be passed

2.3 A REVIEW OF APPROPRIATE LITERATURE ON SALIENCE

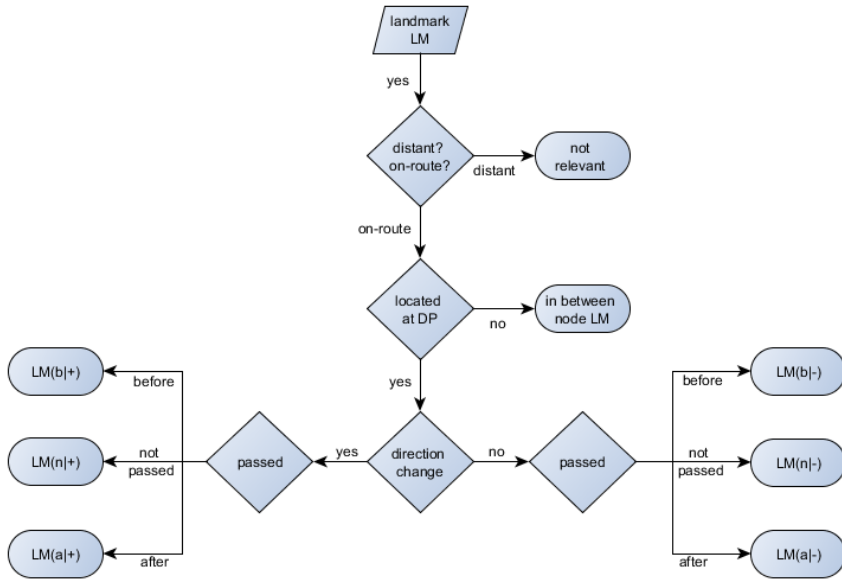


Figure 2.: The different types of LMs based on their location in relation to a DP. The figure is based on Klippel and Winter (2005, p. 350) and was drawn using yEd (cf. yWorks GmbH, 2016).

before or after the direction change. Furthermore, the LM may not be passed at all. Clearly, the location of the LM on a route is most important to *structural salience*. This is particularly true as a route may be seen as a sequence of DPs where direction changes are necessary or not.

Studies with diverse goals draw on the model described, often focusing on specific parts of the model. For example, Winter, Raubal, and Nothegger (2005) focus on the problem of model adaptation to particular contexts. They introduce weight factors for calculating an overall *salience* value as proposed in Raubal and Winter (2002). Furthermore some studies aim towards the automatic assessment of *structural salience* (cf. e.g. Claramunt, 2007; Sadeghian & Kantardzic, 2008). Several studies (cf. e.g. Ishikawa & Nakamura, 2012; Schroder et al., 2011; Sefelin et al., 2005) refer to the model when conducting empirical studies on features, which are important to make an object salient. Despite a vast amount of different factors found in these studies (cf. Chapter 3), there

are commonalities between them: The greater the figure-background contrast of a geographical object, i.e. the more unique one of its features is given its local environment, the more salient an object is. This means, the importance of local context is reinforced by these empirical studies. Further research work based on the model includes but is not restricted to Hansen et al. (2006), Klippel et al. (2009), Fang et al. (2012). Klippel et al. (2009) is concerned with spatial chunking, i.e. a framework for incorporating LMs in cognitively ergonomic route directions based on their *salience*. Hansen et al. (2006) is dedicated to the development of a data structure based on OpenLS, which allows the incorporation of all aspects relevant to the model. Similarly, Fang et al. (2012) provide another Geographic Information Systems (GIS) data model to incorporate LMs. Overall, the model introduced in Raubal and Winter (2002) represents the dominant way of thinking about *salience*.

The Trilateral Relationship—and Another Change in Terminology

The model by Raubal and Winter (2002) was substantially refined by the doctoral work presented in Caduff (2007) and Caduff and Timpf (2008). The key idea of this refinement is the fact that no object is salient *eo ipso*. This means, Caduff and Timpf (2008) stress the importance of context by focusing on the interaction between observer, observed, and surroundings:

“In a navigation context, hence, salience of geographic objects is a property of the trilateral relationship between observer, environment and geographic object. [...] The overall salience of geographic features is defined as a three-valued vector, whereby the components capture perceptual, cognitive, and contextual aspects of geographic objects”(Caduff & Timpf, 2008, p. 264).

Based on this understanding Caduff (2007) proposes a Bayesian network for computing *salience* values which is largely based on visual attention research. The suggested components of the network are (cf. Caduff, 2007, p. 28):

2.3 A REVIEW OF APPROPRIATE LITERATURE ON SALIENCE

EXOGENOUS ALLOCATION OF ATTENTION describes the subconscious way attention is directed to other objects. This means, the intentions of observers are irrelevant, i.e. the features become salient by means of physiological processes beyond conscious control of human beings. Hence, a person's "attention is automatically directed towards these. [...] We will use the term *Perceptual Salience* to refer to effects of attentional capture on a feature's salience" (Caduff, 2007, p. 29).

ENDOGENOUS ALLOCATION OF ATTENTION is meant to be the kind of attention which is intentionally directed to particular features by a person. This means, it is the top-down counterpart to exogenous allocation of attention. Humans are able to consciously direct their attention to features by drawing on prior knowledge about their usefulness to fulfill the current navigational task. "[W]e will use the term *Cognitive Salience* to refer to the endogenous factors that influence salience" (Caduff, 2007, p. 29).

DEPLOYMENT OF ATTENTION is based on the current context. This means, whether someone is driving a car or walking has a major impact on the amount of attention which can be directed to stimuli in search for *salience*. Therefore, this part of the vector is called *contextual salience*.

In summary, Caduff derives *perceptual*, *cognitive*, and *contextual* salience as components of *overall salience*. Figure 3 provides a detailed account of the Bayesian network derived by Caduff (2007, pp. 29–32). As Bayesian networks are used to represent causal networks in general, the model can describe causal relationships between all types of components Caduff uses to model *salience*. He introduces several auxiliary components.

DEGREE OF RECOGNITION "provides a measure how easily the corresponding spatial object can be identified" (Caduff, 2007, p. 54).

IDIOSYNCRATIC RELEVANCE "depends on the amount of knowledge and personal experience navigators associate with a specific number of observations" (Caduff, 2007, p. 55).

2 SALIENCE AS A CONCEPT

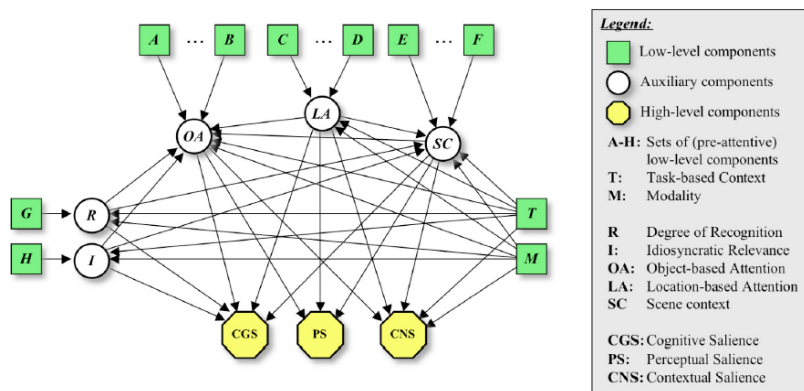


Figure 13 The structure of the Bayesian network used for simulating the salience assessment process. The low-level components are derived directly from input data and serve as evidence. The auxiliary components account for the different types of attention, and the high-level components describe the resulting saliency of the observed spatial object.

Figure 3.: The Bayesian network described and implemented by Caduff (2007, p. 30). For the sake of clarity, the explanation of the original publication is given.

OBJECT-BASED ATTENTION “defines the salience of single objects or groups of objects contained in a scene” (Caduff, 2007, p. 21).

LOCATION-BASED ATTENTION “assesses the potential for attraction of attention of regions across spatial scenes” (Caduff, 2007, p. 20).

SCENE CONTEXT “defines the structure and global semantic characteristics of the scene” (Caduff, 2007, p. 20).

The way that auxiliary and low-level components interact is based on the way that humans process information. With a strong focus on perceptual features only, Caduff bases his whole model and its computation on attention research. Consequently, Caduff (2007) gives a detailed account of the computational steps involved. Using a digital image of a spatial scene, he shows how to find values for these low-level components.

Important Insights

Taken as a whole, four important insights can be derived from the understanding of *salience* in GIScience.

1. The use of the term *cognitive* in favor of *semantic* draws attention to the fact that endogenous allocation of attention is driven by aspects beyond signs and cultural/historical importance. The understanding of *cognitive salience* as presented in Caduff and Timpf (2008) is close to that of Sorrows and Hirtle (1999). It is important to note that their understanding is not very different from Raubal and Winter (2002), who explicitly refer to Sorrows and Hirtle (1999): "Our notion of semantic attraction is similar to that of cognitive attraction [. . . (Sorrows & Hirtle, 1999)], which focuses on the meaning of a feature" (Raubal & Winter, 2002, p. 249). Motivated by these facts, I will use the term *cognitive salience* throughout this thesis.
2. The *overall salience* of objects has been revealed to be a multidimensional construct. At least five different sub-dimensions of *overall salience* must be acknowledged. Visual features, route-dependent (structural) and cognitive aspects are common to the principal models. Furthermore, prototypicality is known to be important (cf. Sorrows & Hirtle, 1999) but at the same time difficult to capture (cf. Raubal & Winter, 2002). The fifth dimension is, *visibility in advance* as introduced by Winter (2003).
3. Context was revealed to have a major influence on *salience*. This holds true from two different angles. First, the navigational mode, i.e. whether a person drives a car or walks, must be acknowledged. Second, the local environment in which the object is embedded is highly important.
4. *Salience* is not inherent to an object but a function of the object itself, the observer and the local surroundings of the particular object. This is a major reason for the importance of conducting in-situ experiments (cf. section 3.2.3). To date, no study has tried to derive a set of survey items suitable to measure all of the dimensions of *salience* in an in-situ manner.

2.4 TOWARDS A PRELIMINARY DEFINITION OF SALIENCE

The presented review reveals two major properties of *salience* various disciplines agree upon:

1. *Salience* is not an inherent physical property of a geographical object, but rather a property ascribed to an object by an observer.
2. *Salience* is dependent on the context, i.e. the local environment plays an important role in whether an object becomes salient or not.

However, there are several aspects with respect to which major differences exist across different fields. The focus in GIScience related research is on the nature of *salience* in pedestrian navigation for urban settings. This yields a major difference in terms of dimensionality, i.e. in GIScience *salience* is a multidimensional construct incorporating route dependent aspects as well as visual ones. In contrast, computer vision research deals solely with visual aspects. Moreover, whereas a rule like “the more frequent the more salient” holds more or less true in automatic summarizing research, the opposite is true in research on *salience* from a GIScience perspective. Here, the uniqueness of an object given all of its features in the local neighborhood is of major importance for the object’s *overall salience*. Overall, *salience* is synonym to prominence in a given local neighborhood in GIScience as well as in automatic summarizing research. It is, however, based on different interpretations of the impact the frequency has.

Based on these differences and commonalities, I can give the following operational definition of *salience*.

Salience

Given a local environment an observer is in, *salience* is the degree to which an object, persistent enough to be used in route instructions, draws the average observer’s attention. This degree is evoked by

1. visual features the objects has (*visual salience*),
2. the degree of prototypicality it shows (*prototypicality*),

3. how identifiable it is when approached (*visibility in advance*),
4. the ease with which it may be integrated into a route description (*structural salience*) and
5. the degree as to which it can evoke prior knowledge about the object (*cognitive salience*).

This definition has major consequences for this doctoral work.

1. Any type of object, persistent enough to be used in a route description, must be included in the sample, i.e. e.g. buildings as well as post boxes. As a consequence, the objects to be rated during the experiment must be fixed apriori because prior empirical evidence suggests that participants tend to neglect the prerequisite of temporal persistence although they were explicitly made aware of this requirement (cf. e.g. Ishikawa & Nakamura, 2012).
2. The empirical method taken has to make sure that the observer is able to discern an object's surroundings. This is, among others (cf. section 3.2.4), one of the reasons why an in-situ setting is chosen to conduct the experiment.

Finally, a note on the way *visibility in advance* (cf. Winter, 2003) as a sub-dimension of *overall salience* is understood seems advisable. It stems from a counterargument given by Schroder et al. (2011). They dismiss *visibility in advance* as a dimension based on the argument that only visible stimuli can be used to guide a person (cf. Schroder et al., 2011, p. 434). While this is certainly true (cf. Xia et al. (2011), who reveal visibility to be the most important aspect of *salience*) I still think visibility in advance has an impact on *salience* if it is understood correctly as recognizability from afar. Bidwell et al. (2005) found strong evidence that LMs in unknown environments must be first and foremost recognizable, a property that relies mostly on the visual features in a given context. Imagine a Cathedral located next to a much smaller building with a red-colored façade in a neighborhood of white-washed houses. Clearly, like any other object, neither the Cathedral nor the red house,

have any *visual salience* at all, unless they are visible from the direction an observer approaches the scene. But, the fact that the Cathedral will be identifiable from a greater distance contributes to its *salience* beyond its *visual salience*, despite that fact that both are related (cf. Miller and Carlson (2011) who report on the strong influence *visual salience* has on object recognition). Therefore, *visibility in advance* must be included in the structural model leading to *overall salience*, as objects that “are identifiable early on along a route are more useful than those that can only be spotted at the very last moment” (Richter & Winter, 2014, p. 142).

2.5 CHAPTER SUMMARY

This chapter has reviewed prominent research on the two major concepts of this doctoral work, landmarks and *salience*. Starting with a short history of the term LM the chapter continued to review the understanding of *salience* in IS, computer vision and GIScience research. The review has resulted in an operational definition of *salience*. The definition will be used throughout the remainder of this thesis.

USE OF LANDMARKS AND SALIENT OBJECTS AS SYNONYMS

The history of the term landmark is closely related to Kevin Lynch’s seminal work on the imageability of a city (cf. Lynch, 1960). He identified LMs to be one of four major elements structuring a city. Although Lynch mentions the existence of global and local LMs, the current usage of the term in research on pedestrian navigation is almost without exception local. This involves a shift of understanding, treating LMs and SOs as synonyms. Given this history the two terms will be used interchangeably throughout this thesis and only local LMs will be considered.

SALIENCE IN IS

In IS research *salience* is particularly important in the field of automatic summarizing. In this domain it is the local context of the document/sentences that make an entity salient. If an entity occurs very often in highly salient sentences these render the entity itself salient.

SALIENCE IN COMPUTER VISION

Computer vision research tries to mimic the way the human visual system calculates *salience*. Local context is a major factor in these computations as *salience* does not exist eo ipso.

SALIENCE IN GISCIENCE

Major models of *salience* were built around the turn of the century, including the description of ways *salience* may be calculated. The sub-dimensions of *salience* identified include visual features, prototypicality, cognitive aspects, structural (i.e. route dependent aspects) and visibility in advance. A major enhancement of these models was the understanding of *salience* as attributed to an object by an observer given the local environment.

SALIENCE IN THIS DOCTORAL WORK

An object in a given environment is said to be salient if it draws the attention of an observer. As a precondition only those objects are considered which have a minimum of temporal persistence to be useful in route instructions. The degree as to which the observer's attention is drawn has five sub-dimensions which are referred to as *visual saliency*, *structural saliency*, *cognitive saliency*, *prototypicality* and *visibility in advance*.

3

HOW CAN SALIENCE BE MEASURED?

3.1 INTRODUCTION

The first part of the chapter deals with experimental design. By summarizing the approaches taken in the literature and discussing the advantages and disadvantages of each design an argument is built for the need to conduct a large-scale in-situ study which takes different types of objects into account. The second part of this chapter is concerned with data analysis and modeling. It motivates the use of Structural Equation Models (SEMs) as means to measure and understand the components contributing to *salience*. SEMs are introduced and an overview of commonalities and differences of Partial Least Squares Path Modeling (PLS-PM) and Covariance-based Structural Equation Modeling (CB-SEM) is given. Having justified the use of PLS-PM, a detailed account of its estimation procedure is given. The third part of the chapter details the measurement model used to assess *salience*. Questions for all LVs will be presented based on reviews of earlier studies into dimensions adding to an object's *salience*. A focus will be given to visual dimensions as they are particularly widespread. Finally, the results of a pre-study on proper phrasing of the survey questions will be presented.

3.2 A REVIEW OF METHODS PROPOSED IN THE LITERATURE

In order to provide a reason for the use of an in-situ experiment to test the fit of SEMs a short review of work done so far is necessary. The research reported focuses on pedestrian navigation in urban settings, which has been the preferred mode and terrain in scientific publications regarding LMs from the early days on. Only recently, contexts beyond cities are investigated (cf. e.g. Kettunen et al., 2013). Moreover, studies concerning the use of LMs in car navigation systems exist (cf. e.g. Burnett,

1998; Burnett et al., 2001). Whereas similar overviews of LM identification approaches for LMs are given elsewhere (cf. e.g. Duckham et al., 2010), the focus of the following review is on experimental designs. Basically, three different approaches may be distinguished: automatic approaches relying on existing databases, lab-based approaches and in-situ-experiments are used to assess *salience*. I will provide the prototypical steps of each of these methods accompanied by a selection of citations implementing it. The overview aims to show two major disadvantages among others:

1. Current attempts do not take non-buildings as objects into account.
2. To date, no study which builds a coherent model by taking prior dimensions systematically into account has been published.

3.2.1 *Automatic Assessment*

In their recent book on LMs Richter and Winter (2014) mention two data sources for automatic assessment of local LMs (cf. also Sadeghian and Kantardzic (2008) for remarks on automatic landmark detection).

GEOGRAPHIC DATA can be exploited to determine LMs by using the structure of its attributes.

(GEOGRAPHIC) INFORMATION RETRIEVAL is used to mine LMs from unstructured mostly textual or pictorial data.

The most notable line of research on exploitation of geographic data is represented by the doctoral work of Elias (2006). Drawing on official geographical data sources she derives 18 attributes. They cover a very broad range of aspects, including building type, function, number of edges, number of direct neighbors etc. Based on these measures, Elias uses a modified form of Iterative Dichotomiser 3 (ID3) (cf. Quinlan, 1986) to find LM-candidates for a given route. This machine learning method is based on two well-known measures, information entropy and information gain.

Assume a source of symbols S the following assumptions hold for:

1. The number n of symbols s_i , which may be produced by the source S , is known.
2. A fixed probability p_i for each symbol s_i is given.
3. The symbols are independent, i.e. $P(s_i|s_{i-1}) = P(s_i) = p_i$

Given these conditions, the self-information of the occurrence of a symbol s_i is given as:

$$I(s_i) = \log\left(\frac{1}{p_i}\right) = -\log(p_i) \quad (1)$$

Equation (1) describes the amount of information known, once realization s_i of random variable S has occurred, i.e. the smaller its probability p_i the higher the self-information of event s_i is.

The entropy of any such source is then given by the renown equation provided by Shannon (1948, p. 11).

$$H(S) = K \cdot \sum_{s_i} p_i \cdot I(s_i) \quad (2)$$

The degree as to which $H(S)$ decreases if a symbol s_{n+1} was added is known as the information gain achieved. Although classification results were plausible to users, human suggestions showed major differences to those calculated. This incompatibility may be due to neglecting possibly relevant attributes (cf. Elias, 2006, pp. 123–124), which were not available in the existing databases. Furthermore, it is important to note that only buildings and only intersections were used throughout this study which diminishes the generalizability.

Another noteworthy approach to gaining LM-knowledge from geographic properties is the POI-based strategy employed by Duckham et al. (2010). They use data provided by the Yellow Pages to identify POIs, which may serve as LMs. In difference to other approaches, they aim to derive the degree of landmarkness for classes of POIs instead of drawing on specific features of individual objects.

“[There are] two independent factors to consider:”

- (i) “How suitable a typical instance of a POI category is as a landmark; and”
- (ii) “How likely it is that a particular instance of a POI category is typical” (cf. Duckham et al., 2010, p. 33).

They assessed the values of their results based on 24 different routes in Melbourne, but did not do any human subject testing. In a sense the method described by Duckham et al. (2010) is a bridge to methods used in geographic information retrieval. Richter and Winter (2014, pp. 146–151) name two different general approaches, exploiting textual vs. pictorial data and a combination of these. Tezuka and Tanaka (2005) reveal that spatial context plays a crucial role when LMs are extracted from web documents. If text mining methods take spatial context into account the results achieved are improved significantly. Mummidi and Krumm (2008) try to find clusters of pins potentially denoting a single POI in Microsoft Bing Maps, thereby drawing on Term Frequency Inverse Document Frequency (TFIDF) measures for n-grams of pin annotations. In line with the approach of combining textual and pictorial data, several approaches try to combine data from tags, geographic position and pictures available on public data sources in order to find LMs (cf. e.g. Kennedy & Naaman, 2008). According to Richter and Winter (2014, p. 149) the most sophisticated of these approaches is presented by Crandall et al. (2009). They make use of the “scale of observation, i.e. the fact that on different levels of scale different effects are observable” (cf. Richter & Winter, 2014, p. 149). Crandall et al. (2009) build clusters of locations where photos were taken and derive labels for these by means of TFIDF for this cluster’s tags. Having found these they try to find a canonical image for this particular LM by using a similarity measure of images.

Overall, promising approaches to automatic detection of LMs have been recently proposed. All of them are based on available data sources and require no user-involved empirical effort. Moreover, data on LMs derived from these sources covers large geographical spaces. However, not all of the approaches are free of manual effort. More severely, none of the approaches is able to gain an insight into which features are preferred by human observers.

3.2.2 *Lab-Based Experiments*

Three different types of lab-based experiments are prevalent in research on LMs and their *salience*: Virtual Reality (VR), Cognitive Map (CM), and computer based experiments.

Virtual Reality

VR experiments are run in a similar fashion to their in-situ counterparts (cf. section 3.2.3).

1. Define an (extremely) low-stimulus route in a VR environment using (geometrical shapes as) LMs.
2. Display a video of the route to the participants.
3. Once they have seen the whole route, they are asked to rewalk the route in VR.

Studies using the VR-method include e.g. Brunyé et al. (2012), Hidayetoglu et al. (2012), and Peters, Wu, and Winter (2010). VR experiments allow for full control of all stimuli presented, leading to a very high degree of repeatability. However, due to the artificial environment it is questionable as to whether the choices of the stimuli can be transferred to real-world scenarios.

Cognitive Mapping

A second kind of lab-based experiments is based on the use of CMs. The setup is typically as follows:

1. Participants with detailed knowledge of a certain place are chosen.
2. Both, origin and destination of a route are provided to participants.
3. A fictitious person *E* is described to participants, including *E*'s local knowledge. Participants are then asked to provide directions to *E*.

4. Participants may either draw a map to provide the directions or describe the route in written form.

Studies using this method encompass e.g. Daniel, Przytula, and Denis (2009); Denis (1997); Denis et al. (1999); Michon and Denis (2001). This kind of experiment has a major advantage: Minor effort is necessary to conduct CM experiments. Participants are, however, required to remember features; but, there is evidence challenging the idea that LMs, which are easy to memorize are comparably useful in navigational tasks (cf. Miller & Carlson, 2011).

Computer based

Finally, a third lab-based approach uses computers.

1. Photos, maps etc. are presented to participants on a computer monitor.
2. Participants are either asked
 - a) to identify LMs on the screen (sometimes accompanied by the request to provide a name for the LM), or
 - b) to undergo an eye-tracking study in which high probability of fixation areas are identified based on eye movement.

Several studies use this method to reveal LMs and the features contributing to their salience (cf. e.g. Fontaine et al., 2005; Schroder et al., 2011; Snowdon & Kray, 2009). Clearly, these experiments are both, easy to conduct and easy to evaluate. They are in fact a good method to come up with an initial choice of LM candidates in high stimuli areas. However, the disadvantage is the difference between a person's field of vision in-situ and the chosen picture detail. This may have a major impact on *salience* estimations, because of the known dependency of *salience* on the local environment (cf. e.g. Partala et al., 2010).

3.2.3 *In-Situ-Experiments*

In-situ experiments typically comprise the following steps.

1. Choose one or more routes.
2. Experimenter and participants walk along the route, together. Participants are asked to choose LMs and provide reasons for their choice whilst walking.
3. During a second walk, participants are asked to describe the route to a fictitious person.

The in-situ method is used regularly, e.g. by Iachini and Logie (2003); Ishikawa and Nakamura (2012); Kettunen et al. (2013); Michon and Denis (2001). This kind of experiment represents the gold standard for data acquisition on everyday use and selection of LMs because these experiments are as close as possible to everyday navigation scenarios. Moreover, they provide an insight as to what extent a spatial configuration leads to difficulties regarding orientation. That being said, such studies come with considerable cost, both in performing and evaluating the experiment.

3.2.4 *Criticisms and Their Impact on Empirical Design*

Given these general approaches a number of disadvantages regarding a variety of aspects become obvious⁹.

NUMBER OF PARTICIPANTS i.e. very small samples are reported in studies, including e.g. only $n = 20$ participants like Ishikawa and Nakamura (2012) do.

NUMBER OF OBJECTS AND ROUTES i.e. objects and routes used in studies often lack scalability. The reasons for choosing certain routes of buildings are often not even mentioned (cf. e.g. Nothegger et al., 2004). Moreover, the number of routes and objects alike is often very small, too. For example, Nothegger et al. (2004) use

⁹ Some of these aspects were already published in Kattenbeck (2015a).

one intersection and nine objects to empirically test the viability of the salience model derived by Raubal and Winter (2002).

NON-BUILDINGS To date, empirical studies and theories often ignore that theoretical claims as well as empirical evidence for salient objects other than buildings exist (cf. Sadeghian & Kantardzic, 2008). Ishikawa and Nakamura (2012) and Schroder et al. (2011), however, are noteworthy exceptions. The authors of both studies explicitly include non-buildings in their experimental setups.

WEAK THEORETICAL GROUNDING Empirical studies on *salience* estimations make a reference to Raubal and Winter (2002) regularly. However, the focus of these studies is often not on theory-testing in a narrower sense. This is particularly true for in-situ studies which often relate to other empirical studies but do not make systematic use of the dimensions reported earlier (cf. e.g. Schroder et al., 2011). To date, none of the known studies has performed a large scale empirical evaluation of the published theories based on the *salience* dimensions previously found.

NO REAL WORLD LIKE DATA ACQUISITION None of the discussed studies uses a method of data acquisition which would be applicable in real-world scenarios. This is particularly true for VR approaches due to their very nature. However, it remains a matter of fact even in in-situ studies, where think-aloud methods or several rewalks of a route are used. As Duckham et al. (2010, p. 28) put it: “[P]rocedures for identifying landmarks are not yet tested in practice, and hence not readily available”.

The methodological setup in this doctoral work has been designed to address these issues. By taking these problems into consideration I justify my choices below.

THEORETICAL GROUNDING THROUGH MODEL BUILDING

Structural models proposing connections between the different constructs based on the theories are built (cf. section 5.3). The indicators used in the measurement models (cf. section 3.3.3) are derived from a thorough literature review, emphasizing visual

3.2 A REVIEW OF METHODS PROPOSED IN THE LITERATURE

dimensions which the literature suggests to be particularly important.

INCLUSION OF BUILDINGS AND NON-BUILDINGS

The study will be conducted in an urban environment. A consequence, thereof, is that approximately (approx.) two thirds of objects are buildings (cf. section 5.2.2). The remaining third are non-buildings (cf. section 4.2.2 for details on how objects were sampled). It is important to note that intersections are not objects in the sense of this thesis. From my perspective, it cannot be avoided to exclude intersections as most of the MVs used for *visual salience* and *cognitive salience* are not applicable to these.

NUMBER OF OBJECTS, ROUTES AND PARTICIPANTS

A suitable sample size of objects was determined a priori (cf. section 4.2.1) and objects were randomly chosen based on geographical coordinates (cf. section 4.2.2). This means, the objects, which ought to be rated by participants, were predefined in order to ensure data suitable for statistical analysis. If people were allowed to use any object for route instructions, a diverse range of objects would have been used, which would likely result in many unique references (cf. Ishikawa & Nakamura, 2012). From a survey-based measurement perspective this is undesirable because of the random error inflation induced by single ratings. The sample of random objects was randomly split into tuples, where each tuple was rated on a single route (cf. section 4.2.3). To counteract measurement bias, each route was walked twice by two different participants. This means, a large set of participants is used in order to ensure that weights are generalizable in terms of both objects and participants. Moreover, no participant rated objects on two routes (cf. section 4.4.1).

REAL-WORLD LIKE THROUGH IN-SITU DATA ACQUISITION

Although in-situ experiments have clear drawbacks with respect to effort in time and participants, an in-situ approach is taken throughout this thesis. There are several reasons to do this:

1. There is a known bias of recognizability compared to memorizability (cf. e.g. Michon & Denis, 2001). However, the

salience of an object must render the object recognizable in order to be of value for route instructions. Hence, cognitive mapping experiments are likely to provide invalid results to test the importance different dimensions have in predicting *salience*.

2. “In building recognition, the location, nearby environment, and building shapes and landmarks also become important” (Partala et al., 2010, p. 431). This claim reinforces the importance of the local context stressed by Caduff and Timpf (2008). However, if a picture-based data acquisition method would have been used a bias would result when *salience* ratings are based on these because pictures can only show a particular fragment of the environment.
3. Miller and Carlson (2011) argue for preferences in probands to use spatial features if the task focuses on navigation—in opposition to visual aspects used, if they are asked to recognize certain objects. Presenting the stimulus on a picture in the in-situ experiment (cf. section 4.3.1), whilst participants have to walk a route will provide a sound balance of both aspects.

Overall, leading participants on a route and asking them to rate objects using a survey is as close as possible to real-world crowd-sourcing scenarios. Therefore, I acknowledge the desire for test data acquisition methods in real-world scenarios (cf. Duckham et al., 2010). Moreover, a step towards the crowd-sourced data acquisition of landmark knowledge claimed by Richter (2013) is done (cf. section 7.2.3).

3.3 INTRODUCING STRUCTURAL EQUATION MODELING, PLS PATH MODELING, AND A MEASUREMENT MODEL FOR SALIENCE

The previous section outlined major design decisions for the empirical study, which will be used to collect data. This section will now go into detail of measuring and analyzing. It provides a rationale for using Structural Equation Models (SEMs) as a mode of analysis for the data collected. Richter (2013) points to a major research gap in terms of the

empirical assessment of the influence each subdimension has on *salience*. SEMs are particularly suitable to assess the degree of influence each subdimension has. As there are different analysis methods for SEMs this section justifies my selection, too. Based on this justification, the PLS-PM algorithm will be described in detail. These theoretical considerations lead to the need of a measurement model. Consequently, I will detail the MVs used to collect data in this chapter. Finally, the results of a pre-study on an adequate way to phrase questions are presented.

3.3.1 *Why Use Structural Equation Models—Pros and Cons and a Decision*

All current models of *salience* share one important aspect. *Salience* is always viewed as having multiple subdimensions. The theories discussed lead to a model including multiple relationships between multiple constructs. This means, multivariate techniques, incapable of examining multiple relationships at the same time are rendered useless. A method is needed which allows for the use of all available information concurrently. In contrast to factor analysis, multiple regression or MANOVA approaches, Structural Equation Modeling has these capabilities. Since the mid 1990s SEM techniques have become increasingly widespread in social science research (cf. Hair et al., 2010, p. 642). The relations between several LVs in a structural model can be assessed simultaneously accompanied by their measured counterparts (cf. Gefen, Straub, & Boudreau, 2000). In contrast to exploratory factor analysis, where no measurement model specification is required at all (cf. Hair et al., 2010, p. 641), SEM analysis requires a specification of dependencies according to theory. Using latent, i.e. not directly measured, variables to build a model is sensible for two reasons. First, using multiple indicators to measure a single variable reduces measurement error. Second, depending on the measurement method, measurement error present in the constructs can be accounted for (cf. Hair et al., 2010, p. 635). It is important to note, that variables not directly observable are very widespread. Haenlein and Kaplan (2004, p. 284) give a detailed account:

“McDonald (1996) stressed that a variable can be called observable ‘if and only if its value can be obtained by means of a real-world

3 HOW CAN SALIENCE BE MEASURED?

sampling experiment' (p. 239). Therefore, any variable that does not correspond directly to anything observable must be considered as unobservable (Dijkstra, 1983). This definition makes it obvious that only a handful of relevant variables, such as age and gender, can be considered as observable, whereas 'the effects and properties of molecules, processes, genes, viruses, and bacteria are usually observed only indirectly' (S. Wold, 1993, p. 138)".

Taken as a whole, SEM as an approach of data analysis fits the goals of this thesis (cf. section 1.1) perfectly.

A Note on SEM Terminology

As, to date, SEM is a rarely used data analysis method in Information Science, some notes on SEM in general and on terminology in particular are advisable. All SEMs consist of two parts. The structural model part describes the relationship between the LVs, whereas the measurement model details the link between each of the LVs and its MVs used to provide a measure for the current construct. In figures, a clear distinction is made between both groups of variables (cf. fig. 4), as LVs are depicted using circles and MVs using boxes. Arrows are used for relationships between LVs and between LVs and MVs. A LV is called *endogenous* if and only if arrow heads from other constructs are pointing to it. Contrastingly, constructs without any arrow heads from other constructs pointing to them are referred to as *exogenous*. The direction of arrows between MVs and LVs indicates the type of measurement. Arrow heads pointing from MVs to LVs denote formative measurement, whereas reflective measurement is shown in the opposite manner. Formatively measured variables are referred to as *causes* or *causal-formative indicators*, whereas their reflectively measured counterparts are commonly referred to as *items* or *indicators*.

Partial Least Squares Path Modeling (PLS-PM) vs. Covariance-based Structural Equation Modeling (CB-SEM)

Generally speaking, two different techniques are widely employed to assess SEMs.

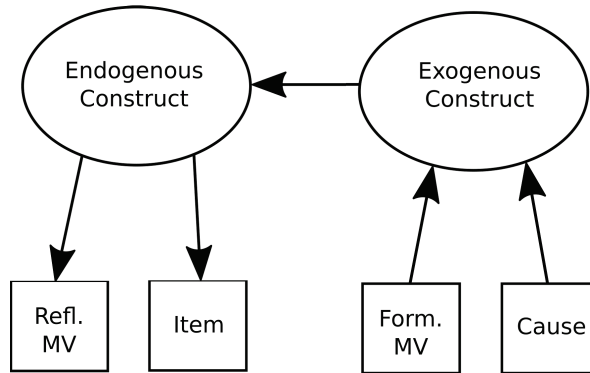


Figure 4.: An example of a structural model used to explain the SEM terminology. This figure was drawn using Inkscape (cf. The Inkscape Team, 2016).

COVARIANCE-BASED SEM tries to find those coefficients, which allow the empirical covariance matrix to be reproduced as well as possible. Several assumptions, e.g. multivariate normality of the data must be fulfilled in order to successfully apply CB-SEM (cf. Chin & Newsted, 1999, p. 309). Reflective measurement models and MIMIC-models can be estimated and both, recursive and non-recursive structural models may be specified. The approach is also known as *LISREL* (cf. Haenlein & Kaplan, 2004, p. 285), which is actually the name of the first tool with which CB-SEM could be calculated (Jöreskog, 1971).

VARIANCE-BASED SEM is focused on prediction. This means, it tries to estimate coefficients which maximize the amount of variance explained in the endogenous construct (cf. Hair, Ringle, & Sarstedt, 2011, p. 140). This approach is also known as Partial Least Squares Path Modeling (PLS-PM). “The estimation procedure is named *partial* [sic] since it solves blocks [of LVs and its associated MVs] one at a time by means of alternating single and multiple linear regressions” (Vinzi, Trinchera, & Amato, 2010, p. 52). PLS-PM was introduced in H. O. A. Wold (1975) and is a non-parametric method, suitable only for recursive models. This means that no bidirectional arrows between two or more LVs can exist. PLS-PM can be used to estimate formative, reflective and MIMIC-models.

In general, SEM is known to be a large sample size method. As of a sample size of $n = 400$ the results for both approaches are practically identical—given that the assumptions of CB-SEM are met. One important difference between the approaches, however, is the availability of a global goodness-of-fit criterion. Global goodness-of-fit can be assessed in case of CB-SEM using several different χ^2 -based measures. However, PLS-PM lacks a similar global model fit criterion. Instead, bootstrapping (cf. Efron, 1979) is used to assess the significance of estimated values. Moreover, PLS-PM is known to induce a bias during model estimation. It makes extensive use of LVs scores which are, in turn, derived from the values of their corresponding MVs. This means that the measurement error present in these is propagated onto the LVs. In contrast, the covariance-based approach allows to explicit model error in LVs. As a consequence, estimations are by design inconsistent in PLS-PM estimations (cf. Fornell & Cha, 1994). In fact, PLS-PM estimations are only consistent at large, i.e. both, the number of indicators per LV and the sample size must converge to infinity (cf. Haenlein and Kaplan (2004, p. 292), who refer to McDonald (1996, p. 248) and Lohmöller (1989)). In this case the error present in the LV scores will be zero and the bias will disappear. This property has been known as PLS-PM bias as it yields overestimation of weights and loadings in measurement models and underestimated structural relationships (cf. Hair Jr. et al., 2014, p. 79). However, there is evidence that this bias can be neglected in many real-world scenarios if proper sample sizes are used (cf. e.g. Reinartz et al., 2009). Furthermore, PLS_c was developed in order to overcome this disadvantage among others (cf. section 6.2.3).

Given these properties, an informed decision with respect to which of the approaches should be taken must be based on the research goals of this thesis. Hair et al. (2011) provides useful rules of thumb upon which to base a final decision. One of the important goals of this thesis (cf. section 1.1) is to find out, which of the features revealed in earlier studies is most important when predicting an object's *overall salience*. This is the first time that theories on *salience* have been combined to form a coherent model and this model ought to be analyzed as a whole. Furthermore, *visual salience* will be modeled formatively (cf. section 3.3.3). This is particularly useful in order to gain insight into the importance

of different visual features (cf. Fassott & Eggert, 2005, p. 47). Another advantage of the method is the absence of assumptions regarding the distribution of the data (cf. e.g. Cassel, Hackl, & Westlund, 1999). This is of particular importance for data collected in-situ as these data are more likely to be non-normally distributed. In fact, the strictness of the assumptions in CB-SEM were one of the reasons why Herman Wold developed the PLS method (cf. Hair et al., 2011, p. 140). According to these preconditions, PLS-PM is most suitable to reach the goals of this doctoral work, particularly as the degree of influence on the construct level may be assessed, too (cf. Hair Jr. et al., 2014, p. 78).

Finally, it is important to note, that PLS-PM must be distinguished from PLS Regression, although both techniques were developed almost simultaneously (cf. Mateos-Aparicio Morales, 2011). In Ordinary Least Squares (OLS) regression projections on the linear space built by independent variables are used to determine the model's coefficients. Hence, a hyperplane yielding a minimum of variance in both independent variables (IVs) and dependent variable (DV) is found. Contrastingly, PLS regression is based on projections of IVs and DV onto a new space each. The goal is "to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space" (Wikipedia, 2015). However, there are occasions where substituting the default OLS regression by PLS Regression methods is advantageous. These cases include situations when LV scores are missing, LVs are highly correlated or the number of observations is low for the MVs of the most complex structural equation (cf. Vinzi et al., 2010, p. 55). By default PLS-PM uses OLS regression to fit the model and this regression method will be used for all estimations reported throughout this thesis.

3.3.2 *How PLS Path Modeling Works*

Please note: The notation and the formulae in this subsection are taken from Vinzi et al. (2010). All figures presented are done by the author.

The Structural Model

Figure 5 introduces major aspects of describing PLS Path Modeling. Each LV ξ has p MVs assigned to it building one of Q blocks. Each of

3 HOW CAN SALIENCE BE MEASURED?

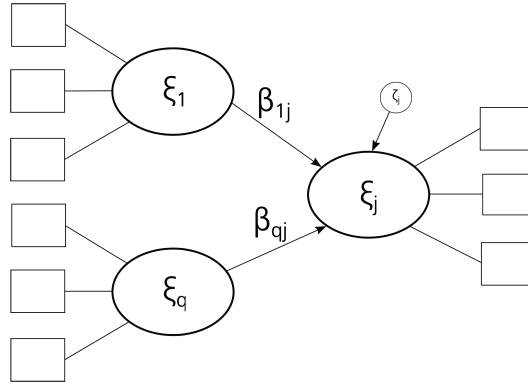


Figure 5.: An example structural model introducing major variable names for the explanations given. The figure was drawn using Inkscape (The Inkscape Team, 2016).

the MVs was measured on N objects. These measurements are used to estimate the relationship between the Q blocks. Given this setup, the structural model part can be written as eq. (3)¹⁰.

$$\xi_j = \beta_{0j} + \sum_{q:\xi_q \rightarrow \xi_j} \beta_{qj} \xi_q + \zeta_j \quad (3)$$

It focuses on the relationship between endogenous (ξ_j in fig. 5) and exogenous (ξ_q in fig. 5) variables. For each endogenous variable ξ_j a path coefficient β_{qj} can be estimated, describing the relationship between the q -th exogenous and the j -th endogenous construct. Any error in estimating the endogenous construct based on *all* its associated exogenous constructs is given by ζ_j . β_{0j} represents the intercept, i.e. a real number specific to the estimation of ξ_j .

The Measurement Model Parts

Given the structural model part we can turn to the measurement model part. I would like to give a general introduction into different types of measurement models before continuing with the explanation regarding

¹⁰ The explanations are based on (Vinzi et al., 2010, p. 49).

the algorithm. Three important types of models must be distinguished (cf. e.g. Fornell & Bookstein, 1982): reflective measurement, formative measurement, and a combination of these called Multiple Indicators, Multiple Causes (MIMIC). Formative and reflective measurement models are fundamentally different with respect to the role the LVs play.

REFLECTIVE In reflective models each of the measured variables is supposed to measure the underlying construct equally well. They ought to reflect the construct's value, which cannot be measured directly. Hence, arrows are directed from LVs to MVs in reflective measurement model parts.

FORMATIVE In contrast to reflective models, measured variables are presumed to cause the manifestation of the LV. As a consequence arrows are directed from MVs to LVs.

MVs are considered as effect indicators, i.e. they “share [...] [a] common cause” (DeVellis, 2012, p. 12) in case of reflective measurement, which is, therefore, often referred to as *common factor model*. “[L]atent constructs cause the measured variables and the error results in an inability of the construct to fully explain these measured variables. Thus, the direction of the arrows is from latent constructs to measured variables and error terms are associated with each measured variable” (Hair et al., 2010, p. 749). As a consequence, all reflective indicators used to measure a single LV should be correlated. This means, changes in one MV must go along with positively correlated changes in all of the other MVs as the change is supposed to be caused by a change in LV's score. If one of its MVs would have been dropped, the meaning of the construct would still remain unchanged. Contrastingly, in formative measurement the causal-formative indicators determine a construct's level, i.e. they “determine the same outcome” (DeVellis, 2012, p. 12). This means, the error is a result of missing one or more of the indicators forming the construct, i.e. the list of causes used is supposed to be exhaustive (Jarvis, MacKenzie, & Podsakoff, 2003)¹¹. However, changes in one MV do not imply that the other MVs show a change in a similar direction or any change at all. There is no need for MVs to be correlated at all.

¹¹ But see section 6.2.1 for an updated view on this claim.

In contrast, large multicollinearity of causes may yield weak, i.e. unreliable, results due to the multiple regression approach used to estimate the model (cf. Hair et al., 2011, p. 75). This may lead to situations, where neglecting causes results in changing the construct’s meaning and keeping the causes may weaken the estimation results. The major advantage of formative models of measurement is their suitability for assessing the degree as to which a LV is influenced by indicators. However, a major drawback is that a formative model’s parameters cannot be estimated in case the model consists of a single formative construct, only. Employing a MIMIC-model is a common method to overcome this disadvantage, i.e. reflective indicators suitable to measure the formative construct must be introduced. This is not necessary in models involving more than one LV. The so-called nomological network LVs are embedded in (cf. Cronbach & Meehl, 1955) is sufficient to estimate all parameters of the formatively measured construct in these cases. Under these circumstances the indicators can be used for a so-called convergent validity analysis suitable to provide empirical evidence for the applicability of any formative assessment (cf. section 5.4.3). Based on these general distinctions, the reflective and formative parts of the model are treated differently.

ASSESSING THE REFLECTIVE MODEL PART ¹²In case of reflective measurement, simple regression of the LV on each of its indicators is applied (cf. eq. (4)). This means, for each of the p MVs in the q -th block, the outer loading λ_{pq} is assessed. Standardized outer loadings will provide

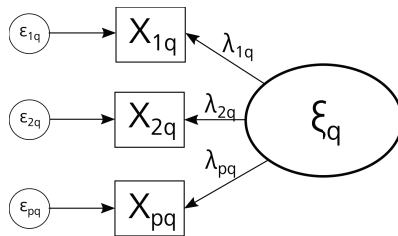


Figure 6.: A simple example of reflectively measuring a construct, thereby introducing major variables of the estimation process. The figure was drawn using Inkscape (cf. The Inkscape Team, 2016).

¹² These explanations are based on Vinzi et al. (2010, pp. 49–50).

correlations between the MVs and the LV in the q -th block. Any error in the measurement eventually leading to imprecise predictions of x_{pq} is modeled by ϵ_{pq} , i.e. on the MV-level. λ_{p0} thereby represents the intercept.

$$x_{pq} = \lambda_{p0} + \lambda_{pq}\xi_q + \epsilon_{pq} \quad (4)$$

A major assumption of this modeling error is given in eq. (5). This assumption is known as predictor specification and is important to yield reliable results in OLS regression scenarios. For each block Q the modeling errors are assumed to be uncorrelated with the current ξ_Q and the expected value of ϵ_{pq} is zero.

$$E(x_{pq}|\xi_q) = \lambda_{p0} + \lambda_{pq}\xi_q \quad (5)$$

FORMATIVE MODEL PART ¹³In contrast to reflective measurement items, formative MVs are not supposed to *measure* a single construct. Instead they are viewed as forming the LV they are associated with (cf. the direction of errors in fig. 7), i.e. each of the MVs adds something unique to the LV. Hence, homogeneity of causal-formative indicators is neither

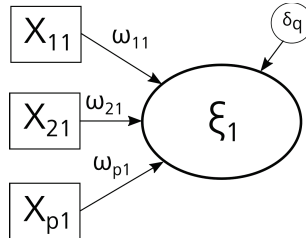


Figure 7.: A formative measurement example introducing major variables used throughout this description. The figure was drawn using Inkscape (cf. The Inkscape Team, 2016).

assumed nor desired. Instead, the LV is modeled as a linear combination (cf. eq. (6)).

$$\xi_q = \sum_{p=1}^{P_q} \omega_{pq}x_{pq} + \delta_q \quad (6)$$

¹³ These explanations are based on Vinzi et al. (2010, p. 51).

For each of the q LVs the sum of ω_{pq} of the p MVs associated to it is calculated. As explained earlier (cf. section 3.3.2) MVs are supposed to be causes of their corresponding LV. As a consequence completeness of contents is assumed. In order to reflect this assumption the error term δ_q is introduced. It models the error in measuring the q -th LV resulting from the fact that not all causes p are included in the q -th block. Having said this, the predictor specification for formative measurement models is given in eq. (7), i.e. the expected value of the true score for a LV given its x_{pq} indicators equals the sum of its weighted MVs.

$$E(\xi_q | x_{pq}) = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} \quad (7)$$

The Core of the Algorithm

¹⁴Regardless of the measurement type employed the goal of a PLS-PM analysis is to estimate the scores of the LVs as a linear combination. Equation (8) provides the so-called weight relation the calculation is based on. Standardized scores $\hat{\xi}_q$ for each of the LVs ξ_q are computed by using the outer weights of variable p in the q -th block and the value of each MV. They are achieved by alternating two different steps, outer and inner estimation, until convergence ¹⁵. Convergence is achieved if the difference between the outer weights calculated during two subsequent iterations is less than a predefined threshold usually set to $\Delta = 10^{-5}$ according to (cf. Hair Jr. et al., 2014, p. 82).

$$\hat{\xi}_q = \sum_{p=1}^{P_q} w_{pq} x_{pq} \quad (8)$$

¹⁴ The explanations are based on Vinzi et al. (2010, pp. 51–55).

¹⁵ A formal proof of convergence for models is yet to be shown and is impossible for certain cases (cf. Henseler, 2010, p. 107). Henseler shows that there are cases where convergence cannot be achieved combining Mode A with one of the weighting schemes factorial or path (cf. the explanations of the inner estimation step).

STEP 1: OUTER ESTIMATION The whole estimation process is based on standardized MVs. On initializing the algorithm, random outer weights are chosen. During the outer estimation (cf. eq. (9)) the standardized outer weight for each LV is estimated, such that the left side of the equation is proportional to the standardized right side. Using \pm indicates the importance of the signs in terms of specification of the measurement model for the current LV.

$$v_q \propto \pm \sum_{p=1}^{P_q} w_{pq} x_{pq} = \pm X_q w_q \quad (9)$$

STEP 2: INNER ESTIMATION During the subsequent inner estimation phase the chosen weighting scheme is important. Based on this schema eq. (10) is used to obtain a LV score, i.e. a standardized inner estimate \mathfrak{D} of LV ξ_q by taking all its connected LVs into account. The schema is used to determine $e_{qq'}$, i.e. the so-called inner weight.

$$\mathfrak{D}_q \propto \sum_{q'=1}^{Q'} e_{qq'} v_{q'} \quad (10)$$

Three different schemes are distinguished, all of which are based on the connection between a LV q and q' :

CENTROID Use the same sign as the correlation between the outer estimate of the current variable q and the outer estimate of q' (cf. Vinzi et al., 2010, p. 53).

FACTORIAL Use the correlation coefficient between the outer estimate of the current variable q and the outer estimate of q' (cf. Vinzi et al., 2010, p. 53).

PATH Use the regression coefficient between q and q' if the current LV is an endogenous construct. In case q is an exogenous LV use the correlation coefficient (cf. Vinzi et al., 2010, p. 53).

Vinzi et al. (2010, p. 53) recommend using the path option because it is the only scheme which acknowledges the specification of the measurement

model, i.e. whether reflective or formative measurement is employed for a particular LV. It is important to note that two different ways of estimating the LV-values exist. Lohmöller (1989) and H. O. A. Wold (1982) provide different methods to estimate LV scores. Henseler (2010, pp. 112-113) provides an example using three LVs. Table 2 shows the calculations for iteration 1 in this case. Wold’s approach uses the most recent estimations available to calculate LV scores whereas Lohmöller employs those values derived in the last run. Lohmöller’s approach can be calculated using matrix algebra. This increased efficiency makes it particularly widespread in tools used to assess PLS-PM models, although Wold’s original attempt is known to converge faster.

Iteration	Lohmöller’s procedure	Wold’s procedure
i	\dots $\xi_1^{(i)}, \xi_2^{(i)}, \xi_3^{(i)}$	\dots $\xi_1^{(i)}, \xi_2^{(i)}, \xi_3^{(i)}$
	$\xi_1^{(i+1)} = f\left(\xi_1^{(i)}, \xi_2^{(i)}, \xi_3^{(i)}\right)$	$\xi_1^{(i+1)} = f\left(\xi_1^{(i)}, \xi_2^{(i)}, \xi_3^{(i)}\right)$
i + 1	$\xi_2^{(i+1)} = f\left(\xi_1^{(i)}, \xi_2^{(i)}, \xi_3^{(i)}\right)$	$\xi_2^{(i+1)} = f\left(\xi_1^{(i+1)}, \xi_2^{(i)}, \xi_3^{(i)}\right)$
	$\xi_3^{(i+1)} = f\left(\xi_1^{(i)}, \xi_2^{(i)}, \xi_3^{(i)}\right)$	$\xi_3^{(i+1)} = f\left(\xi_1^{(i+1)}, \xi_2^{(i+1)}, \xi_3^{(i)}\right)$
	$\xi_1^{(i+1)}, \xi_2^{(i+1)}, \xi_3^{(i+1)}$	$\xi_1^{(i+1)}, \xi_2^{(i+1)}, \xi_3^{(i+1)}$

Table 2.: The different ways of estimating LV scores based on Lohmöller (1989) and H. O. A. Wold (1982), respectively. As an example, three LVs are shown. The table is taken from Henseler (2010, p. 112).

STEP 3: UPDATE OUTER WEIGHTS ¹⁶Based on this estimation of the LV score outer weights w_{pq} are updated according to one of two modes. *Mode A* means to use the regression coefficient of the simple regression of the observed value of the MV, x_{qp} , on the inner estimate ϑ_q LV it belongs to.

$$w_{pq} = cov(x_{qp}, \vartheta_q) \tag{11}$$

This regression coefficient becomes a simple covariance (cf. eq. (11)) when standardized values for ϑ_q are used as depicted in eq. (10). If the

¹⁶ Based on Vinzi et al. (2010, pp. 53–54).

observed values are standardized, as well, the correlation coefficient is used. For *Mode B*, a multiple regression approach is taken (cf. eq. (12)). X_q denotes the standardized observed values x_{pq} , which were additionally scaled to $\sqrt{1/N}$ of the manifest variables in the current block q . This means, the vector of outer weights for block q is updated using the regression coefficients resulting from the multiple regression of the inner estimate ϑ_q on the observed values of all the MV associated with it (X_q).

$$w_q = (X_q' X_q)^{-1} X_q' \vartheta_q \quad (12)$$

Given these definitions, it becomes evident that *Mode A* is to be preferred in case of reflective measurement, whereas *Mode B* suits formative measurement models best. Finally, correlations between MVs of a block q may endanger results produced (cf. Vinzi et al., 2010, p. 54). This is, multicollinearity may render regression coefficients insignificant (cf. Mateos-Aparicio Morales, 2011, p. 2306). Substituting OLS regression by PLS regression may be used to overcome these issues (cf. Vinzi et al., 2010, p. 55).

STEP 4: CALCULATING LV SCORES UPON CONVERGENCE ¹⁷In case convergence is achieved eq. (8) is used to calculate LV scores. This means, PLS-PM uses aggregated MVs prone to measurement error to establish construct scores. As a result, inconsistency of these scores cannot be avoided, which is the reason of the PLS-PM bias (cf. section 3.3.1)¹⁸. Subsequently, the so-called path coefficients which describe the structural model, i.e. the relationship between constructs, are calculated based on the score of the current LV and the endogenous LV scores of all constructs related to it.

Vinzi et al. (2010, p. 56) summarize the whole algorithm using pseudo-code, based on “the following options: centroid scheme, standardizes latent variable scores, OLS regressions”.

¹⁷ Based on Vinzi et al. (2010, p. 55).

¹⁸ This drawback is overcome by the advent of PLS, cf. section 6.2.3.

Algorithm 1 A pseudo-code representation of the non-consistent PLS Path Modeling algorithm as provided by Vinzi et al. (2010, p. 56).

Input: $X = [X_1, \dots, X_q, \dots, X_Q]$, i.e. Q blocks of centered manifest variables;

Output: $w_q, \hat{\xi}_q, \beta_j$;

for all $q = 1, \dots, Q$ **do**

 initialize w_q

$$v_q \propto \pm \sum_{p=1}^{P_q} w_{pq} x_{pq} = \pm X_q w_q$$

$$e_{qq'} = \text{sign}[\text{cor}(v_q, v_{q'})]$$

$$\vartheta_q \propto \sum_{q'=1}^{Q'} e_{qq'} v_{q'}$$

 update w_q :

 (a) $w_{pq} = \text{cov}(x_{qp}, \vartheta_q)$ for Mode A (outwards directed model)

 (b) $w_q = (X_q' X_q)^{-1} X_q' \vartheta_q$ for Mode B (inwards directed model)

end for

[... the for-loop is] **repeated until convergence** on the outer weights is achieved, i.e. until:

$$\max\{w_{pq, \text{current iteration}} - w_{pq, \text{previous iteration}}\} < \Delta$$

where Δ is a convergence tolerance usually set at [... $\leq 10^{-5}$]

Upon convergence:

(1) for each block the standardized latent variable scores are computed as weighted aggregates of manifest variables:

$$\hat{\xi}_q \propto X_q w_q$$

(2) for each endogenous latent variable $\xi_j (j = 1, \dots, J)$, the vector of path coefficients is estimated by means of OLS regression as:

$$\beta_j = (\hat{\Xi}' \hat{\Xi})^{-1} \hat{\Xi}' \hat{\xi}_j$$

where $\hat{\Xi}$ includes the scores of the latent variables that explain the j -th endogenous latent variable ξ_j , and $\hat{\xi}_j$ is the latent variable score of the j -th endogenous latent variable

3.3.3 *The Measurement Model*

The last two sections detailed important theoretical aspects of SEM in general and PLS-PM in particular. The algorithm makes a clear distinction between those parts of the model dealing with relationships between LVs, i.e. the structural model, and the relationships between reflective indicators, causal-formative indicators and their corresponding constructs, i.e. the measurement model. It is important to keep in mind, what measurement essentially is. It is

“the process by which a concept is linked to one or more latent variables, and these are linked to observed variables’ (Bollen, 1989, p. 180). Once we have the latent variable that represents the construct in a model, we need to ask how the indicator is related to the latent variable. The dominant assumption is that the latent variable affects the indicator so that the latter is an effect or reflective measure. But another possibility is that the indicator affects the latent variable as with causal-formative indicators. In either case, the indicator is a measure of the latent variable, even though the nature of their relationships is different” (Bollen & Diamantopoulos, 2015, pp. 7–8).

Accordingly, for each MV, a question is derived in order to capture the meaning of a particular LV. This section details these questions for each LV, based on the definition of the construct given first. In general, the number of MVs per construct was based on the following rules (cf. Hair et al. (2011, p. 701)).

1. Avoid less than three MVs per construct.
2. Use four MVs per LV.
3. Use three MVs only if there are other constructs having more than three.

Accordingly, less than four indicators were used in two cases only. Providing only three indicators for *prototypicality* seems to be reasonable as

prior studies even neglect this construct. Similarly, for the key target construct, *overall salience*, no more than three MVs were provided. All other LVs used in the SEM have more than three indicators each, counterbalancing the potential lack of indicators for *overall salience* and *prototypicality*. It is important to note, that all questions were phrased and asked in German language and are only translated for ease of reading. In its English language version the measurement model was first presented in Kattenbeck (2015a, p. 10), i.e. the definitions for constructs and the translations of questions are taken from there.

Overall Salience

Overall Salience (OVSAL)

“The overall salience of geographic features is defined as a three-valued vector, whereby the components capture perceptual, cognitive, and contextual aspects of geographic objects” (Caduff & Timpf, 2008, p. 264).

Var	Question	ToM
ov_att	To what extent does this object draw your attention? <i>Wie stark zieht dieses Objekt Ihre Aufmerksamkeit auf sich?</i>	r
ov_por	How suitable is this object to be used as a point of reference? <i>Wie gut eignet sich dieses Objekt als Bezugspunkt?</i>	r
ov_mem	How memorable is this object? <i>Wie gut können Sie sich dieses Objekt einprägen?</i>	r

Table 3.: The items used to measure *overall salience*. Column Var gives the variable codes used, ToM the type of measurement used, where r means reflective and f formative measurement. Items are given in the order they were presented to participants.

Overall salience is the key target construct all subdimensions are somehow related to. It is measured using the three questions presented in table 3. The items aim to capture different facets of salience as defined by Caduff and Timpf (2008).

*Prototypicality*Prototypicality (PRO)

“[...] [t]hat is, how typically they represent a category” (Sorrows & Hirtle, 1999, p. 43).

Var	Question	ToM
p_exa	To what extent is this object suitable as an example of objects belonging to the category you named? <i>Wie gut eignet sich dieses Objekt als Beispiel für Objekte der von Ihnen vergebenen Kategorie?</i>	r
p_img	To what extent does this object represent your impression of such objects? <i>Wie gut entspricht dieses Objekt Ihren Vorstellungen eines solchen Objekts?</i>	r
p_sim	How often do you encounter similar objects? <i>Wie häufig sehen Sie ähnliche Objekte?</i>	r

Table 4.: The items used to measure *prototypicality*. Column Var gives the variable codes used, ToM the type of measurement used, where r means reflective and f formative measurement. Items are given in the order they were presented to participants.

In general, *prototypicality* is known to be difficult to measure. Reflecting this, major models, such as Raubal and Winter (2002), do not provide measures for this dimension. Nevertheless, three items covering different notions of *prototypicality* are derived for this thesis. I draw on hints by Sorrows and Hirtle (1999), who introduced *prototypicality* to be one of the dimensions of *salience* as well as I take inspiration from Rosch et al. (1976) and Mervis and Rosch (1981). Whereas p_exa and p_img try to gain an insight into the user’s impression of objects, p_sim is a consequence of an object’s prototypicality: The more prototypical an object is, the more often it would be encountered.

*Visibility in Advance*Visibility in Advance (ADV)

The degree as to which an object (at a potential DP) may be seen from the direction it is approached at (cf. Winter, 2003).

3 HOW CAN SALIENCE BE MEASURED?

Var	Question	ToM
a_dis	To what extent can one easily refer to this object from afar? <i>Wie gut kann auf dieses Objekt auch aus größerer Entfernung verwiesen werden?</i>	r
a_vis	Given the current route, to what extent were you able to see this object from a distance? <i>Inwieweit ist dieses Objekt auf der aktuellen Route schon seit Längerem sichtbar?</i>	r
a_per	To what extent is this object generally perceptible from afar? <i>Wie gut ist dieses Objekt aus der Ferne wahrnehmbar?</i>	r
a_sui	In the context of the current route to what extent is this object suitable to explain the route? <i>Wie gut eignet sich dieses Objekt, um den Verlauf der aktuellen Route zu beschreiben?</i>	r

Table 5.: The items used to measure *visibility in advance*. Column Var gives the variable codes used, ToM the type of measurement used, where r means reflective and f formative measurement. Questions are given in the order they were presented to participants.

As explained above (cf. section 2.4), *visibility in advance* has a strong focus on recognizability. This aspect is reflected by the way items were phrased (cf. table 5). a_sui and a_vis are explicitly referring to the current route. Contrastingly, a_per and a_dis are worded generally.

Cognitive Salience

Cognitive Salience (COS)

“[...] The processing of information is based on prior knowledge, while intentions and strategies of the observer are in control of the allocation of attention. In our framework, we will use the term Cognitive Salience to refer to the endogenous factors that influence salience” (Caduff & Timpf, 2008, p. 255).

As mentioned earlier (cf. section 2.3.3), I prefer the term *cognitive* to *semantic*. From my perspective, the latter is subsumed by the earlier. To me, the term *cognitive* stresses the fact that this dimension of *salience* draws on aspects a person already has in mind, when viewing an object. The questions used to measure *cognitive salience* try to reflect both, the broader definition as given by Caduff and Timpf (2008) and the narrower definition by Raubal and Winter (2002). c_eas is used as an item reflecting *cognitive salience* as a whole because objects which play a major role in

Var	Question	ToM
c_per	To what extent do you have personal memories concerned with this object? <i>Inwieweit verbinden Sie mit diesem Objekt persönliche Erinnerungen?</i>	r
c_his	To what extent does this object's appearance suggest it to be historic? <i>Wie historisch wirkt dieses Objekt auf Sie?</i>	r
c_wor	To what extent do you regard this object to be worthy of preservation? <i>Inwieweit ist dieses Objekt Ihrer Meinung nach erhaltenswert?</i>	r
c_cus	To what extent is the current use of the object obvious? <i>Wie klar ersichtlich ist die aktuelle Nutzung dieses Objekts?</i>	r
c_pus	To what extent is the former use of the object obvious? <i>Inwieweit lässt sich erkennen, dass das Objekt ursprünglich eine andere Funktion erfüllte?</i>	r
c_eas	How easy was it for you to label this object? <i>Wie leicht fiel es Ihnen, eine Bezeichnung/Benennung für dieses Objekt zu finden?</i>	r

Table 6.: The items used to measure *cognitive salience*. Column Var gives the variable codes used, ToM the type of measurement used, where r means reflective and f formative measurement. Items are given in the order they were presented to participants.

the knowledge acquired by an observer will be straight forward to label. The variables c_his and c_wor reflect the cultural and historical dimension indicated by Raubal and Winter (2002). Idiosyncratic relevance, as suggested by Caduff (2007) (cf. section 2.3.3), was measured using the variable c_per. Finally, c_cus and c_pus are included because of their importance for building knowledge about a particular object. If the prior or current usage or both are obvious, this will contribute to the meaning an object has to an observer and increases object-based attention as viewed by Caduff (2007).

Structural Salience

Structural Salience (STS)

“Objects are called structurally salient if their location is cognitively or linguistically easy to conceptualize in route directions” (Klippel & Winter, 2005, p. 347).

As explained earlier, the location of LMs in relation to DPs is crucial to the understanding of *structural salience* as a concept (cf. section 2.3.3). The questions used to measure *structural salience* reflect this viewpoint. s_imp and s_dir aim to the suitability an object shows to identify a

3 HOW CAN SALIENCE BE MEASURED?

Var	Question	ToM
s_eas	How easy is it for you to refer to this object in a route description? <i>Wie leicht fällt es Ihnen, auf dieses Objekt in einer Routenbeschreibung zu verweisen?</i>	r
s_lor	How easy is it to describe this object's location as part of the current route? <i>Wie einfach ist es, die Lage dieses Objekts auf der Route zu beschreiben?</i>	r
s_imp	To what extent is this object located at an important location within the current route? <i>Inwieweit liegt dieses Objekt an einem wichtigen Punkt dieser Route?</i>	r
s_dir	To what extent may this object be suitable to determine whether this is an appropriate route or a change in course is required? <i>Wie gut lässt sich dieses Objekt dazu verwenden, den Ort einer Richtungsänderung zu identifizieren oder den korrekten Routenverlauf zu bestätigen?</i>	r

Table 7.: The items used to measure *structural salience*. Column Var gives the variable codes used, ToM the type of measurement used, where r means reflective and f formative measurement. Items are given in the order they were presented to participants.

decision point. The two remaining indicators, *s_eas* and *s_lor*, are used to assess the degree as to which the object may easily be integrated into route directions. This again reflects the taxonomy provided by Klippel and Winter (2005): In route directions the location of an object relative to a decision point is often included simultaneously, e.g. “turn right in front of the Cathedral”.

Visual Salience

Visual Salience (VIS)

“The features of contrast with surroundings, prominence of spatial location, and visual characteristics that make the landmark particularly memorable” (Sorrows & Hirtle, 1999, p. 45).

Compared to other subdimensions of *salience* a vast number of different dimensions have been named in earlier studies as being important for *visual salience* (cf. table 8; a similar, yet less exhaustive overview is given by Sadeghian and Kantardzic (2008, p. 273)). There are several reasons why *visual salience* may be *prima inter paria*. There is strong empirical evidence that non-blind people prefer visual navigation cues (cf. Janzen & Turennout, 2004). Furthermore, visual attributes can be employed for different user groups because they are physical features LMs have (cf. Röser et al., 2011) and, hence, do not rely on prior knowledge about

the local environment. This means, visual aspects are most universally usable with respect to different user groups. One can imagine that there are different groups of users who use navigation systems for various reasons. These users may have diverging levels of knowledge regarding the local environment within which they are navigating. As a consequence, certain types of route instructions may be unsuitable for users with low-knowledge (e.g. “pass the courthouse”), whereas as long as the user is not visually impaired, visual aspects are generally appropriate (e.g. “pass the yellow building with the large clock on it”).

Dimension	References
age	Raubal & Winter, 2002; Schroder et al., 2011
architecture	Gärling, Böök, & Lindberg, 1986
cleanliness	Raubal & Winter, 2002
colour	Hamburger & Röser, 2011; Lazem & Sheta, 2005; Nothegger et al., 2004; Raubal & Winter, 2002; Schroder et al., 2011
company	Sefelin et al., 2005
condition	Schroder et al., 2011
construction	Schroder et al., 2011; this refers to construction material
height	Elias, 2003a, 2003b; Lazem & Sheta, 2005
length	according to Sadeghian & Kantardzic, 2008 proposed but not used in experiments
location	Sorrows & Hirtle, 1999
shape	Caduff & Timpf, 2008; Hamburger & Röser, 2011; Raubal & Winter, 2002
signage	Duckham et al., 2010; Ishikawa & Nakamura, 2012; Nothegger et al., 2004; Partala et al., 2010; Sefelin et al., 2005
size	Caduff & Timpf, 2008; Matsumoto, Satake, Kanda, Imai, & Hagita, 2012; Nothegger et al., 2004; Raubal & Winter, 2002
texture	Nothegger et al., 2004; Partala et al., 2010; Raubal & Winter, 2002
visibility	Xia et al., 2011
width	Lazem & Sheta, 2005

Table 8.: Overview of visual dimensions named in existing studies. The dimensions are given in alphabetical order and are assigned to their corresponding references.

This means, measuring *visual salience* thoroughly is particularly important. The measurement model presented in table 9 is different from the ones just presented. It shows causal-formative indicators for *visual salience*. The aspects revealed in prior studies are all modeled as formative causes. According to Bollen (2002) in most cases a clear decision made by researchers leads to the modeling of a construct in either a reflective or formative manner. Consequently, employing formative measurement for *visual salience* needs to be justified, particularly because misspecification may have a major impact on the reliability of results (cf. Bollen & Lennox, 1991). Several aspects must be taken into account:

MUTUALLY INTERCHANGEABLE Formative causes must not be mutually interchangeable (cf. Jarvis et al., 2003, p. 203). From my point of view, the dimensions given in table 8 are not interchangeable. In fact, they are rather loosely related aspects (cf. Fassott & Eggert, 2005, p. 47), but all of them contribute to *visual salience*.

IMPLIED CAUSALITY A second indicator for the appropriateness of formative measurement in case of these dimensions is given by the causality implied (cf. e.g. Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003). The dimensions revealed in earlier studies are characteristics contributing to *visual salience*. For example, color and condition both rather form *visual salience* than they are a consequence of these.

THOUGHT EXPERIMENT Imagine one were able to change a construct's score directly. In case of reflective measurement, where all indicators share a common theme, a change in the same direction for all indicators must be the result (cf. Chin, 1998). This is not reasonable in this case, for example, color and condition are both rather distinct aspects (i.e. they are not correlated). Consequently, they are not likely to covary both with *visual salience* if the construct's score could be changed.

Taking all these arguments into account, I model *visual salience* as a formative LV. This decision has two consequences. First, all the dimensions revealed earlier were acknowledged in the measurement model of the current study (cf. table 9)—except for dimensions *architecture* and

Var	Question	
	<p>Please find below several visual attributes. For each of these, please indicate the extent to which the named visual attribute contributes to an object's salience given its surroundings. <i>Nachstehend finden Sie eine Reihe visueller Aspekte. Bitte beurteilen Sie für jede dieser Dimensionen, wie sehr diese dazu beiträgt, dass dieses Objekt in seiner Umgebung Ihre Aufmerksamkeit auf sich zieht (ein Stern bedeutet „gar nicht“, 5 Sterne bedeutet „sehr“). Wie stark zieht dieses Objekt in seiner Umgebung Ihre Aufmerksamkeit auf sich wegen seiner ...</i></p>	
v_cin	intensity of colour <i>Farbintensität</i>	f
v_mot	motion (e.g. flashing, flow) <i>Bewegung (z.B. Blinken, Fließen etc.)</i>	f
v_col	tone <i>Farbton</i>	f
v_loc	location (e.g. raised, very close to street) <i>Lage (z.B. erhöht, Nähe zur Straße)</i>	f
v_siz	size <i>Größe</i>	f
v_sha	shape <i>Form</i>	f
v_con	condition (e.g. new, dirty, etc.) <i>Zustand (neu, schmutzig, usw.)</i>	f
v_sig	signs attached <i>Schilder daran</i>	f
v_hei	height <i>Höhe</i>	f
v_wid	width <i>Breite</i>	f
v_len	length <i>Länge</i>	f
v_are	area <i>Fläche</i>	f
v_pat	pattern <i>Oberfläche</i>	f
v_mat	material (as far as identifiable) <i>Material (soweit erkennbar)</i>	f
v_age	<p>To what extent is this object salient as a result of how old it looks? <i>Wie stark fällt dieses Objekt wegen seines augenscheinlichen Alters auf?</i></p>	f

Table 9.: The questions used to formatively measure *visual salience*. Column Var gives the variable codes used, ToM the type of measurement used, where r means reflective and f formative measurement. Questions are given in the order they were presented to participants.

3 HOW CAN SALIENCE BE MEASURED?

company. This doctoral work takes buildings and non-buildings into account and these dimensions are suitable for buildings only, i.e. it is reasonable to neglect them. A large number of questions regarding *visual salience* must be answered by participants, consequently. In order to avoid fatigue in participants an adequate way to ask for ratings of these dimensions was examined in a pre-study (cf. section 3.3.4). Furthermore, it is important to note that participants were asked to visually estimate *v_mat* and *v_age*. That is, no knowledge of building material or knowledge of building constructional history is required. A fence might look like if it were built from marble but it was actually only a paint effect. Similarly, buildings may look like they were built in the 18th century but in fact may not be. Beyond these formative indicators, four reflective items (cf. table 10) were included to achieve an overidentified model¹⁹ and to enable a so-called convergent validity analysis (cf. section 5.4.3). All reflective indicators are phrased to acknowledge the effect *visual salience* has on observers.

Var	Question
v_loo	To what extent does the appearance of this object draw your attention? <i>Wie stark zieht dieses Objekt Ihre Aufmerksamkeit wegen seines Aussehens auf sich?</i>
v_odd	How unusual is the appearance of this object? <i>Wie ungewöhnlich ist das Aussehen dieses Objekts?</i>
v_eye	How eye-catching is this object? <i>Wie stark sticht dieses Objekt ins Auge?</i>
v_rec	How recognizable is this object? <i>Wie gut ist dieses Objekt wiedererkennbar?</i>

Table 10.: The four reflective questions used to enable a convergent analysis for *visual salience*. Column *Var* gives the variable codes used. Questions are given in the order they were presented to participants and these questions were presented to participants before any of the formatively measured variables were shown.

¹⁹ Bollen and Davis (2009a, 2009b) provide a thorough discussion of identification of models which make use of causal-formative indicators.

3.3.4 About the Phrasing of Questions

According to the advice given in (cf. DeVellis, 2012, pp. 73–113) the wording of questions had been developed iteratively. Having revealed the visual dimensions used in prior studies, a large pre-study ($n = 71^{20}$) was performed (cf. Kattenbeck, 2015b) as part of another study (cf. Ohm et al., 2016). The goal of this pre-study was to get an insight into whether or not participants are fatigued by recurring phrases in questions about visual dimensions. Furthermore, I tried to discover whether participants base their judgments on the local surroundings, even though the questions do not explicitly suggest to do so. A route of 1.1 km length through the Old Town of *Regensburg* chosen for the seminar work of students was used. Participants were led by student experimenters on this route and asked to rate eleven objects along the route. These objects were revealed to be salient to pedestrians in another pre-study and were randomly chosen from this larger set.

 Question

How noticeable is the appearance of this object? [global item]

Wie stark fällt das Aussehen dieses Objekts auf?

How noticeable is the color of this object? [col]

Wie stark fällt die Farbe dieses Objekts auf?

How noticeable is this object's location? [loc]

Wie stark fällt die Lage dieses Objekts auf?

How noticeable is the size of this object? [siz]

Wie stark fällt die Groesse dieses Objekts auf?

How noticeable is the shape of this object? [sha]

Wie stark fällt die Form dieses Objekts auf?

How noticeable is the condition of this object? [con]

Wie stark fällt der Zustand dieses Objekts auf?

How noticeable is this object due to a sign (commercial sign, company logo, etc.)? [sig]

Wie stark fällt dieses Objekt auf Grund eines Schildes (Reklametafel, Firmenlogo, usw.) auf?

Table 11.: The questions regarding visual dimensions during the pre-study. The questions are phrased in a general way, i.e. no reference to the local surroundings was made. Abbreviations used in fig. 9 are given in square brackets.

A qualitative analysis of the data uncovered three key findings.

²⁰ In fact, 19 out of $n = 90$ participants had to be excluded due to missing data.

3 HOW CAN SALIENCE BE MEASURED?

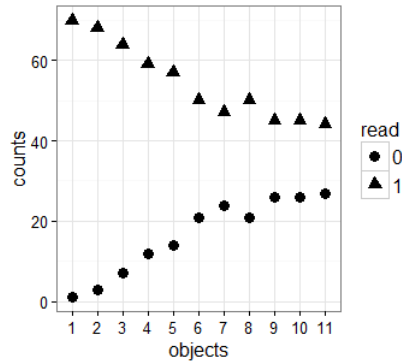


Figure 8.: Counts of answers to the question, whether participants read the whole question ($read = 1$) or not ($read = 0$). $n = 71$ due to missing data for 19 participants. The figure was taken from (Kattenbeck, 2015b, p. 316).

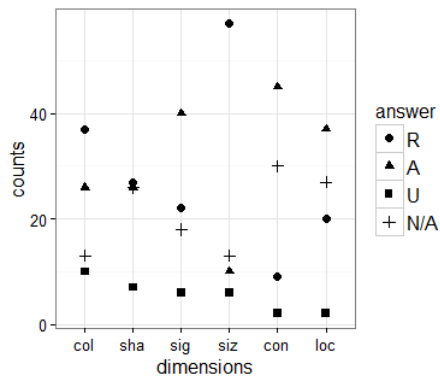


Figure 9.: For each of the visual dimensions the counts representing the reported way of judgment are given. R means *relative* and A means *absolute judgment*, U denotes cases where no distinction was explicitly made. N/As result from different sources, i.e. either technical issues or non-cooperative behaviour of participants. The figure is taken from (Kattenbeck, 2015b, p. 316).

SHORTER QUESTIONS ACCOMMODATE PARTICIPANTS.

Figure 8 shows a clear trend. Participants tend to stop reading all the questions word by word, i.e. there is a correlation between time and counts. For objects 1–6 the number of participants who do not fully read each of the questions increases. Conversely, the number of participants who read the whole question decreases. The explanations given reveal a precise reason. Participants tend to stop reading in case they feel to know the question well ($P(\text{well} - \text{known}|\text{no}) \geq 0.55$ as of object 3). As a consequence, a decision was made to present questions about visual dimensions using a star-rating system (cf. section 4.3.1). This presentation mode will be more suitable to avoid fatigue in participants.

TELL PARTICIPANTS TO BASE JUDGMENTS ON SURROUNDINGS.

The local environment was used inconsistently by participants. Figure 9 shows counts for each of the visual dimensions with exemption of the global item (cf. table 11). The wording did not explicitly encourage participants to use the local environment as a basis for judgments. The majority of users used local surroundings to establish dimensions, such as color and size, but not for signage, condition and location. This suggests that, explicitly encouraging participants to use the local surroundings will enhance the validity of the results. This is particularly true because it does not matter whether e.g. the tone of a building's façade is loud or rich in an absolute manner. It is the local environment that suffices to make an object salient, i.e. a feature needs to be outstanding from its local background (cf. section 2.4).

DISTINGUISH BETWEEN TONE AND INTENSITY.

A qualitative analysis of the comments given has revealed a major difference made by participants. Whereas some reported rating the tone others reported rating intensity for dimension color. As a result, it was separated into tone (v_col) and intensity (v_int) for the main study in order to avoid undesirable variance due to mixing these two different concepts.

3.4 CHAPTER SUMMARY

This chapter was concerned with the development of SEMs to measure the *salience* of objects. Beyond measurement model building, this includes an informed decision about experimental setup.

MAJOR SHORTCOMINGS OF APPROACHES

A thorough review of different approaches of empirically measuring *salience* revealed several major shortcomings. Typically, they are underpowered in terms of number of objects and users. Furthermore, they neglect non-buildings as stimuli and are not well grounded in theory about *salience*. In particular, features of objects revealed to be important in earlier studies are often not taken into account systematically.

IN-SITU EXPERIMENTS AS PRIMA INTER PARIA

To counterbalance the aforementioned shortcomings the decision was taken to conduct a large-scale in-situ experiment. It is based on a random sample of buildings and non-buildings with an a priori calculated sample size. The in-situ approach is as close as possible to real-world navigation scenarios in an experimental context. In particular, it counterbalances the known difference between memorizability and recognizability of objects.

RATIONALE TO USE SEM IN GENERAL AND PLS-PM IN PARTICULAR

SEMs show an unique advantage: The relationships between LVs on the one hand and MVs and the LVs they are associated with on the other can be examined simultaneously. Hence, SEM is a suitable means to get to know about the importance of different variables given the context of other variables. Using PLS-PM ensures that *visual salience* can be modeled formatively, as it allows to incorporate this type of measurement without building a MIMIC-model.

PLS-PM ALGORITHM

Having made the decision to use PLS-PM a detailed explanation of its algorithm was given. It is based on an iterative approach, alternating between outer and inner estimations of coefficients until convergence is achieved.

THE NATURE AND DESIGN OF MEASUREMENT MODELS

Based on dimensions contributing to *saliency* revealed in earlier studies, a measurement model was derived for each LV used to model *saliency*. Particular emphasis was put on *visual saliency* estimation because of its usefulness for different user groups. The visual features revealed in earlier studies are causal-formative indicators of *visual saliency*. Consequently, reflective indicators were added in order to allow for an analysis of convergent validity for *visual saliency*.

PRE-STUDY RESULTS

A pre-study revealed a star-rating like approach to be a proper way of asking questions on visual dimensions, therefore avoiding fatigue in participants. Moreover, the importance to distinguish between tone and intensity was found.

4 EMPIRICAL METHOD

4.1 INTRODUCTION

Based on the theoretical insights given above, the empirical method taken will be detailed in this chapter. It starts with considerations about a sample size in terms of objects which is suitable for a SEM analysis. Drawing on this argument, the method taken to derive a random sample of objects is presented. The way routes were built and sample size considerations regarding a suitable number of participants form the last part of this first section. The second section introduces the apparatus used to conduct the experiments. This includes the DB-Schema as well as the way stimuli and questions are presented. The last part of the apparatus section details the sense of direction scale used and provides a rationale for the demographic data collected. The third and final section of the chapter explains the procedure taken to collect data.

4.2 OBJECTS, PARTICIPANTS, ROUTES

Considerable effort was taken to establish a suitable sample size of objects to be rated. As this study tries to identify important dimensions for accurately modeling *salience*, it is crucial to collect ratings for a diverse range of objects. Consequently, sample size calculations are performed with respect to the number of objects needed.

4.2.1 *Considering a Suitable Number of Objects*

Generally, SEM analyses are known to be a high sample size technique (cf. Hair et al., 2010, p. 661). As a first step, the minimum number of

objects needed was estimated using the so-called “ten times” rule (cf. Barclay et al., 1995; Hair Jr. et al., 2014). Ensuring a sample of objects large enough to conduct a PLS-PM analysis yields a sample size of $n_{min} = 150$. This number equals ten times the maximum number of formative causes. It represents the lower bound of objects needed. In order to yield a more well-grounded sample size, an a priori sample size calculation was performed. As CB-SEM is known to require higher sample sizes than PLS-PM, the calculation was based on CB-SEM requirements. This also ensures the possibility of analyzing the data using methods other than PLS-PM. The following parameters were chosen:

EFFECT SIZE As this is the first time a SEM for *salience* is to be tested, a medium effect size ($eff = 0.3$) according to Cohen (1988) was assumed.

POWER By convention the level of statistical power was set to $\beta = 0.8$.

SIGNIFICANCE LEVEL The conservative level of $\alpha = 0.01$ was applied.

LATENT VARIABLES As mentioned above (cf. section 3.3), the number of LVs in the complete model equals $num_{LV} = 6$.

MEASURED VARIABLES The number of variables used in the measurement model equals $num_{MV} = 39$.

Using a sample size calculator provided online (cf. Soper, 2015) a required sample size of $n_{obj} = 360$ was calculated.

4.2.2 A Set of Randomly Selected Objects

Regarding the object selection, a wide variety of objects must be ensured because variance resulting from the inclusion of salient (cf. e.g. fig. 10) and non-salient (cf. e.g. fig. 11) objects is needed for statistical reasons. There is evidence that participants will choose objects unsuitable for use in PNSs if they are free in their choices. For example, Ishikawa and Nakamura (2012) report on participants who use cars parked next to the street, although they were instructed to ensure usefulness of objects for route instructions, which may be given to others in two or three days time.

4.2 OBJECTS, PARTICIPANTS, ROUTES



Figure 10.: The Cathedral is a good example of a salient building in *Regensburg's* Old Town. Its average salience rating was 4.11 out of 5. The picture was taken by the author.



Figure 11.: An example of a non-salient building with an average salience rating of 1.59 out of 5. The picture was taken by the author and a red frame was added in order to avoid misconceptions using GIMP (cf. The GIMP Team, 1997-2013).

Furthermore, for obvious reasons participants cannot be asked to name objects they feel are non-salient. Finally, it is advisable to include non-buildings (cf. e.g. fig. 12) because there is empirical evidence for their importance (cf. e.g. Schroder et al., 2011) as well as there are theoretical arguments for their inclusion (cf. e.g. Sadeghian & Kantardzic, 2008).



Figure 12.: A circuit breaker panel located next to street is an example of a non-building included in the sample. The picture was taken by the author and a red frame was added in order to avoid misconceptions using GIMP (cf. The GIMP Team, 1997-2013).

Considering these arguments, there is a need to select a random sample of objects including both buildings and non-buildings in an urban environment. In order to achieve this goal, an initial set of 400 geographical coordinates was generated using *GNU R* (cf. R Core Team, 2013) and the package *random* (cf. Eddelbuettel, 2015). Figure 13 shows the district those coordinates must be located in. For each of these random coordinates, the following approach was applied:

1. If a building was hit by at least one point, this building was chosen, i.e. multiple hits were counted as one hit.
2. Any points which hit properties not open to the public (e.g. gardens etc.) were neglected.

4.2 OBJECTS, PARTICIPANTS, ROUTES

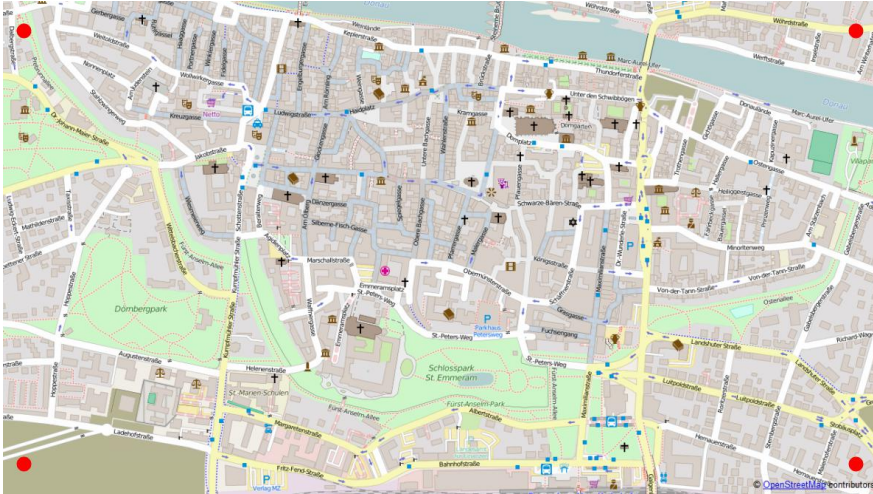


Figure 13.: The area randomly selected geographical coordinates had to be located in. The red dots give the *NE*, *NW*, *SE* and *SW* borders. They are given as $x_{min} = 1345000$, $x_{max} = 1347900$, $y_{min} = 6276900$, $y_{max} = 6278500$ in EPSG:3857. The figure is based on OpenStreetMap (cf. Haklay & Weber, 2008) and was created using QGIS (cf. QGIS Development Team, 2014).

3. Any points which hit the large river *Donau* were neglected with the exception of one such point, therefore including the river *Donau* as an object in the sample.
4. All other points were visited by the author in person. In order to select an object, a lot was drawn at each point, determining one of the cardinal directions *N*, *NE*, *E*, *SE*, *S*, *SW*, *W*, *NW*. A compass was used to point at the direction chosen and the closest object having a minimum of temporal persistence was chosen. This means, parking cars were neglected, but road construction zones were included.
5. A picture was taken of each object included in the sample.

Using this approach the 400 geographical coordinates sampled yielded 320 objects. Hence, another 80 coordinates were chosen randomly and the above procedure was repeated. This yielded $n_{rated} = 369$ random objects, overall.

4.2.3 Routes

Three major steps were taken in order to be able to assume the independence of ratings:

1. I have chosen the set of objects to be rated on a single route randomly.
2. A minimum walking distance of approx. 200 m between subsequent objects was required on a given route.
3. Within one trial, i.e. on a particular route, no more than six or seven objects must be rated.
4. Temporal effort was constrained to approx. 60 minutes to minimize participant fatigue.

The decision to include steps 2–4 resulted from a small pretest. The test showed that participants become frustrated when the distance between objects is too long. Furthermore, there was a noticeable increase in fatigue when participants were asked to rate 9 objects or more. Finally, a duration of approx. 60 minutes turned out to be an acceptable period of time for participants. Taking all of these arguments into account random tuples of objects were built, resulting in 330 of 366 objects distributed to $n_r = 55$ routes. For each tuple the actual routes to be walked were built manually, trying to build a circular course, if possible. The actual direction the route was to be walked was randomly chosen. Having built these routes, the remaining objects were assigned manually to routes. If more than one route was feasible, the actual route was chosen randomly. Having assigned all objects to a single route each route was again walked by the experimenter and photographs for each of the objects on the route were taken from the line of sight when walking the route. This step was necessary due to findings of Lee, Kwong, Pun, and Mack (2001), reporting on an evaluation of photos of LMs in navigation systems. For 20 routes one of its objects was already part of another route as routes may have route segments in common. For each of these routes, random coordinates were generated within the area given by the route. Again, the procedure described above (cf. page 78) was applied to choose an object. Route IDs, starting points and number of objects to

be rated in routes are given in table 12. Figure 14 shows a sample route and its corresponding object locations.

Meeting point	route count	avg. objects
Alter Kornmarkt (an der Litfaßsäule)	1	7
Altes Rathaus (vor dem Eingang zur Touristinformation)	1	6
Arnulfplatz (vor der Arnulf-Apotheke)	2	7
Bahnhofsvorplatz (unter der großen Wanduhr)	5	7
Bismarckplatz (vor der Zufahrt zur Tiefgarage)	6	7
Café am Peterstor (am Peterstor 1a)	5	6
Dachauplatz (beim zur Zeit leeren Brunnen)	9	7
Dachauplatz (vor Restaurant Scholz)	5	7
Emmeramsplatz (vor dem Schlosseingang)	2	6
Ernst-Reuter-Platz (vor dem Europa-Brunnen)	4	7
Fischmarkt (vor den steinernen Fischbänken)	2	7
Neupfarrplatz (vor der Sparkasse)	2	6
Pustetpassage (neben dem Eingang zu Bücher Pustet)	2	7
Restaurant Jalapenos (Schottenstraße 4)	6	7
Stadtamhof (am Fuß der Steinernen Brücke)	2	7
vor der Wurstkuchl	1	7

Table 12.: All starting points for the 55 routes built. *route count* gives the number of routes starting at this particular meeting point. *avg. objects* gives the mean number of objects for the routes.

4.2.4 Number and Acquisition of Participants

To counterbalance random error, a minimum number of two ratings per object from two different persons was collected. Any calculations were then based on the mean of both ratings. As 55 routes were randomly built, $n_{part} = 110$ participants are needed. This procedure will foster the validity of the PLS-PM results because the influence of personal preferences will be reduced due to the large sample size. Participants were recruited by the author in person by visiting university lectures of several different disciplines or personal contact. In order to minimize bias stemming from specific personal requirements potential participants must not suffer from color-blindness or any physical disabilities. They were required not to have taken part in prior experiments about pedestrian navigation, too (which may have been the case due to the strong focus on

4 EMPIRICAL METHOD



Figure 14.: The course of route 22 and the geographical coordinates of object locations. It started at the *Café am Peterstor (am Peterstor 1a)* and was 1.28 km long. The figure is based on OpenStreetMap (cf. Haklay & Weber, 2008) and was created using QGIS (cf. QGIS Development Team, 2014).

pedestrian navigation the Chair for Information Science at the University of Regensburg has). A Samsung Galaxy Tab 4 10.1 WiFi was raffled off between all participants. Moreover, in case they were university students, they were offered course credit for participation²¹. Persons willing to participate were asked to provide their full name and e-mail address on a list. Participants were asked via e-mail to choose their preferred time slot using a doodle-survey (cf. Doodle AG, 2016). This e-mail referred to the importance of the prerequisites and gratifications. A maximum number of four slots of 1.5 hours length was available per day, i.e. experiments took place between 09:30am and 04:00pm in order to ensure comparable daylight conditions. No timeslots were offered between November 21st, 2014 and December 02, 2014 or between December 21st, 2014 and January 7th, 2014. All appointments were confirmed one day ahead of time, naming the meeting point and kindly asking participants to e-mail the author if they worried about finding

²¹ The recruitment of participants was first described in (Kattenbeck, 2015a).

the meeting point. This was done in order to avoid an occasion of bad orientation performance before the experiment started. If participants were worried another meeting point would have been chosen²² in order to avoid biased answers to the SoD-survey participants were required to fill in at the meeting point (cf. section 4.4.2).

4.3 APPARATUS

This section provides details concerning the application used to conduct experiments and the SoD and demographic data survey.

4.3.1 *Application*

I adapted a derivate of an early prototype smartphone application of the navigation project called URWalking at the University of *Regensburg* (cf. Kattenbeck et al., 2013) for the purpose of data acquisition. Originally, most of the client-side application was developed by Robert Jackermeier. Data communication mechanisms were joint work of Manuel Müller, Christina Ohm and the author based on code provided by Richard Schaller. The application was run on Android 4.4 on a Google Nexus 7 (2013) tablet. The application was connected to the Internet using a LTE connection. Participants were free to input data using a stylus provided by the experimenter or use their fingers. The tool was able to load questions and further experiment data from the DB described below. Furthermore, the answers provided were stored in an XML format. In principal, the application has the capabilities to send these files to the server automatically via an LTE connection. In order to prevent loss of data, however, a copy of all XML-files was additionally stored on the device. In fact, these files were manually uploaded once all experiments were conducted.

DB-Design

The DB was initially designed by the author to enable a variety of features. Many of these features, including the possibility to store land surveying

²² In fact, only one participant wanted to choose another meeting point.

data, were not used in this thesis and will not be described²³. The part of the database most important for the current study is the one required to store survey data. The database allows to create surveys which have multiple questions per section. Each question can only have one scale, i.e. users can either choose from several scale points or, alternatively, the question requires a freetext answer. Participants can take part in several experiments. Several different surveys may be answered by a single participant during one experiment. All answers given to questions are registered as separate answer versions. The sample of objects used in this thesis is stored in the database including its geographical coordinates.

Presentation of Stimuli

As some of the questions of the *salience* questionnaire refer to naming, potential experimental bias resulting from textual references to objects had to be minimized. To address this issue, a picture was used to communicate to participants which object they were expected to rate next. The particular object was marked by a red frame (using GIMP (cf. The GIMP Team, 1997-2013)) to avoid potential misconceptions resulting from perspective. The object to be rated next was presented (cf. fig. 15) on a screen accompanied by a button “Rate now” (*Objekt bewerten*). A click on “Rate now” stored the current GPS-position, which was calculated based on a step detector implemented by Robert Jackermeier, in a separate file and displayed the first three questions of the questionnaire. In order to avoid a bias resulting from participants potentially rating the picture instead of the real world object, participants were not able to review the picture showing the stimulus once they pressed “Rate now”.

Presentation of Questions

Overall, $num_{allq} = 46$ questions were asked per object and only $num_{MV} = 39$ of these were used to assess the structural models throughout this thesis. The remaining five questions were asked in order to gain a specific

²³ For a thorough description of how land surveying data was included in the database Kattenbeck et al. (cf. 2013).

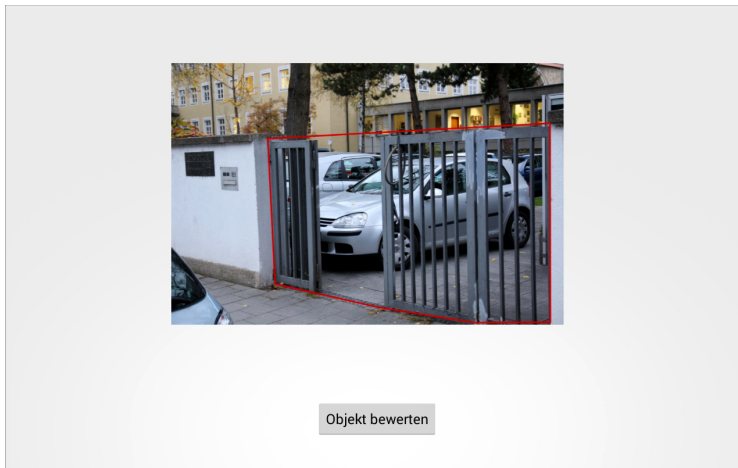


Figure 15.: An example screen which shows a picture with a red-framed object participants were required to rate next. The picture was taken by the author and a red frame was added in order to avoid misconceptions using GIMP (cf. The GIMP Team, 1997-2013).

name for the object etc. (cf. appendix A.1.2). It is important to note that an option “other reason [why the object is salient]” was not included for any of the questions used to assess models throughout this thesis. This decision is based on the results of Partala et al. (2010), who found that participants, if they are allowed to, make use of this option very often thereby reentering answers already possible by pre-compiled answers. Although several questions were placed on one screen (cf. table 13), participants were required to answer all questions on a given page before they were able to reach the next page of the survey. The item codes refer to table 3, table 4, table 5, table 6, table 7, and table 9 (cf. pp. 60–67). Participants were allowed to go back and forth between pages and change any of the answers given to any question in the current survey. However, once they had completed a survey and clicked on “Done” (*Fertig*) on screen 8, they were not able to review their judgments. Three different ways of presenting questions and scales to participants were applied.

Screen	Questions
1	ov_att ov_por ov_mem
2	p_cat p_exa p_img p_sim v_loo v_odd v_eye v_rec
3	v_cin v_mot v_col v_loc v_siz v_sha v_con v_hei v_sig v_wid v_len v_are v_pat v_mat
4	v_age s_dis s_vis s_per
5	s_pic
6	s_eas s_sui s_lor s_imp s_dir
7	c_per c_his c_wor c_cus c_cst c_pus c_pst
8	c_hid c_kid c_oid c_eas

Table 13.: The way questions used to measure variables were spread across different screens. Bold-faced question codes refer to the tables given on pp. 60–67. Non-bold questions were not included in the measurement model (cf. appendix A.1.2).

SENSE OF DIRECTION QUESTIONS

In line with Münzer and Hölscher (2011) a seven-point rating scale is presented to users (cf. fig. 16). Only the leftmost and the rightmost scale points are named, where the left point reads “completely reject” (*lehne stark ab*) and the very right scale point “completely agree” (*stimme stark zu*).

Bitte beantworten Sie folgende Fragen zu Ihrem Orientierungssinn:

Wenn ich durch eine unbekannte Stadt laufe, dann weiß ich, aus welcher Richtung ich gekommen bin und in welche Richtung ich mich bewege.

lehne stark ab stimme stark zu

Wenn mich jemand in meiner Stadt nach dem Weg fragt, dann stelle ich mir meine Stadt wie auf einer Karte vor und ermittle daraus den Weg.

lehne stark ab stimme stark zu

Wenn ich mich durch ein großes Gebäude bewege, dann stelle ich mir dabei eine Art Plan oder Grundriss (Überblicksansicht) vor.

lehne stark ab stimme stark zu

Ich bin sehr gut darin, von meinem gegenwärtigen Standort aus Richtungen zu anderen Orten

Figure 16.: A seven-point rating scale was used to assess SoD according to the so-called *Fragebogen räumliche Strategien* (cf. Münzer & Hölscher, 2011).

SALIENCE SURVEY

In principle, participants were asked to provide answers to rating questions on a five point rating scale. The literature suggests that, although quite common, questions requiring participants to agree/disagree with a statement (e.g. "This object is eye-catching." (*Dieses Objekt sticht ins Auge.*)) lead to biased results because "they encourage acquiescence, the tendency to agree irrespective of item content" (Schaeffer & Presser, 2003, p. 80). Consequently, direct questions (e.g. "How eye-catching is this object?" (*Wie stark sticht dieses Objekt ins Auge*)) were asked and ratings were based on an intensity scale provided by Rohrman (1978). The five scale points are: "not at all" (*gar nicht*), "rarely" (*kaum*), "medium" (*mittelmäßig*), "considerable" (*ziemlich*), "very much" (*sehr*). As long as no scale point was selected, only the leftmost and rightmost scale points were given (cf. fig. 17). Once participants had selected a scale point, the corresponding value was displayed to right of the scale.

VISUAL DIMENSIONS

On screen three participants were asked to evaluate 15 visual dimensions with respect to the degree each of them contributes to this object's *salience* given its local environment (cf. fig. 18). A pre-study (Kattenbeck, 2015b) suggested, that participants get increasingly frustrated, if equally phrased questions are presented for each of the visual dimensions. Hence, a general introduction to these was given and participants were asked to use a star-rating-like approach for each of the dimensions. The way of presenting the names of scale points remained unchanged.

4.3.2 *Measuring Sense of Direction*

Several ways of assessing a person's sense of direction have been proposed (cf. the literature review given in Hegarty et al. (2002, pp. 426–428)). Of these methods, a self-report SoD scale is particularly suitable given the in-situ setting and the time constraints in terms of overall experiment time needed. Two up-to-date scales available in German

4 EMPIRICAL METHOD

Wie stark zieht dieses Objekt Ihre Aufmerksamkeit auf sich?
gar nicht sehr **ziemlich**

Wie gut eignet sich dieses Objekt als Bezugspunkt?
gar nicht sehr **kaum**

Wie gut können Sie sich dieses Objekt einprägen?
gar nicht sehr

← Zurück Seite 1/8 Weiter →
2/3 Fragen beantwortet

Figure 17.: Saliency survey questions measured using a rating scale were presented using radio buttons. On selection the meaning of the chosen scale point was displayed next to the meaning of the rightmost scale point.

Nachstehend finden Sie eine Reihe visueller Aspekte. Bitte beurteilen Sie für jede dieser Dimensionen, wie sehr diese dazu beiträgt, dass dieses Objekt in seiner Umgebung Ihre Aufmerksamkeit auf sich zieht (ein Stern bedeutet "gar nicht", 5 Sterne bedeutet "sehr").

Wie stark zieht dieses Objekt in seiner Umgebung Ihre Aufmerksamkeit auf sich wegen seiner...

Farbintensität ★★★★★ ziemlich

Bewegung (z.B. Blinken, Fließen etc.) ★★☆☆☆ kaum

Farbton ★★★★★ ziemlich

Lage (z.B. erhöht, Nähe zur Straße) ★★★★★

Größe ★★★★★

← Zurück Seite 3/8 Weiter →
3/14 Fragen beantwortet

Figure 18.: Questions regarding *visual saliency* other than (visual) age must be rated using convenient a star-rating system.

could have been used. The well-known Santa Barbara Sense of Direction Scale (Hegarty et al., 2002) was initially developed in English language. There is, however, no clear decision concerning its applicability using a German language translation. Whereas Montello and Xiao (2011) provide empirical evidence for reliable and valid translations to several languages, Münzer and Hölscher (2011) doubt the coherence of the German version. They developed a German language scale called *Fragebogen räumliche Strategien* (FRS) (cf. appendix A.1.1). This scale was preferred over the German translation of the Santa Barbara Sense of Direction Scale due to the unclear status of the empirical evidence at that time. The FRS consists of 19 items and comprises three factors: knowledge of cardinal directions, allocentric strategies to mental mapping and strategies to egocentric orientation combined with the global self-assessment of a person's SoD.

4.3.3 Demographic Data

To control possible confounding variables, participants were asked to provide demographic data. All questions asked aimed to minimize the bias in *salience* estimation of objects. The following list provides the dimensions and a rationale, why these may be important (cf. appendix A.1.3 for the German language questions used to measure these).

SEX Due to the known differences with respect to spatial cognition between women and men (cf. Colluccia & Louse, 2004), women and men may differ in the way they judge salience, too.

YEAR OF BIRTH Age may be an important confounding variable, because of evidence for strong differences in orientation abilities (cf. e.g. Jansen-Osmann, Schmid, & Heil, 2007). These may have a particular impact on *structural salience* estimation.

COUNTRY OF BIRTH Participants not born in Germany may be used to environments shaped differently to European cities.

KNOWLEDGE ABOUT PLACE A person's knowledge about a local neighborhood may have an effect on several dimensions of *salience*. For this reason, Ishikawa and Nakamura (2012) use a neighborhood

completely unfamiliar to participants. I use a different approach explicitly asking participants how familiar they are with *Regensburg*. Participants with intimate knowledge of the city may well judge objects differently to those who know less. For example, whereas the former may recognize a building as the location of the local observatory and, consequently, associate a particular meaning with this building, the latter may not be able to create any association.

FREQUENCY OF PEDESTRIAN NAVIGATION IN REGENSBURG

Participants must provide data about the frequency with which they walk through *Regensburg*. Those walking often may be more affine to objects potentially useful for the explanation of routes.

FREQUENCY OF USAGE OF MOBILE PEDESTRIAN NAVIGATION SYSTEMS

Participants who make frequent use of PNSs may have encountered LM-based navigation. As a consequence they may think differently about the degree to which an object may be useful in route instructions.

BORN IN OR RAISED IN REGENSBURG Familiarity is known to play an important role with respect to spatial cognitive style (cf. Piccardi, Risetti, & Nori, 2011). Participants, who grew up in *Regensburg* may be far more familiar with the city than persons, who moved to the city, consequently. Moreover, evidence was found for the influence familiarity has on the preference of certain types of LMs (cf. Quesnot & Roche, 2015).

CURRENT LOCATION PERSON LIVES AT Participants who live in *Regensburg* may have a more detailed knowledge about the city. Therefore, their CM may differ from those persons not living in *Regensburg*.

COURSE OF STUDY OR CURRENT JOB POSITION A person's profession or course of study may have an impact on the way visual and cognitive dimensions are rated. For example, architects may have a perspective on the shape of buildings which is different to the one of computer scientists.

4.4 PROCEDURE

4.4.1 *Assigning Routes Randomly*

Routes were randomly assigned to participants by the end of the week prior to their participation. This method ensured that each route was walked once before any of the routes was walked for the second time. If someone declined to participate after receiving the confirmation e-mail, her/his route was considered for the next week's random assignment. During the last week of experiments, all routes were reassigned randomly to all remaining participants if someone declined to participate. As some of the routes crossed squares, which are crowded in December due to Christmas markets the first trials for the 12 routes affected were conducted between November 12th, 2014 and November 20th, 2014.

4.4.2 *Conducting Experiments*

A research log was kept, including all instances where planned participants declined to take part, reasons for the declines and particular circumstances of trials (cf. appendix A.3.2). All participants were guided by the author. At the meeting point participants were instructed on the reasons why personal data and SoD-data were gathered (cf. appendix A.3.1). Furthermore, participants were advised that they were guided on the route and do not have to remember the route. Having filled in both surveys, the picture depicting the first of the objects to be rated was used to give further instructions. While walking along the route, participants were required to identify the object highlighted by a red bounding box on the picture in the local environment. If they spotted the object and were able to answer a survey concerning the object given its surroundings, the participant and I would stop and they were asked to complete a survey regarding this particular object. Participants were made explicitly aware that they need not stop anywhere near the location from which the picture was taken but where they were sure to be able to rate the object in its local environment. This methodological setup aligns with the findings of Xia et al. (2011), where visibility was found to be most important for LMs. Furthermore, participants were precisely instructed not

to rate the photo shown, but base their ratings solely on the actual object in its real environment. Finally, they were told that they had to carry the tablet because of a localization system implemented to keep track of locations participants chose to stop which relied on counting the steps taken. In the case participants had questions regarding the procedure, these were answered by the experimenter before the participant started their walk. During the rating of the first object, several explanations regarding some of the questions were given (cf. appendix A.3.1). Once the first object was rated, participants were told that distances between objects may differ. The overall number of objects to be rated was named upon request. Once the last object was rated, participants were logged out automatically and informed about future e-mails to be sent by the experimenter, regarding the results of the study and the tombola. All details provided may be found in appendix A.3.1.

4.5 CHAPTER SUMMARY

This chapter has provided a thorough explanation of the empirical method taken in this thesis. The main contributions concern the following topics:

SAMPLE SIZE

Rigorous methods for sample size calculation were applied. Instead of typical rules of thumb for PLS-PM, the sample size was calculated to fit conservative CB-SEM needs. As 6 LVs formed the structural model and $num_{MV} = 39$ MVs were used, a sample size of $n = 360$ objects is applicable. Due to avoid random error in data, each object must be rated by two participants, at least. As 55 routes were randomly built, $n_p = 110$ participants are needed.

SAMPLING OBJECTS AND ROUTES

Considerable empirical effort was taken to generate a random sample of objects. 480 random geographical coordinates were created and the building hit or the closest object located in a randomly drawn cardinal direction was used. This resulted in a set of $n_{rated} = 369$ objects to be rated. Tuples of objects were built, leading to 55 routes, overall.

PERSON-RELATED MEASUREMENTS

The German language self-report scale provided by Münzer and Hölscher (2011) was used to measure SoD. Regarding demographic data, beyond standard measures, such as age and sex, knowledge about the place among others was assessed in order to control these potentially confounding variables.

APPLICATION AND DATABASE DESIGN

A derivate of a project-related Android application was used as a basis to develop an application suitable for data acquisition. Surveys were loaded from a remote DB. The application used pictures to communicate stimuli to participants. Objects were highlighted using a red bounding box on the pictures presented. Questions asked in the *salience* survey had to be rated on a five point rating-scale, where only the meaning of the leftmost and the rightmost scale point was given and the selected scale point was made explicit.

PROCEDURE

Recruitment of participants was done in lectures not given by the experimenter and via personal contact. Participants were free to choose one of four 1.5 hour timeslots available per day. All participants were personally guided by the author of this thesis. Routes were randomly assigned to trials. At the starting point, participants had to fill in both the SoD and demographic data survey. Participants were then guided on the route and had to identify the object presented on the screen in the environment. They were asked to fill in the survey from a position where they were able to view the object in its local environment.

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.

BOX AND DRAPER (1987, p. 74)

5

EMPIRICAL RESULTS

5.1 INTRODUCTION

In this chapter empirical results of the study will be presented. First, some descriptive statistics regarding trials and participants, objects, and routes are given. Second, the severity of the influence snowy weather had on data quality will be assessed, as this weather condition may have had a particular large impact. The main part of this chapter is dedicated to the estimation and fit of the proposed measurement model, a derivative of the measurement model for *visual salience* and two different structural models combining the LVs. In principle, three different approaches may be distinguished in use of SEMs (cf. Hair et al., 2010, pp. 646–647).

CONFIRMATORY MODELING STRATEGY

Employ a single model and test whether the relationships proposed provide an adequate model fit.

COMPETING MODELS STRATEGY

Compare several models with each model corresponding to another, often competing theory.

MODEL DEVELOPMENT STRATEGY

Use a rather loose framework of basic model setups and develop a model by iteratively employing the estimation results to change the original model.

While all of these approaches provide viable results, the competing model strategy is particularly useful because “[i]t has been shown that for any proposed structural equation model, at least one other model exists with the same number of parameters but with different relationships portrayed that fits at least as well as the proposed model” (Hair et al., 2010, p. 647). The SEM application strategy employed to analyze

data throughout this thesis utilizes a mixture of model development and competing models comparison. Accordingly *model I*, which makes extensive use of assumptions of independence and was presented in section 5.3, will be fitted and its fit will be assessed. These results are used to adapt *visual salience* estimation and to hypothesize on mediation. Consequently, a so-called higher-order component model for *visual salience* and a structural model making extensive use of mediation will be developed and assessed.

5.2 DESCRIPTIVE STATISTICS

In order to describe participants, objects investigated and routes, some descriptive statistics are given first. A particular focus is placed on the influence snowy weather had on ratings of objects. This is of major importance as snow cover may cause a severe bias on any *visual salience* dimensions.

5.2.1 *Participants and Trials*

As mentioned in section 4.2.4, the maximum number of trials per day was fixed to four in order to ensure comparable daylight conditions. All trials were performed by the author. Overall, experiments were conducted on 48 days between November 12th, 2014 and February 19th, 2015 with a median of $\bar{x} = 2$ trials per day (cf. table 46 in appendix A.2). On average, trials had a duration of $\bar{x} = 60 \text{ min}$ ($SD = 12 \text{ min}$, $MIN = 38 \text{ min}$, $MAX = 113 \text{ min}$). Only 9 of the 12 trials, which were affected by Christmas markets (cf. section 4.4.1) were conducted between November 12th, 2014 and November 20th, 2014 due to the fact that participants canceled appointments. The remaining three trials were shifted to days after the Christmas holidays. Overall, $n_{raw} = 119$ persons participated. Eight out of the 119 trials had to be excluded from the analysis due to technical issues. More precisely, on six occasions one or more buildings were not shown during the trial leading to incomplete data. The seventh case resulted from massive discontinuity of cell reception. Overall, $n_{def} = 112$ trials were used for the analysis. $n_{fem} = 68$ of them were females and $n_{non-stud} = 20$ of them were non-students. Each of the

$n_r = 55$ routes was walked by two participants, i.e. each object was rated two times by two different raters. On occasion, one route was walked three times with all the trials usable. The age of participants ranged between 18 and 65 years ($\bar{x}_{age} = 25.46$ years) and only four participants were non-native speakers. However, these had a native-like proficiency of German language. Participants turned out to be a fairly homogenous group in terms of SoD ($\bar{x} = 4.10, \tilde{x} = 4.16, SD = 1.00$). Similarly, their knowledge of *Regensburg* ($\bar{x} = 3.30, \tilde{x} = 3.00, SD = 0.93$) and the frequency they walk through *Regensburg* did not vary greatly ($\bar{x} = 3.46, \tilde{x} = 3.00, SD = 1.06$). The vast majority of participants were currently living in *Regensburg* ($n_{liv} = 81$), whereas only two grew up in *Regensburg*. As it turned out, participants used PNS on mobile devices rather infrequently ($\bar{x} = 2.02, \tilde{x} = 2.00, SD = 1.05$). In summary, potentially confounding variables are not a threat to the reliability of the results as the participants were a relatively homogenous group with respect to these. Whether inequalities with respect to gender and age exist may be examined in detail in the future (cf. section 7.3).

5.2.2 Objects

During the experiments, nine objects of the $n_{rated} = 369$ rated objects turned out to be problematic for different reasons.

ONE RATING On one occasion only one rating was received for a particular object. The technical reasons could not be found.

CONFUSING On three occasions participants reported a severe confusion about the location of the object given the current route. Ratings for both objects were neglected during analysis in order to avoid bias in *structural salience* questions.

PHOTO For two objects photos were accidentally taken from the wrong side of the tracks. This means, the photo showed the object from the opposite direction to that, which it was approached from. As a consequence, participants were reportedly confused and both objects are excluded from the analysis in order to avoid bias in *salience* ratings.

DOUBLETTES When building the routes and taking the pictures of objects the same objects were twice unintentionally included from different angles. This happened on three occasions and did not become clear until the first walk of each of the routes. Consequently, which of the doublettes was included was chosen randomly during data analysis.

Please note: For the sake of comparability of trials none of the affected objects were removed during the experiment phase. This means, each trial per route was conducted using the same number of objects. The exclusion of objects was done during data analysis, only. Hence, it was based on an overall number of $n_{obj} = 360$ objects. On average, each participant was required to rate $\bar{x}_{nobj} = 6$ objects on a given route. Table 14 provides an overview of categories the objects belonged to. Buildings and non-buildings were included in order to counterbalance one of the major shortcomings in existing studies (cf. section 3.2.4). The

No.	category	count
1	Shipping Pier	1
2	Squares	1
3	Stairs	1
4	Ships	1
5	River	1
6	Non-Permanent Structures	2
7	Fountain	2
8	Alley	3
9	Bus Stop	3
10	Bridge/Tunnel	3
11	Part of a Building	4
12	Statues and Monuments	7
13	Green Spaces	8
14	Trees	9
15	Car Parks	14
16	Street Furniture	70
17	Single Building	232

Table 14.: Categories of objects included in the sample. The counts show the situation where certain objects were removed already.

majority of the categories (i.e. numbers 2, 6, 10, 11, 12, 13, 15, 16, 17) was taken from Schroder et al. (2011, p. 433). The counts indicate that buildings were most frequent, as the experiments were conducted in a city center. The remaining 120 objects were non-buildings. Street furniture (e.g. walls, fences, post boxes, street lamps) subsumes lots of different objects and is the second most frequent category, overall. Those categories, which were not derived from existing literature, are rather infrequent, but not all of them are specific to the particular town in which the experiments were conducted. Alleys, fountains, bus stops and stair(way)s are to be found in many city centers. Similarly, trees may be found in all kinds of parks and other spaces. There are, however, some peculiarities because they are used as climbing trees for children (cf. fig. 19) or because individual trees became part of the sample due to their species, which was stressed using signs (cf. fig. 20). However,



Figure 19.: The climbing tree located in the *Dörnbergpark* is one example of specific trees included in the sample. The picture was taken by the author.

some of the categories are incorporated due to peculiarities of the local environment, i.e. numbers 1, 4, 5 are sampled as a direct consequence of the river Danube being part of the town. Table 15 provides an overview of the time needed to answer a survey for a single object. The values presented were calculated for the second through seventh object only. This means, the first object of each trial was neglected, because this object



Figure 20.: The black walnut tree located in the *Dörnbergpark*, too, is another example of specific trees included in the sample. The picture was taken by the author.

was accompanied by explanations given by the experimenter/author. The figures indicate an increasing speed in giving answers, most likely occurring from the fact that questions were repeated for each object. Furthermore, the standard deviations are quite similar. This suggests that participants acted consistently with respect to time, i.e. quickly answering persons remained quick and vice versa.

Regarding the survey length, one may question the utility of more than 40 questions per object. However, none of the participants complained about the length of the survey. There are three explanations for this positive result: One can be found in the thorough pre-study about the phrasing of questions regarding *visual salience*, which revealed a star rating system as a reasonable approach (cf. section 3.3.4). Second, splitting the whole survey across several screens is advisable. As the number of questions per page differed (cf. section 4.3.1) the length of the survey was less obvious to users. Third, the conversation between

	2nd	3rd	4th	5th	6th	7th
mean	302	269	263	240	239	233
median	290	250	250	228	224	220
sd	84	79	81	78	76	77

Table 15.: Mean, median and standard deviation of time needed to finish a survey for one particular object in seconds. In order to minimize bias the first object was neglected because several explanations were given by the experimenter/author whilst it had been rated.

experimenter and participants had a positive effect, as the topics were different to those the experiment was about. Thus, the walking distance between subsequent objects became recreational.

5.2.3 Routes

On average, each of the routes comprised a walking distance of 1.50 km ($SD = 0.36$ km, $MIN = 0.74$ km, $MAX = 2.360$ km). The routes were not mutually exclusive, but no pair of routes was equal. This means, two different routes may have had route segments in common but different objects were to be rated on these routes. Routes were of comparable complexity regarding their number of decision points. On average, a route had $\bar{x}_{dp} = 1.22$ decision points per 100m route length ($SD = 0.35$, $MIN = 0.51$, $MAX = 2.00$). In order to get an impression for the difference, fig. 21 shows the least complex route in this sense and fig. 22 shows its most complex counterpart.

Table 47 in appendix A.2.2 provides a detailed overview of each route, including the number of objects to be rated on this route, the starting point of the route, its length and the mean duration of its trials.

5.2.4 Influence of Weather Conditions on Ratings

In case of heavy rain, snowfall or fog trials did not take place. This was in order to minimize bias due to vision constraints. Nevertheless, five of the trials took place using an umbrella during the whole trial and on four occasions an umbrella had to be used for parts of the route (cf. the re-

Based on the argument that differences stemming from weather conditions are inevitable I try to examine whether the *snow sample* ratings are more different from the *non-snow sample* ones than the *non-snow sample* ratings are in itself. Therefore, the bootstrapping procedure was run twice (cf. fig. 23).

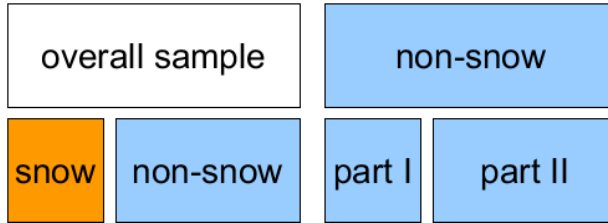


Figure 23.: The way subsamples were built for the bootstrapping analysis of snowy weather conditions. Figure was drawn using yEd (cf. yWorks GmbH, 2016).

In the first run, stratified resamples were drawn using *snow* and *non-snow* as strata. In the second run the *non-snow sample* was split into two subsamples. The sample sizes were chosen to mirror the fraction of *snow* (part I) to *non-snow* sample sizes (part II). Splitting the *non-snow sample* was done in order to avoid bias eventually resulting from different sample sizes when using the *non-snow sample* as a whole. For each iteration new split samples were found. For each of the $k = 19$ visual dimensions used, each of the $B = 10000$ iterations comprised the following two steps:

1. Draw a resample from each subsample having the same size as the current subsample with replacement.
2. Compare the means of the current subsamples using *Hedges' g* to find an unbiased estimated of the true effect size δ , thereby assuming homogeneity of variance, i.e. using the pooled variance (cf. Cumming, 2012) to find *Hedges' g*.

The results are shown in fig. 24. BCa-CIs for each of the visual dimensions are shown by means of a so-called forest plot. The blue error bars show the mean value, lower and upper bound of the 95%-BCa-CI for *Hedges' g* values in the *snow/non-snow* condition, whereas the green error bar reflects the *non-snow/non-snow* case. The variable names correspond

to section 3.3.3. The approximately equal mean of *Hedges' g* values for the *non-snow* condition is to be expected given the random splits applied in this condition.

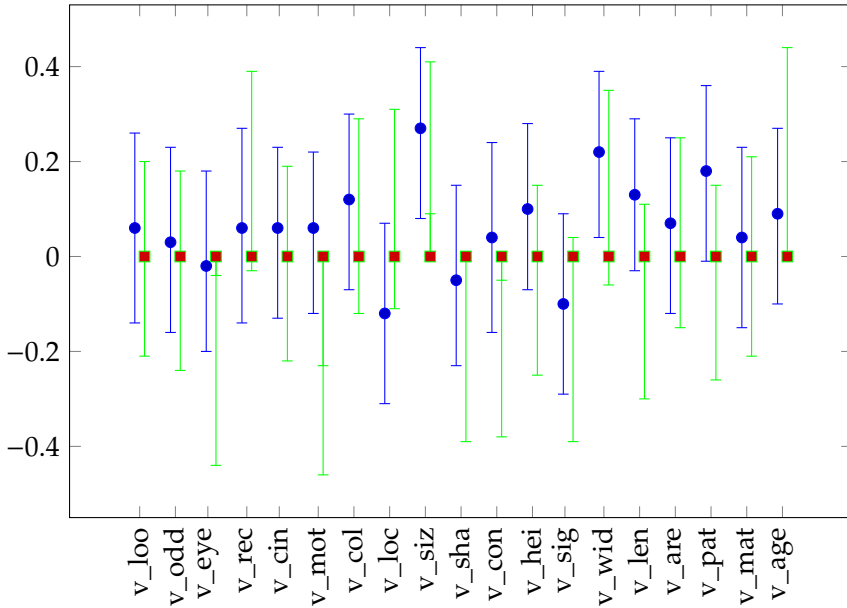


Figure 24: A forest plot showing CIs of *Hedges' g* values comparing the *snow* and *non-snow* sample ($K = 10000$ resamples). For each dimension the *snow* sample error bar is shown in blue and the *non-snow* error bar is shown in green. The figure was created using TikZ (cf. Feuersänger & Tantau, 2015).

It is interesting to note, that on some occasions the random splits for the *non-snow* condition yield an average *Hedges' g* value higher than the upper bound of the BCa-CI. This is an indication that some objects were rated particularly high for this variable. Following the suggestions of interpretation given by Kline (2013, p. 45), the overlap of the BCa CIs suggests there is empirical evidence that the snow cover did not have a more severe effect on ratings than other weather conditions had. There are, however, three exceptions to this general impression. *v_mot* and *v_con* differ significantly between conditions. Regarding the latter, snow may have had a positive impact, i.e. its white color may have enhanced the impression of good condition. Why *v_mot* was affected is rather

difficult to explain. Random effects may, however, play a role, as for this variable the mean of the BCa-CI in the non-snow condition is much higher than the upper bound of the CI. This is a strong indicator for existing outliers regarding this variable. Finally, a third large difference is found for variable *v_eye*. This may be attributed to the fact, that the pictures shown to participants were not taken during winter time. Accordingly, some features of objects may have been hidden by snow others may not and both may have had an impact on how eye-catching an object is. In sum, only 3 out of 19 variables show a rather large difference between conditions. In combination with the fact, that only 65 ratings overall (i.e. approx 8%) were done during the snow-cover phase, I take the result of this analysis as justification for the inclusion of the data collected between January, 30th 2015 and February, 3rd 2015.

5.3 A FIRST STRUCTURAL MODEL

Given this description of the data on which the analysis is based this section presents an initial structural model for the connections between relevant LVs.

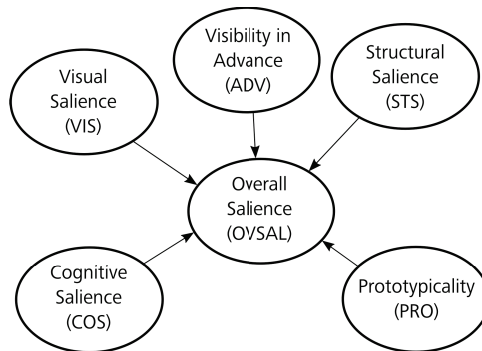


Figure 25.: The proposed structural model part depicting how *overall saliency* is modeled to be formed by *visual saliency*, *structural saliency*, *cognitive saliency*, *prototypicality* and *visibility in advance*. This model is referred to as *model I*. The figure was drawn using Inkscape (cf. The Inkscape Team, 2016).

Figure 25 shows that this model makes extensive use of the assumption of independence (cf. section 5.5.2 for an alternative model). I will

refer to this model as *model I*. The goal is to keep the initial model as simple as possible to gain a first insight into its ability to predict *overall salience*. The structural model proposes *overall salience* to be formed by *visual salience*, *structural salience*, *cognitive salience*, *prototypicality*, *visibility in advance* i.e. all of these are hypothesized to have a positive influence on *overall salience*. This means that *overall salience* is supposed to be an endogenous construct formed by the exogenous LVs *visual salience*, *structural salience*, *cognitive salience*, *prototypicality*, and *visibility in advance*. This is reasonable because all current approaches use a multi-dimensional vector to model *salience* per se.

Based on the explanations on measurement models given in section 3.3.3, the overall model is presented in fig. 26. The LVs proposed are given depicted using circles, whereas rectangles are used for MVs. For the sake of clarity, only the formative model part of *visual salience* is shown, as the reflective items are used for the convergent validity assessment, only (cf. section 5.4.2). All other constructs are measured in a reflective manner as indicated by the direction of arrows.

5.4 ESTIMATING STRUCTURAL EQUATION MODELS USING PLS PATH MODELING

This section presents the estimation of *model I*. All major steps and scores to assess the model's qualities are detailed theoretically and applied to *model I*.

5.4.1 General Steps and Parameters

Using PLS-PM for model assessment comprises four steps.

1. The whole model must be estimated.
2. The reflective models are assessed in order to ensure thorough measurement of the LVs involved.
3. The formative model part is examined in order to show that this type of measurement is appropriate for *visual salience*.

5 EMPIRICAL RESULTS

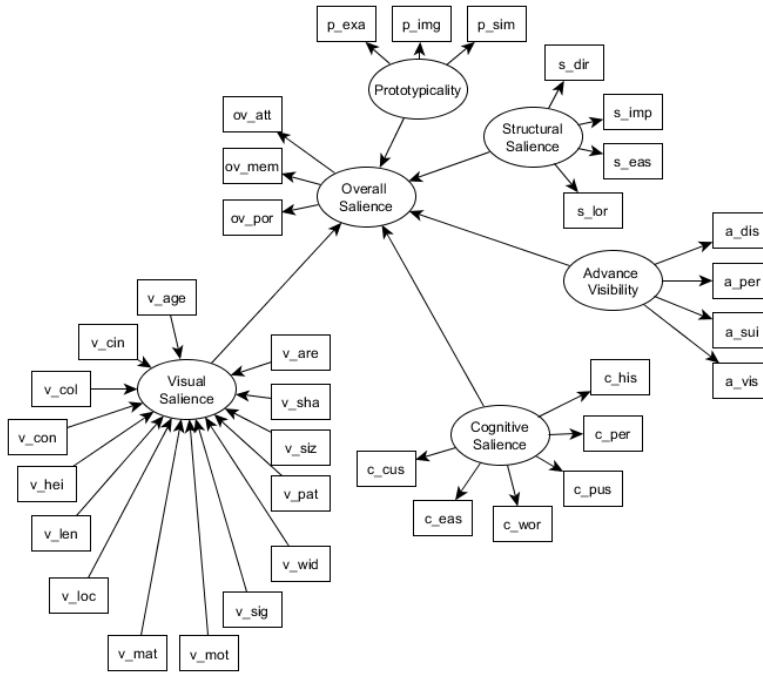


Figure 26.: The first structural equation model analyzed, depicting LVs using circles and MVs using rectangles. The figure was created using yEd (yWorks GmbH, 2016). The variable codes adhere to those introduced earlier (cf. section 3.3.3).

4. The assessment of the structural model will reveal the predictive quality of the model proposed

All of these steps were conducted using SmartPLS software (Ringle, Wende, & Becker, 2015). In order to estimate the model, several parameters must be chosen to run the algorithm (cf. section 3.3.2). The values used are based on the advice given in Hair Jr. et al. (2014, p. 82).

STANDARDIZATION The estimation of the model can either be based on raw or z-standardized values. The latter was used to ease the interpretation of values.

WEIGHTING SCHEME The weighting scheme used has a major impact on the inner estimation stage. As formative and reflective measurement is used within a single model the path weighting scheme was used, as this will take the nature of the model into account.

INITIAL WEIGHTS As outer weights are updated during each iteration they may be set to arbitrary initial weights. In this case +1 was used as initial weight.

NUMBER OF ITERATIONS Due to the lack of proofs of convergence of the algorithm, a maximum number of iterations must be given. The number of iterations was set to 300.

STOP CRITERION Due to its iterative nature, a stop criterion must be provided. The stop criterion was set to 10^{-5} . This means, convergence is achieved if the weights of the last run and the weights of the current run differ by less than 10^{-5} at maximum.

SIGNIFICANCE LEVEL For all hypothesis tests a significance level of $\alpha = 0.01$ was applied and the 99%-CI will be reported, consequently.

Based on this setup, convergence was achieved after 7 iterations. Given the estimation, each of the model parts and values must be assessed in terms of quality. As there is no global goodness of fit criterion in PLS-PM, this is done using several different quality indices.

5.4.2 *Assessing the Reflective Measurement Model*

In assessing the reflective measurement model parts, several different scores are usually used. Each of these is described theoretically before their values are given for the current model.

Internal Consistency

Due to the nature of reflective measurement where all items are supposed to reflect a single construct's value (cf. section 3.3.2), the degree as to which this assumption holds true must be assessed. Usually two different measures are employed. The classical Cronbach's α is given

in eq. (13). For each block q consisting of a LV and its associated MVs denoted as P_q its homogeneity is assessed using the correlation between the p -th measured variable in block q and the remaining p' MVs of this block.

$$\alpha = \frac{\sum_{p \neq p'} \text{cor}(x_{pq}, x_{p'q})}{P_q + \sum_{p \neq p'} \text{cor}(x_{pq}, x_{p'q})} \times \frac{P_q}{P_q - 1} \quad (13)$$

Despite its widespread use in particular in (covariance-based) SEM and psychological research in general (cf. Vinzi et al., 2010, p. 50) Cronbach's α is known to only provide a lower bound for reliability assessment. This is due to its main assumption: each measured variable is supposed to be equally important to measure the constructs value (Chin, 1998). Consequently, Cronbach's α will not be reported in this thesis. Instead, composite reliability (CRel) will be used, which overcomes the shortcoming of the traditional measure.

$$\rho = \frac{\left(\sum_{p=1}^{P_q} \lambda_{pq} \right)^2}{\left(\sum_{p=1}^{P_q} \lambda_{pq} \right)^2 + \sum_{p=1}^{P_q} (1 - \lambda_{pq}^2)} \quad (14)$$

Equation (14) shows the composite reliability as proposed by Werts, Linn, and Jöreskog (1974). It makes extensive use of the standardized outer loadings λ_{pq} of each MV P associated with a particular LV Q . The composite reliability is based on model results instead of correlations. It takes the variance induced by measurement error into account because the squared standardized outer loading equals the proportion of variance explained by the LV the current MV is linked to²⁴. Composite reliability takes values in $[0; 1]$. Table 16 provides an overview of commonly used interpretations of values. Why very high levels of composite reliability are undesirable needs further explanation. It stems from the domain sampling model assumption that each of the indica-

²⁴ cf. section 5.4.2 for more information about the importance of the squared standardized outer loadings.

range	applicability	source
< 0.60	insufficient composite reliability	Hair Jr. et al. (2014)
$0.60 < x < 0.70$	exploratory research	Nunnally and Bernstein (1994)
> 0.7	acceptable	Fornell and Larcker (1981)
$0.70 < x < 0.90$	advanced expl. or confirm.	Nunnally and Bernstein (1994)
> 0.95	undesirable	Hair Jr. et al. (2014)

Table 16.: Overview of interpretations used to judge composite reliability values.

tors measures its LV equally well. This does not imply that indicators are redundant. In contrast, a construct's MVs aim to measure different aspects of it. Hence, very high composite reliability values are not desirable (cf. e.g. Rossiter, 2002). Furthermore, a higher level of error term correlations may occur in these cases (cf. e.g. Drolet & Morrison, 2001; Hayduk & Littvay, 2012), which has a major impact on the estimations (cf. section 3.3.2 for the assumptions about error terms in reflective measurement models of PLS-PM).

Convergent Validity

Convergent validity is very closely related to the sampling model of items in reflective measurement. All indicators used to reflectively measure a LV are supposed to measure the same construct. Consequently, it is desirable that they share a large proportion of variance, a characteristic commonly referred to as *convergence*. Three measures are widespread to assess indicator reliability: statistical significance, outer loading threshold and Average Variance Extracted (AVE). The statistical significance and the outer loading threshold are commonly referred to as indicator reliability. The explanations of each of these measures given below are based on Hair Jr. et al. (2014, pp. 102–104).

STATISTICAL SIGNIFICANCE The outer loading of any indicator should be significant. As its value may be rather low, further measures must be applied.

OUTER LOADING THRESHOLD Commonly, a threshold of $\lambda \geq 0.708$ is applied for outer loadings. In order to justify this threshold value the concept of communality must be introduced. The commu-

nality of an indicator is given by squaring its standardized outer loading. This value represents the proportion of variance present in the indicator which is explained by the LV. A desirable lower threshold for this amount of variance is 50%. A value less than this threshold would mean that the error variance in the measurement is higher than the proportion of variance explained by the LV. As $\sqrt{0.5} \approx 0.707$ the threshold for outer loadings is set to 0.708, which is often relaxed to 0.70 (cf. Hair Jr. et al., 2014, p. 103). However, newly developed scales may lead to weaker outer loadings (cf. e.g. Hulland, 1999). As a consequence the influence a deletion of items with $0.4 \leq \lambda \leq 0.7$ has on composite reliability and AVE should be evaluated. If either of them meets their threshold values if the item is deleted, the indicator should be removed. Items with outer loadings $\lambda < 0.4$ should be eliminated regardless the impact their removal has (cf. Hair et al., 2011).

AVE Whereas communality is an indicator-level criterion, AVE is its construct level counterpart. It is defined as the average of all squared outer loadings of those items P_q associated with a single reflectively measured construct ξ_q . In case of standardized values of indicators, AVE is given by eq. (15)

$$AVE(\xi_q) = \frac{1}{P_q} \sum_{P_{q1}}^{P_{qp}} \lambda_{qp}^2 \quad (15)$$

Its threshold value is set to 0.5, indicating that on average the LV explains 50% of the variance present in each of its indicators, on average. This means, variance induced by measurement errors must be less than half of the variance present in the indicator.

Discriminant Validity

So far the criteria presented were all concerned with the degree to which an indicator is linked to its LV. Discriminant validity is a measure which deals with the degree as to which a LV is unique in a given model. This means, discriminant validity is achieved if each of the constructs has

as little in common with all remaining LVs as possible. Two different measures are commonly employed to judge discriminant validity.

LOW CROSS LOADINGS Cross loadings are those outer loadings an indicator has on LVs it is not connected to. As each MV is associated with one and only one LV in a given measurement model its outer loading ought to be higher than any of its cross loadings. In this case discriminant validity is achieved. Cross loadings are known to be a non-conservative criterion to establish discriminant validity, i.e. they tend to confirm the desirable characteristic easily (cf. Hair et al., 2011). As a consequence, a second score is usually assessed simultaneously.

FORNELL-LARCKER-CRITERION The Fornell-Larcker-Criterion (Fornell & Larcker, 1981) is based on the standardized outer loadings of each construct and the correlation of each LV with all other constructs. Suppose we want to examine the Fornell-Larcker criterion for a LV ξ_q . First of all, the AVE for ξ_q is calculated as the mean of squared standardized outer loadings of all indicators P_q of that construct. Now, the correlation between ξ_q and all the remaining LVs is calculated (denoted as r_{ij}). The Fornell-Larcker-Criterion is met if eq. (16) is fulfilled.

$$AVE(\xi_q) > \max((r_{ij})^2) \quad \forall i \neq j \quad (16)$$

A Well Fitting Reflective Measurement Model

Table 17 presents the results of the reflective measurement model assessment based on the scores described above. The figures indicate a very reasonable fit for the LVs *overall salience*, *structural salience*, and *visibility in advance*. The outer loadings of all the MVs for these constructs are well above 0.7 and highly significant. Furthermore, these indicators show excellent indicator reliability. Although almost undesirably high, I regard the very high composite reliability values to be a hint regarding excellent reliability of the reflective indicators proposed, particularly given the average variance extracted. Table 18 gives the square root of the AVE on the diagonal and uses the lower left triangle to display the correlations between a LV (column) and all other LV (rows). For all columns

the \sqrt{AVE} is greater than any of the other values in this column. Consequently, the Fornell-Larcker-Criterion was met. Similarly, the cross loadings criteria for the reflective indicators in the model were generally met. Table 19 shows the outer loadings achieved on each construct for each of the MVs. With exception of *c_cus*, *p_sim*, and *a_sui* each reflective indicator shows a reasonable result. As *c_cus* and *p_sim* show non-significant, weak outer loadings (i.e. $\lambda < 0.4$ according to Hair Jr. et al. (2014, p. 104)) and high cross loadings these will be removed from the model. The deletion of *c_cus* is in line with observation made during the trials. Many participants found it very hard to name the current usage (if any). The need to delete *p_sim* may stem from inconsistent

LV	Ind.	OL	IRel	CRel	AVE	DVal
Overall Salience (OVSAL)	ov_att	0.938***	0.880	0.951	0.866	Y Y
	ov_mem	0.906***	0.820			
	ov_por	0.947***	0.896			
Prototypicality (PRO)	p_exa	0.893***	0.797	0.621	0.492	Y N
	p_img	0.814***	0.662			
	<i>p_sim</i>	-0.128n.s.	0.016			
Cognitive Salience (COS)	<i>c_cus</i>	-0.119n.s.	0.014	0.698	0.351	Y N
	c_eas	0.473***	0.224			
	c_his	0.809***	0.655			
	c_per	0.463***	0.214			
	c_pus	0.513***	0.263			
	c_wor	0.858***	0.736			
Structural Salience (STS)	s_dir	0.870***	0.757	0.924	0.753	Y Y
	s_eas	0.888***	0.789			
	s_imp	0.845***	0.714			
	s_lor	0.868***	0.753			
Visibility in Advance (ADV)	a_dis	0.917***	0.841	0.928	0.764	Y Y
	a_per	0.913***	0.843			
	a_sui	0.840***	0.706			
	a_vis	0.823***	0.667			

Table 17.: Outer Loadings (OL), Indicator Reliability (IRel), Composite Reliability (CRel), Average Variance Extracted (AVE) and Discriminant Validity (DVal; Fornell-Larcker-Criterion | Cross Loadings) of all reflectively measured variables. Indicators with highly significant ($p < 0.001$) outer loadings are boldfaced. Please note: This table was first published in Kattenbeck (2015a), where significance levels were not indicated using stars.

interpretations across different participants (cf. section 6.3.1). *a_sui* will be retained, because of its significant and high outer loading on *visibility in advance*.

Given these results I would like to draw the attention to those reflective indicators having $0.4 < \lambda < 0.7$, i.e. *c_eas*, *c_per*, *c_pus* must be further assessed. According to Hair Jr. et al. (2014, p. 103) it is advisable to check whether the deletion of any such indicator with a low outer loading has an desirable effect on AVE. This means, if the deletion yields an effect on AVE such that the threshold is met, the indicator ought to be removed. Table 20 shows that no deletion yielded an improvement of AVE to meet the threshold. Accordingly, all indicators may be retained. However, this result will be double checked with respect to composite reliability. The figures reveal slight improvement over the threshold for composite reliability in all three cases. As the rise in AVE is reasonable for *c_per* and *c_pus* and the values are close to the threshold, taking content validity into account may provide a solution. *c_pus* will be deleted for a similar reason to *c_cus*. Its outer loading is significant but rather weak, composite reliability and AVE increase. More importantly, during experiments participants often claimed it to be particularly hard to determine a prior usage and some of them even refused to do so. *c_per* should be removed because of the increase in AVE and CRel. Beyond this reason, content validity is not likely to suffer from a deletion as this indicator refers to very personal experience. Although it may contribute strongly to *cognitive salience* if someone has personal mem-

	ADV	COS	PRO	OVSAL	STS
ADV	0.87				
COS	0.44	0.59			
PRO	0.40	0.29	0.70		
OVSAL	0.76	0.56	0.45	0.93	
STS	0.82	0.45	0.38	0.76	0.87
VIS	0.66	0.58	0.43	0.75	0.63

Table 18.: The diagonal shows \sqrt{AVE} . The lower left triangle gives the correlations between constructs. Clearly, the Fornell-Larcker-Criterion was met for all LVs as the value in the diagonal is highest for each column.

5 EMPIRICAL RESULTS

MV	ADV	COS	PRO	OVSAL	STS	VIS
c_cus	-0.01	-0.12	0.09	0.00	-0.04	-0.06
c_eas	0.33	0.47	0.20	0.35	0.37	0.25
c_his	0.28	0.81	0.19	0.38	0.27	0.48
c_per	0.22	0.46	0.07	0.22	0.16	0.22
c_pus	0.09	0.51	0.07	0.20	0.13	0.30
c_wor	0.41	0.86	0.29	0.52	0.41	0.54
ov_att	0.70	0.55	0.42	0.94	0.67	0.73
ov_mem	0.68	0.51	0.42	0.91	0.70	0.66
ov_por	0.73	0.49	0.41	0.95	0.75	0.70
p_exa	0.33	0.22	0.89	0.34	0.32	0.35
p_img	0.28	0.22	0.81	0.29	0.23	0.26
p_sim	-0.17	-0.12	-0.13	-0.23	-0.18	-0.22
s_dir	0.67	0.36	0.31	0.62	0.87	0.54
s_eas	0.80	0.45	0.39	0.77	0.89	0.62
s_imp	0.63	0.34	0.27	0.56	0.84	0.47
s_lor	0.70	0.41	0.35	0.65	0.87	0.52
a_dis	0.92	0.47	0.38	0.71	0.73	0.64
a_per	0.91	0.34	0.37	0.62	0.65	0.56
a_sui	0.84	0.41	0.38	0.76	0.87	0.62
a_vis	0.82	0.30	0.26	0.50	0.55	0.46

Table 19.: The outer loadings of all reflective indicators in the model for all constructs. For each row the value must be highest for the construct the indicator was associated with in order to achieve discriminant validity based on cross loadings. The MVs which do not meet cross loadings criteria are not bold-faced. The highest loading achieved per MV is bold-faced.

	c_eas	c_per	c_pus
AVE	0.351 (0.351)	0.488 (0.351)	0.467 (0.351)
CRel	0.767 (0.698)	0.781 (0.698)	0.767 (0.698)

Table 20.: The resulting AVE and CRel if the indicator is removed. Values estimated when the MV is included are given in brackets.

ories associated with a particular object, the generalizability of these memories to others may not be given. Finally, c_eas will be kept for two reasons. Firstly, AVE remained unchanged when deleting this MV. Secondly, the answers given for this indicator are valuable in terms of

route descriptions. If an object can be named easily its denomination may be useful in route instructions, as other participants may be likely to be able to make sense of such references. This is, e.g., reflected by the results of Schroder et al. (2011), who report “name” to be referred to most frequently.

Overall, the reflective measurement model suggested shows a very reasonable fit. Hence, the reflectively measured LVs provide a sound basis for assessing the structural model. Before the fit of the structural relationships proposed between constructs can be examined, the formative measurement model of *visual salience* must be analyzed.

5.4.3 *Assessing the Formative Measurement Model*

²⁵Causal-formative indicators are not assessed based on correlations as reflective indicators are. This is a result of formative measurement theory, where content validity is essential. This means that causes must be exhaustive to cover any facet of a given LV. Exhaustiveness can be ensured by a thorough literature review, which helps to include all important parts of a construct to be measured formatively (cf. e.g. Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003). Furthermore, a second assumption rendering internal consistency assessments useless in formative measurement is that causes are error free (cf. e.g. Diamantopoulos & Sigauw, 2006; Edwards & Bagozzi, 2000). Consequently, quality factors different to those of reflective measurement must be employed. These are:

CONVERGENT VALIDITY is established by analyzing the degree of correlation between a formatively measured construct and its reflectively measured counterpart.

ABSENCE OF COLLINEARITY is fundamental for the suitability to include a set of causes. That is, redundant causal-formative indicators are not desirable in formative measurement models.

SIGNIFICANCE AND RELEVANCE of causal-formative indicators must be ensured regarding the contribution of a cause in a relative

²⁵ This section is based on the advice given in Hair Jr. et al. (2014, pp. 118–166).

(i.e. given all other causes) as well as an absolute manner (i.e. regardless the other causes).

Convergent Validity

In PLS-PM a convergent analysis is a widespread method to assess the appropriateness of formative measurement. It is done based on a submodel. Figure 27 shows a sample submodel to examine convergent validity. Construct X is formatively measured on the left hand side whereas the same Construct X is measured reflectively on the right hand side. Both LVs are connected using a single path (cf. e.g. Diamantopoulos & Winklhofer, 2001). If formative measurement is appropriate, a high path coefficient (i.e. > 0.80) is required. This means, the formative measurement model must be able to explain at least 64% of the variance present in the reflective measurement model part (cf. Chin, 1998).

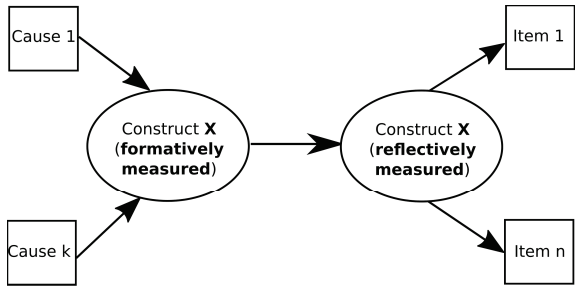


Figure 27.: A sample submodel used to assess convergent validity. A similar figure is given in Hair Jr. et al. (2014, p. 122). The figure was created using the Inkscape (cf. The Inkscape Team, 2016).

The specific model applied in order to examine the appropriateness of formatively measuring *visual salience* is given in fig. 28. The model incorporates the reflective MVs to *visual salience* by associating them with a new construct.

Hence, this submodel consists of two constructs. The formatively measured LV VIS_F and its reflectively measured counterpart VIS_R. If measuring *visual salience* formatively is appropriate, the path coefficient between VIS_F and VIS_R will be high. However, this analysis must

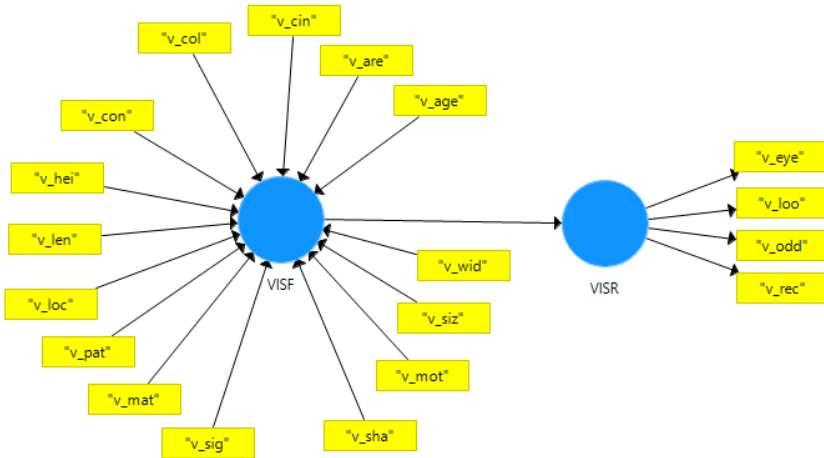


Figure 28.: The convergent validity model used to assess whether formatively measuring *visual salience* is appropriate. This figure was first published in Kattenbeck (2015a, p. 6). The figure was created using the software Smart-PLS (Ringle et al., 2015).

be based on a sound reflective measurement model for VIS_R. It is important to note that the indicators used in the reflective measurement model of the LV are usually not part of the SEM. Hence, either existing scales are used or single item measurement is utilized in order to ensure shorter surveys (cf. Hair Jr. et al., 2014, p. 122). As no prior developed scales exist, four items were included in the survey to enable the convergent analysis. Of course, this was done at the expense of asking four more questions. According to the criteria presented by Diamantopoulos et al. (2012) single item measurement is not appropriate in this case, primarily because sample size is large and a high path coefficient is to be expected between the formative construct and its reflective counterpart. Moreover, a valid and reliable way of reflectively measuring *visual salience* is particularly important because examining the appropriateness of formative measurement to *visual salience* is one of the research goals of this thesis. Consequently, as a first step the MVs associated with VIS_R must be assessed in order to ensure a sound reflective measure-

LV	Ind.	OL	IRel	CRel	AVE	DVal
VISAL_R	v_eye	0.939***	0.882			
	v_loo	0.941***	0.885	0.950	0.827	Y Y
	v_odd	0.867***	0.752			
	v_rec	0.888***	0.789			

Table 21.: Assessment of the four reflective indicators of *visual salience*. Indicators with highly significant ($p < 0.001$) outer loadings are bold-faced. Please note: This table is taken from Kattenbeck (2015a, p. 5).

ment model. Applying the same indices as above, table 21 shows the results achieved. All criteria are met by the reflective MVs proposed for *visual salience*. The reflective measurement model part of the submodel provides a sound basis in terms of valid and reliable measurement. Using the same algorithmic setup as employed for the overall model, the algorithm revealed a path coefficient of 0.81 for the structural relationship between VIS_R and VIS_F. Applying a bootstrapping analysis with $K = 5000$ resamples the coefficient is highly significant ($t = 47.113$, $p < 0.001$, *Std. Err.* = 0.017). This suggests that strong empirical support exists for the formative specification of measuring *visual salience*. Based on this result, the causal-formative indicators can be further evaluated based on *model I*.

Examining Multicollinearity

The need to avoid multicollinearity in formative measurement models stems from the regression based approach employed, i.e. high levels of correlation between LVs will increase the standard error (SE) of the regression coefficient estimated (cf. Weiber & Mühlhaus, 2010, p. 208). If the causal-formative indicators of a construct are (highly) correlated the significance tests used to show their deviance from zero are more likely to become insignificant. Moreover, severe incorrectness may occur due to sign changes resulting from correlations. Hair Jr. et al. (cf. 2014, pp. 123–124) provide an example: Suppose a formatively measured construct has two causes and both are slightly correlated with it. Additionally assume their multicollinearity is higher than any of the correlations with the LV itself. In this case, a negative sign will result in the final estimation

of weights for the cause which shows weaker correlation with the LV. This situation is undesirable as the causes are supposed to form not to weaken the construct. Two measures are commonly used to assess issues of multicollinearity:

$$TOL_{cf} = 1 - R_{cf}^2 \quad (17)$$

$$VIF_{cf} = \frac{1}{TOL_{cf}} \quad (18)$$

Let x_{11} denote the first causal-formative indicator of construct one and therefore x_{cf} represent the c^{th} cause of the formatively measured construct f . Equation (17) gives the coefficient of determination resulting from a multiple regression where x_{cf} is used as dependent variable and all the other causes of construct f are independent variables. R_{cf}^2 represents the coefficient of determination resulting from such a regression for x_{cf} . Hence, this cause's tolerance represents the proportion of variance present in x_{cf} which cannot be explained by the other causes of construct f . The Variance Inflation Factor (VIF) is based on the tolerance (cf. eq. (18)). It is particularly important because of the meaning of \sqrt{VIF} , i.e. it indicates the severity of the multicollinearity between the indicators. This means, \sqrt{VIF} represents the enlargement of the standard error in estimating regression coefficients. As both measures are closely related the critical values used are corresponding as well. The thresholds commonly referred to $TOL \geq 0.2$ and $VIF \leq 5$ for non-severe levels of multicollinearity (cf. Hair et al., 2011). If these thresholds are not met, three remedies may be employed. One can either remove one or more of the affected causes from the model or these causes are used to form a so-called higher-order construct or the collinear MVs are used as a composite *indicator* (e.g. by using their average). However, the first and the third way of counterbalancing multicollinearity may have a large impact on content validity because they delete causes from models which are based on the assumption of exhaustiveness of MVs. Generally speaking, there may be occasions where none of the remedies suggested can be justified. In any such cases the formative measurement model must be dismissed, leading to a respecification of the LV.

cause	VIF
v_age	2.02
v_are	2.71
v_cin	3.60
v_col	3.46
v_con	1.93
v_hei	2.19
v_len	2.09
v_loc	1.51
v_mat	2.38
v_mot	1.14
v_pat	2.11
v_sha	1.91
v_sig	1.16
v_siz	2.91
v_wid	2.27

Table 22.: VIFs for all causal-formative indicators. As all values are $VIF \leq 5$ multicollinearity is not a threat to the results reported.

Table 22 shows that all values are below the common threshold $VIF < 5$, i.e. multicollinearity is not a threat according to Hair Jr. et al. (2014). Indeed, the more conservative threshold $VIF < 3.33$ (Diamantopoulos & Siguaw, 2006) for all of the causes—except for tone (v_col) and intensity (v_cin)—was met. However, as both concepts are closely related, the slightly higher VIF is negligible. In contrast, these values are regarded to be low given the relatedness and do stress the importance of distinguishing tone and intensity. Hence, reliable estimations of the correlation coefficients for all causes can be calculated.

Assessing the Relevance of Causes

The relative importance (i.e. the outer weight) and the absolute importance (i.e. the outer loading) for each cause is found based on bootstrapping (Efron, 1979). The procedure is described based on pseudo-code.

Require: $K ::=$ number of resamples

Require: $S ::=$ original sample

Require: $n ::=$ size of S

Require: ow_{orig} ::= vector of outer weights based on full sample
for all $i < K; i++$ **do**
 $S' \leftarrow \text{drawsample}(\text{from} = S, \text{size} = n, \text{replace} = \text{TRUE})$
 $ow_i \leftarrow \text{estimateOuterWeights}(S')$
end for
for all $j < \text{length}(ow_{orig})$ **do**

$$stderr_j = \sqrt{\frac{\sum_i^K (ow_{orig}[j] - ow_{res}[i,j])^2}{K}}$$

$$t_j = \frac{ow_{orig}[j]}{stderr_j}$$

end for
return t_j

As PLS-PM is a non-parametric estimation method, the significance of the regression coefficients is estimated using K resamples drawn with replacement from the original sample S . Each of the resamples S' is of the same size n as the original sample is. For each of the resamples the outer weights ow_i are determined. These provide an approximation of the true distribution of outer weights in the population. Based on these resamples, the standard error of the estimation can be determined based on the squared difference between original outer weights and their resampled counterparts. The t-value of an outer weight is obtained by dividing the original weight by its standard error. The t-statistic is known to follow a t-distribution with $df = n - 1$. A two-sided test is applied to examine whether the outer weights differ significantly from zero ($ow_{orig}[j] \neq 0$). As the t-distribution approximates $N(0;1)$ very well for large df values, critical t-values are based on $N(0;1)$, i.e. $t_{crit} > 1.96$ in case of $\alpha = 0.05$ and $t_{crit} > 2.57$ if $\alpha = 0.01$ was chosen.

Beyond these significance values it is common to report the CI for outer weights. It is important to remember that reporting a CI does not mean that a given outer weight is with $100 - \alpha\%$ chance between its lower and its upper bound. Instead, it means that one will be able to construct a CI from this particular bootstrap which contains the unknown true value of the outer weight in $100 - \alpha$ of the cases the experiment would be repeated (cf. Steiger & Fouladi, 1997). Three different options are widely used to construct CIs in bootstrapping analyses.

PERCENTILE BOOTSTRAP

For each variable the K estimations are ordered ascendingly. The

interval to conduct a two-tailed test is constructed by the $K * \frac{\alpha}{2}$ -th value and the $(100 - \frac{\alpha}{2}) * K$ -th value of the ordered bootstrapping values for a particular variable. If the interval $[K * \frac{\alpha}{2}; (100 - \frac{\alpha}{2}) * K]$ does not contain zero, a significant result is achieved (cf. Taylor, MacKinnon, & Tein, 2008, p. 245). This kind of bootstrapping method is known to show a large coverage error²⁶ for non-symmetric distributions.

BIAS-CORRECTED BOOTSTRAP

CI's try to overcome the problems of percentile bootstrap methods for non-symmetric distributions. They are based on the number p of bootstrapped outer weights, which are less than the originally estimated value. Let $b = p/K$ be the relative frequency of these instances. The CI's lower bound is given by the s -th element of the bootstrapped outer weights in ascending order. s is found by using the integer part of $K * (2b - z_{\alpha/2})$. Similarly, the upper bound is found based on $K * (2b + z_{\alpha/2})$ (cf. Carpenter & Bithell, 2000; Taylor et al., 2008). Again, if this interval does not contain zero a significant result is achieved. This method is known to yield significant coverage error in rather many cases and is not widely used, consequently (cf. Carpenter & Bithell, 2000, pp. 1153–1154).

BIAS-CORRECTED AND ACCELERATED BOOTSTRAP

Introduced by Efron (1987), this method takes non-symmetry and skewness issues into account. The bounds of the CI are found by using the integer part of $K * \left(b \pm \frac{z_{\alpha/2} \pm b}{1 + a(z_{\alpha/2} \pm b)} \right)$. In this case b is again the relative frequency of instances of bootstrapped outer weights which are lower than the original estimate. According to Carpenter and Bithell (2000, pp. 1154–1155) several different methods for calculating a exist, depending on the distributional assumptions of the bootstrap (if any). In its most fundamental kind, a simple jackknife (cf. Efron, 1979) is used. The advantages of percentile bootstraps apply to this method and the coverage error is significantly decreased. This comes with a disadvantage

²⁶ Coverage error refers to the degree as to which the probability not to meet the bound of a CI is larger than the significance level chosen (cf. Carpenter & Bithell, 2000, p. 1143).

regarding α -levels converging to zero in which cases coverage error may be increased.

Overall, the BCa-method shows an important advantage and it is recommended by Ringle et al. (2015). Consequently, it will be used as a default during the analysis, i.e. unless otherwise stated CIs reported are based on this method. Having said this, the so-called sign indeterminacy of LVs scores (H. O. A. Wold, 1985) is another important property of the PLS algorithm, which plays another crucial role in bootstrapping methodology. That is, the outer weights calculated based on any of the resamples S' may have a different sign than the estimation based on the original sample S yielded. An increased bootstrap SE combined with a mean of a specific outer weight closer to zero will be an immediate consequence. As sign indeterminacy cannot be avoided, three different alternatives to handle it were proposed by Hair Jr. et al. (2014, p. 135):

NO SIGN-CHANGE The signs of estimations based on bootstrap sample S' are taken *as-is*.

CONSTRUCT-LEVEL SIGN CHANGE The signs of all coefficients of a LV which were determined for the original sample are compared with those of the current bootstrap sample. If and only if the majority of signs of S' need to be reversed in order to match the signs in the original sample, the sign change in S' is applied.

INDIVIDUAL-LEVEL SIGN CHANGE Any differences in signs regarding coefficients between estimations based on S vs. S' are harmonized by changing the bootstrap signs accordingly.

Following the arguments provided by Rönkkö (2014), I will not apply any sign changes at all. This means, I will use option *no sign-change*, which will yield conservative significance results (Hair Jr. et al., 2014), but does not contradict the general bootstrapping methodology (cf. section 6.2.4).

Regarding the significance and signs of outer weights, four groups of variables may be distinguished in these results:

GROUP 0 **age** | **color intensity** | **location** | **shape** | **signage** | **size** show significant and positive outer weights.

5 EMPIRICAL RESULTS

Cause	Outer Weights	Std. Err. OW	99%-CI OW (Bootstrap)	Outer Loading	Std. Err. OL	99%-CI OL (Bootstrap)
age (v_age)	0.245***	0.066	[0.063; 0.402]	0.549***	0.053	[0.384; 0.659]
area (v_are)	0.145n.s.	0.079	[-0.058; 0.341]	0.631***	0.046	[0.484; 0.718]
intensity (v_cin)	0.284***	0.083	[0.039; 0.474]	0.601***	0.051	[0.448; 0.699]
tone (v_col)	0.126n.s.	0.083	[-0.079; 0.347]	0.595***	0.052	[0.435; 0.699]
condition (v_con)	-0.185***	0.068	[-0.363; -0.018]	0.400***	0.057	[0.226; 0.529]
height (v_hei)	0.012n.s.	0.072	[-0.189; 0.189]	0.613***	0.051	[0.442; 0.712]
length (v_len)	-0.122n.s.	0.072	[-0.293; 0.069]	0.432***	0.060	[0.259; 0.560]
location (v_loc)	0.285***	0.057	[0.128; 0.418]	0.664***	0.047	[0.510; 0.744]
material (v_mat)	-0.009n.s.	0.068	[-0.179; 0.177]	0.516***	0.052	[0.349; 0.622]
motion (v_mot)	0.020n.s.	0.049	[-0.110; 0.148]	0.209***	0.060	[0.038; 0.348]
pattern (v_pat)	0.128n.s.	0.069	[-0.049; 0.294]	0.573***	0.050	[0.414; 0.675]
shape (v_sha)	0.204***	0.061	[0.034; 0.357]	0.655***	0.046	[0.506; 0.736]
signage (v_sig)	0.206***	0.047	[0.069; 0.319]	0.264***	0.062	[0.098; 0.402]
size (v_siz)	0.342***	0.078	[0.133; 0.527]	0.780***	0.036	[0.649; 0.831]
width (v_wid)	-0.123n.s.	0.069	[-0.298; 0.061]	0.554***	0.054	[0.379; 0.656]

Table 23.: Outer weights (OW), outer loadings (OL) and standard errors of all causes of *visual salience*. *** indicates $p < 0.001$ ($K = 5000$ resamples, option *no sign-change*). Please note: Parts of the table were first published in Kattenbeck (2015a, p. 6).

GROUP 1 **condition** must be discussed because of its significant, but negative outer weight.

GROUP 2 **area | height | tone | pattern | motion** have highly significant outer loadings, but their positive outer weights are insignificant.

GROUP 3 **width | material | length** build the fourth group. All of these show non-significant and negative outer weights combined with positive and highly significant outer loadings.

Whereas group 0 members show results to be expected, group 1 shows an interesting anomaly. **Condition** has a significant but negative outer weight. This means, it has a suppressive effect (cf. Cenfetelli & Basselier, 2009, pp. 696–697) on *visual salience*. Given that all other causes showing significant (and positive) outer weights remain constant, *visual salience* will be decreased by the ratings regarding **condition**. Two causes of group 2 and two causes of group 3 have to be examined in combination. The causal-formative indicators **width**, **length**, **height** and **area** will be deleted from the model. All of these show non-significant, and in the case of **width** and **length**, even negative outer weights. Their

deletion will not affect content validity because all are well-correlated with *size* which shows the highest outer weight of all causes and it is highly significant. This is a strong hint that participants subsumed the non-significant causal-formative indicators using *size*. This result reinforces the importance of testing all causes simultaneously as all these dimensions were revealed in other studies (cf. section 3.3.3). Those studies, however, did not test overall models. Of group 3, *material* as well as *pattern* will be retained despite their non-significant outer weights (cf. Hair Jr. et al., 2014, p. 129). Both show reasonable and significant outer loadings, i.e. taken individually their contribution to *visual salience* is remarkable (i.e. > 0.5). Finally, the insignificant outer weight for *tone* is rather surprising, particularly because it is largely used by participants. In addition to the questions used to estimate the SEM, participants were asked to find a name for the object they were currently rating. The name was required to be usable in route descriptions given to others. 25% of the answers explicitly refer to the *tone* of the object. This is reflected by the high and significant outer loadings of this cause. Hence, *tone* will be retained in the model. Overall, this result points to a major issue. Interpreting the relative importance of causes based on their outer weights is yet to be done with caution because of the large number of causes used to model *visual salience*. According to Cenfetelli and Bassellier (2009) the more causes used, the higher the chance that one or more of these will be rendered insignificant. This is particularly true for uncorrelated causes (cf. Hair Jr. et al., 2014, p. 128). Suppose c causes are given in a formative measurement model and no correlations exist between these. In this case any outer weight has an upper bound of $1/\sqrt{c}$. Increasing c yields both, a decreased maximum possible value and smaller average of outer weights. The chance of insignificant causes increases in such circumstances. This disadvantage can be overcome by means of a second order analysis, i.e. causes are grouped into several so-called lower order components (LOCs), which, in turn, form a so-called higher order component (HOC). Drawing on the formative measurement model of *visual salience* the number of causes revealed in earlier studies equals $c = 15$, i.e. it is rather large. Although multicollinearity is not a threat in case of the current data set (cf. section 5.4.3), the outer weights suggest (cf. table 23) that some correlation must be present as

some of the bootstrapped values exceed $\frac{1}{\sqrt{(15)}} = 0.258$. As a consequence, an adjustment to the formative measurement model by means of a second order analysis will be proposed in section 5.5.1.

5.4.4 *Assessing the Structural Model*

Drawing on these very good results regarding formative and reflective measurement models, the relationships between the LVs can be further assessed. In contrast to CB-SEM where the difference between the model-based covariance matrix and its empirical counterpart may be assessed using various χ^2 -based indices, non-consistent PLS-PM lacks a global goodness-of-fit criterion²⁷. Instead, five aspects are commonly taken into account. In analogy to the formative measurement model, multicollinearity between the LVs must be assessed and the structural relationships will be tested by means of bootstrapping. Furthermore, the coefficient of determination, predictive accuracy and relevance of constructs which are exogenous to others must be examined. These criteria are suited to gain an insight into the predictive capabilities the model has.

Multicollinearity

The need to examine multicollinearity between constructs stems from the OLS regression based approach. Basically, each endogenous LV of a structural model is predicted by all its antecedents. Hence, multicollinearity between predecessor constructs would yield biased regression (path) coefficients. Examining the degree of multicollinearity is again based on the VIF.

Relevance and Significance

Significance of path coefficients is determined using bootstrapping. The calculation and interpretation of t-values including the CIs for path coefficients based on the critical t-values is done in analogy to formative

²⁷ This is a major difference to the latest methodological advancements achieved (cf. section 6.2.3).

model assessment. Path coefficients are standardized, i.e. their values range between ± 1 . The closer a path coefficient is to zero, the weaker the relationship is, i.e. the weaker the effect an exogenous LV has on the endogenous LV. This means, the relevance of a significant relationship can be examined, directly.

Coefficient of Determination

The coefficient of determination R^2 takes values in $[0; 1]$. It is based on the correlations between actual values of the endogenous LV and its predictions. R^2 represents the proportion of variance in the endogenous LV which is explained by its associated predictor LVs. Hence, the coefficient of determination depends on the number of exogenous LVs which are present in a given model. Increasing the number of exogenous LVs will inevitably increase R^2 and vice versa, i.e. adjustments to R^2 are needed to compare models with a different number of constructs. Equation (19) provides a way to calculate the coefficient of determination which takes the sample size n and the number of exogenous LV, q' , used to predict the endogenous LV in the model into account.

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n - 1}{n - q' - 1} \quad (19)$$

It may be used to compare models comprising different numbers of exogenous LVs or having unequal sample sizes or both using bootstrapping. However, the values calculated based on eq. (19) cannot be interpreted in the same way as R^2 is, i.e. R_{adj}^2 can only be used for comparisons.

Predictive Accuracy

The coefficient of determination may be used to assess the degree of influence exogenous LVs have on an endogenous construct one at a time. The effect size f^2 (cf. eq. (20)) compares the difference between the R^2 when a particular exogenous LV is included and when this very construct is excluded to the amount of unexplained variance present in the endogenous LV if all exogenous LVs are included. According to

Cohen (1988) commonly used interpretations of f^2 effect sizes are: *small* ($f^2 < 0.02$), *medium* ($0.02 \leq f^2 < 0.35$) and *high* ($f^2 \geq 0.35$).

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \quad (20)$$

Predictive Relevance

The Q^2 value by Stone (1974) and Geisser (1974) is the last measure to be examined in structural model assessment. Q^2 indicates predictive relevance, i.e. whether the inner model “accurately predicts the data points of indicators in reflective measurement models of endogenous constructs” (Hair Jr. et al., 2014, p. 178). Basically, two different approaches exist to calculate Q^2 (cf. Hair Jr. et al., 2014, p. 183). The so-called cross-validated redundancy approach takes the LV scores of exogenous and endogenous constructs as well as the path coefficients of the inner model into account. Contrastingly, in a cross-validated communality estimation scenario the inner model estimations are neglected with the exception of the LV-score of the endogenous construct currently examined. Based on this difference, the cross-validated redundancy approach is to be preferred in PLS-PM because it uses the path coefficients of the structural model for predictive purposes. This approach to calculate Q^2 is based on blindfolding, comprising the following steps (cf. Hair Jr. et al., 2014, pp. 178–183).

1. Estimate the whole model.
2. Calculate estimations \hat{Y} of the endogenous variable Y examined based on the LV scores of its antecedents and the path coefficients estimated (i.e. do not use the measurement model results).
3. Choose an omission distance D such that $\frac{n}{D} \notin \mathbb{N}$, where n is the sample size and $5 \leq D \leq 10$.
4. Use the estimations \hat{Y} to predict MV-values systematically omitted based on D . For example, if $D = 7$ every 7th value of each MV is neglected in a single blindfolding round, starting from case 1 for MV_1 , and case 2 for MV_2 in the first round and so on.

5. Repeat the blindfolding, i.e. the prediction of neglected MV-values from \hat{Y} values, until each value of all MVs associated with Y was omitted once.

Q^2 is then given by eq. (21), where E_D denotes the sum squared error of predictions of estimated omitted data points. O_D is the sum squared error obtained by predictions simply based on the mean (cf. Chin, 1998, p. 317).

$$Q^2 = 1 - \frac{\sum_D E_D}{\sum_D O_D} \quad (21)$$

“Thus, [...] Q^2 represents a measure of how well-observed [sic] values are reconstructed by the model and its parameter estimates. $Q^2 > 0$ implies the model has predictive relevance, whereas $Q^2 < 0$ represents a lack of predictive relevance” (Chin, 1998, p. 318).

The effect size q^2 (cf. eq. (22)) can be used to compare different impacts exogenous variables have in terms of predictive relevance for a particular endogenous LV (cf. Hair Jr. et al., 2014, p. 183).

$$q^2 = \frac{Q_{included}^2 - Q_{excluded}^2}{1 - Q_{included}^2} \quad (22)$$

The results of q^2 are interpreted as small ($q^2 < 0.02$), medium ($0.02 \leq q^2 \leq 0.35$) and large ($q^2 > 0.35$) (cf. Hair Jr. et al., 2014, p. 184).

The Results

It is important to note that the calculations for the assessment of the structural model are based on the measurement model which is adapted according to the findings above. First of all, the degree of multicollinearity between LVs must be assessed by means of the VIF. Clearly, the figures are low across LVs ($1.51 \leq VIF \leq 3.46$), i.e. multicollinearity is not a threat to the results achieved.

All path coefficients were estimated using the PLS-PM algorithm and tested for significance using bootstrapping (cf. table 24). According to these figures, visual and structural aspects are equally important to *over-*

LV	path coeff.	99%-CI	$R^2_{excl.}$	f^2	$Q^2_{excl.}$	q^2
VIS	0.312***	[0.255; 0.475]	0.679	0.151	0.585	0.092
COS	0.119***	[0.034; 0.191]	0.712	0.032	0.612	0.021
PRO	0.049	[-0.011; 0.160]	0.719	0.007	0.619	0.003
STS	0.312***	[0.157; 0.407]	0.690	0.111	0.593	0.071
ADV	0.224***	[0.064; 0.330]	0.706	0.054	0.607	0.034

Table 24.: Each row shows the path coefficient between the current LV and *overall salience*. *** indicates $p < 0.001$ ($K = 5000$ resamples). Columns f^2 and q^2 represent the effect sizes resulting from a deletion of the current row's exogenous construct. Please note: Parts of this table were first published in Kattenbeck (2015a, p. 7).

all salience. Furthermore, *visibility in advance* and cognitive aspects contribute significantly to *overall salience*. However, *visibility in advance* is more important than *cognitive salience*. Finally, *prototypicality* was rendered insignificant, i.e. it does not help to explain the variance given in *overall salience*. Overall, the model is able to explain approximately 72% of the variance present in *overall salience*. This is a reasonable proportion, particularly given an in-situ experiment. The predictive accuracy achieved is $Q^2 = 0.620$, which is also highly satisfactory. The effect sizes in terms of predictive accuracy (i.e. f^2) and relevance (i.e. q^2) are small (cf. table 24). The largest path coefficients yield the highest predictive accuracy and relevance. *Visual salience* and *structural salience* are important to explain the variance present in *overall salience* and to predict its value.

5.5 ADAPTING THE MODEL

Taken as a whole the initial model shows very reasonable overall fit regarding all of its elements. The structural model is able to explain a large proportion of variance present in *overall salience* with a reasonable predictive accuracy. The reflective measurement models are generally well suited. However, only two indicators remain in case of *prototypicality* and half of the indicators measuring *cognitive salience* have to be dismissed. Estimating *visual salience* formatively is revealed to be an

adequate mode of measurement. Analyzing the outer weights, four causes were removed and based on these results adaptations to the initial model are proposed. First of all, the formative measurement model of *visual salience* will be adapted in order to counterbalance potential bias induced by the number of causes. A reassessment of the model will allow for a better understanding of the relative importance the causes have. In a second step, I will estimate *model II* which involves structural relationships based on mediation and is based on both, prior empirical evidence and theoretical rationale.

5.5.1 *Adapted VISF*

Although four causes are removed from the originally proposed formative measurement model due to high correlations with *size*, the remaining eleven causes are a rather large amount of causes for a single construct. As detailed in section 5.4.3 this increases the chance of getting insignificant results due to the theoretical upper bound of outer weights if no multicollinearity is present. In order to counterbalance this problem, I follow the advice given in Cenfetelli and Bassellier (2009) and develop a formative-formative hierarchical component model for *visual salience*. In doing so, I adhere to the steps provided in Hair Jr. et al. (2014, pp. 229–237). This model does not include the four causes area, length, width, height due to the high correlations with *size*. Figure 29 provides an overview of the LOCs introduced to formatively measure *visual salience* as a HOC.

The corresponding structural model showing the LOCs and their HOC is given in fig. 29. The four LOCs are: *Color*, *Geometry*, *Surface* and *Secondary*. The latter of these constructs subsumes aspects, which are influenced by visual aspects, but which may be misleading. For example, a house may have been built in the 20th century, yet appear much older due to incorporated older building parts. In this example the *visual salience* relating to age may be important if the houses in the local environment are built in the 20th, as well, but do not appear to be older. All LOCs are formative constructs and in turn formatively measure *visual salience*. In order to assess the model, all causes of all LOCs will be reused at the HOC-level in order to estimate the outer weights of causes

at the LOC-level. The reflective measurement model part needs not to be reassessed for this kind of analysis, as the assessment does not take structural relationships between the HOC and other non-LOC constructs into account.

Interpretation of the Results of Outer Weights for LOC-Causes

Table 25 details outer weights and loadings of the MVs at the LOC level. The results clearly stress the importance of the Second-Order-Analysis. A one-sided test of significance was applied, reflecting the fact that positive contributions of each causal-formative indicator are assumed.

Table 25 shows that all outer loadings are significant at the 1% significance level. Furthermore, the adapted model reveals all causal-formative indicators but *condition*, *tone* and *intensity* are significant at the 1% level. Having said this, the interpretation of the results is based on the LOC-level.

COLOR Both, *tone* and *intensity*, turn out to be equally important in an absolute as well as a relative manner. However, both outer weights are only rendered significant at $\alpha = 0.05$. Together with the fact that the equally sized outer loadings are close to 1.0 and

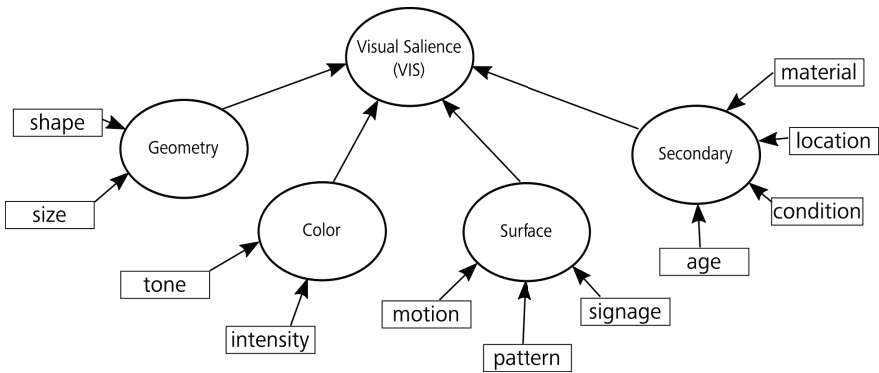


Figure 29.: The second order model grouping causes into four LOCs in order to measure *visual saliency* as a higher order construct. Please note: The remaining constructs were neglected for the sake of clarity. This figure was drawn using Inkscape (cf. The Inkscape Team, 2016).

LOC	Cause	Outer Weight	Std. Err. OW	99%-CI OW (Bootstrap)	Outer Loading	Std. Err. OL	99%-CI OL (Bootstrap)
Color	intensity (v_cin)	0.513*	0.227	[-0.371; 0.884]	0.954**	0.374	[-0.923; 0.998]
	tone (v_col)	0.533*	0.302	[-0.898; 0.843]	0.957**	0.390	[-0.997; 0.995]
Secondary	age (v_age)	0.364***	0.073	[0.197; 0.534]	0.732***	0.045	[0.610; 0.820]
	condition (v_con)	0.141*	0.077	[-0.050; 0.312]	0.642***	0.056	[0.491; 0.751]
	location (v_loc)	0.565***	0.061	[0.415; 0.701]	0.785***	0.045	[0.663; 0.870]
	material (v_mat)	0.271***	0.074	[0.090; 0.428]	0.737***	0.050	[0.594; 0.828]
Surface	motion (v_mot)	0.158**	0.063	[0.012; 0.307]	0.298***	0.073	[0.115; 0.460]
	pattern (v_pat)	0.936***	0.030	[0.844; 0.987]	0.951***	0.024	[0.868; 0.987]
	signage (v_sig)	0.238**	0.078	[0.057; 0.421]	0.261**	0.087	[0.056; 0.459]
Geometry	shape (v_sha)	0.423***	0.066	[0.268; 0.561]	0.804***	0.058	[0.716; 0.891]
	size (v_siz)	0.703***	0.070	[0.579; 0.838]	0.936***	0.058	[0.890; 0.980]

Table 25.: Outer weights (OW)/ outer loadings (OL) and standard errors of all causes of all LOCs modeling *visual salience* as a HOC. *** indicates $p < 0.001$, ** indicates $p < 0.01$, and * means $p < 0.05$, respectively ($K = 5000$ resamples, option *no sign change*).

highly significant a possible explanation may be the interpretation by participants. In contrast to the pre-study the participants of the main study tend not to distinguish tone from intensity. In fact, the distinction may have been an artefact of the way objects were chosen in the pre-study (cf. section 3.3.4). Both dimensions are retained in the model, consequently.

GEOMETRY The outer weights of both dimensions clearly indicate that size is far more important than shape. This does not hold for absolute measures as the outer loadings of both are rather similar.

SECONDARY From a relative point of view, an object's condition no longer contributes significantly to its LOC, as it does not show a significant outer weight at $\alpha = 0.01$. However, its absolute contribution is strong and highly significant. It is consequently retained in the model. Relative to the other significant causal-formative indicators location is most important followed by (visual) age and (visual) material.

SURFACE Pattern is by far the most important aspect in a relative and absolute manner. Of the two remaining causal-formative indicators, the outer weight and loading of cause signage are equally sized. Regarding motion its absolute importance is almost twice

as large as its outer weight is. This means, relative to the other causes, it is least important, but its absolute importance is still larger than the one signage shows.

Overall, the measurement model for all LOCs show an adequate fit and no further causes need to be removed.

Interpreting the Structural Relationships Between LOCs and HOC

Based on this detailed analysis of the importance of MVs to their corresponding LOCs, the structural relationships between LOCs and HOC must be examined. Multicollinearity is not a threat to the results as VIF values of the LOCs are low ($1.156 \leq VIF \leq 2.231$). With exception of Surface all structural relationships are highly significant (cf. table 26).

LOCs	path coeff.	99%-CI
Geometry	0.497***	[0.308; 0.637]
Color	0.382***	[0.224; 0.516]
Secondary	0.282***	[0.087; 0.479]
Surface	0.131*	[0.004; 0.323]

Table 26.: Each row shows the path coefficient between the current LOC and the HOC. *** indicates $p < 0.001$ and * means $p < 0.05$ ($K = 5000$ resamples). LOCs showing highly significant results are bold-faced.

Table 26 shows that dimensions regarding the geometry of an object are most important to *visual salience*, followed by color. This result is particularly notable as one would have expected color to be the most important contributor to *visual salience*. The path coefficient of the LOC Secondary, i.e. aspects which are visually prevalent is third-largest and highly significant, too. It is important to note that it is not useful to analyze the R^2 value achieved for *visual salience* in this case because the causes used to measure the LOCs are used at the HOC level simultaneously. R^2 achieved for the HOC is close to 1.0 due to this design. Consequently, neither f^2 nor q^2 are not reported here. However, this R^2 value is undesirable for any further analysis of the structural model based on the second-order analysis. Following the advice given in

Hair Jr. et al. (2014, pp. 233–234) the LOC scores estimated will be used as MVs from now on. This becomes possible due to the major advantage of PLS-PM, i.e. the availability of LVs. It is important to note that, although its path coefficient is rendered insignificant at the $\alpha = 0.01$ level for the current model, the LV scores of *Surface* will be used as MVs-scores in *model II* in order to stabilize results. The different relationships between subdimensions and *overall salience* may yield different path coefficients between LOCs and their HOC (cf. section 5.5.2).

5.5.2 *Constructs Do Have an Influence on Each Other*

Hypotheses About Cause-Effect Relationships Between Different Dimensions of Salience

The first model presented makes extensive use of the hypothesis that the subdimensions of *overall salience* are independent. As a consequence it proposes as few structural relationships as possible. However, several dependencies between all six LVs (cf. fig. 33) are reasonable based on prior empirical work or from a theoretical point of view. The structural relationships become fairly complex in this case, involving a number of instances of so-called mediation. Let me introduce this concepts first. Generally speaking, two different types of mediation can be distinguished, both of which can be applied to MVs and LVs, alike (cf. Preacher & Hayes, 2008). I will focus on LVs in the description and the figures given because of the importance of LVs in PLS-PM in general and in the model derived for this study in particular.

SIMPLE MEDIATION are cases, where a LV L_1 has a causal relationship to LV L_2 and to LV M which in turn has a causal relationship to LV L_2 (cf. fig. 30 which is based on Hair Jr. et al. (2014, p. 220)). The structural relationship between L_1 and L_2 is called direct effect. It equals the path coefficient between these. The indirect effect of L_1 on L_2 , which is mediated by M is given as product of the path coefficients of $L_1 \rightarrow M$ and $M \rightarrow L_2$. This means, the chain-rule applies to mediation analysis, i.e. the path coefficients of a causal chain are multiplied in order to find the size of the effect (cf. Alwin & Hauser, 1975). This indirect effect indicates the suitability of M

to mediate the impact L_1 has on L_2 . It is important to note that a significant effect of L_1 on L_2 is *not* a prerequisite of mediation. If the direct effect is significant mediation need not exist and vice versa. This means, in order to reveal a mediational effect, a significant indirect effect is sufficient (cf. X. Zhao, Lynch Jr., & Chen, 2010, p. 205). Table 27 provides an overview of the interpretation of significant indirect effects in a simple mediation context. These insights can be extrapolated to multiple mediation contexts.

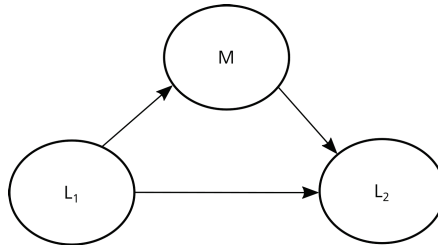


Figure 30.: An example of simple mediation. The direct effect L_1 has on L_2 is hypothesized to be mediated by an indirect effect via a third LV called M . The figure is based on Hair Jr. et al. (2014, p. 220).

Sign. indirect effect	Sign. direct effect	Interpretation
F	F	no evidence found for mediation or direct effect
F	T	no mediation found an omitted mediator may exist
T	F	mediation consistent to theoretical framework identified
T	T	product of direct and indirect effect > 0: Complementary mediation as both effects are positive < 0: Competitive mediation as direct effect acts as a suppressor An omitted mediator may be responsible for the significant direct path

Table 27.: The interpretation regarding the presence of mediation given the significance of direct and indirect effects. Cells of effects show truth values. Table adapted from the decision tree given in X. Zhao et al. (2010, p. 201).

MULTIPLE MEDIATION can be derived from simple mediation by substituting M with k LVs, where the relationship between L_1 and L_2 remains and L_1 shows a causal relationship with each M_k and each M_k has a structural relationship with L_2 (cf. fig. 31). In these cases the total indirect effect is built by summing all specific indirect effects (cf. Cohen et al., 2003, pp. 460–464). In terms of interpretation one must bear in mind a crucial difference to the single mediation context. If multiple mediators are completely uncorrelated the interpretation of single mediation applies for each of these. However, if multiple mediators are correlated, the specific effect of a single mediator of these “represents the ability [...] to mediate the effect of [...] [L_1 on L_2] conditional on the inclusion of the other mediators in the model” (Preacher & Hayes, 2008, pp. 881–882). As specific indirect effects may be positive and others may be suppressors, it is not a prerequisite that the total indirect effect must be significant (cf. Preacher & Hayes, 2008, p. 882). This means, it depends on the hypotheses formulated which effect is examined for significance, i.e. whether a specific indirect effect or the total indirect effect is analyzed.

THREE-PATH MEDIATED EFFECTS can only be hypothesized in multiple mediation contexts. A so-called three-path mediated effect (cf. Taylor et al., 2008) means that one mediator is a dependent variable of a second mediator. The upper part of fig. 32 provides an example for such a scenario. The effect of L_1 on L_2 is supposed to be mediated by the causal chain of $L_1 \rightarrow M_1 \rightarrow M_2 \rightarrow L_2$.

Based on these theoretical explanations, I will detail *model II* (cf. fig. 33) based on the following hypotheses:

- H_1 : The greater an object’s *visual salience* the larger its *overall salience* is.
- H_2 : The greater an object’s *cognitive salience* the larger its *overall salience* is.
- H_3 : The greater an object’s *structural salience* the larger its *overall salience* is.
- H_4 : The greater an object’s *visibility in advance* the larger its *overall salience* is.

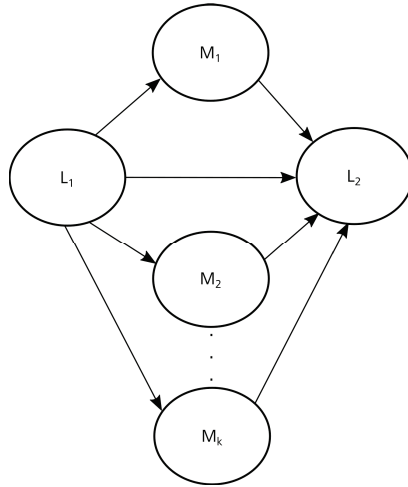


Figure 31.: An example of multiple mediation, i.e. those cases where the causal relationship between L_1 and L_2 is mediated by k variables. Figure adapted from Preacher and Hayes (2008, p. 881).

H_5 : The greater an object's *prototypicality* the larger its *overall salience* is.

H_6 : The greater an object's *visual salience* the easier it can be seen from advance.

H_7 : The greater an object's *visibility in advance* the more suitable it is to be included in route instructions.

H_8 : The greater an object's *prototypicality* the larger its *cognitive salience* is.

H_9 : The greater an object's *cognitive salience*, the easier it can be integrated in route instructions.

H_{10} : The effect *prototypicality* has on *overall salience* is mediated by *visual salience*.

These hypotheses reflect a proposed three-path mediated effect for *visual salience*. Visual aspects become salient at a very early stage of human perception and are consistent across individuals (cf. Caduff, 2007; Itti, 2007). Hence, it is interesting to see, whether and how deeply other

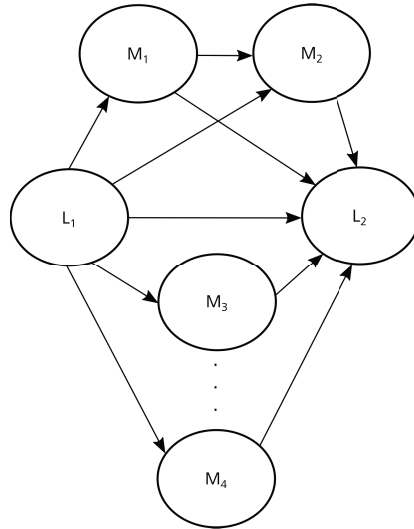


Figure 32.: An example of multiple mediation including a three-path mediation, (cf. upper half of the picture). The figure is based on Taylor et al. (2008, p. 243).

subdimensions are affected by it. The positive impact *visual salience* has on *overall salience* is modeled to be mediated by *visibility in advance*, which in turn has a positive influence on *structural salience*, which is positively related to *overall salience*, too. A rationale to propose a positive influence of *visibility in advance* on *structural salience* can be based on the understanding of *visibility in advance* (cf. section 2.4). If an object can be recognized from afar it can be more easily integrated into route instructions. Similarly, visual features contribute to the degree as to which an object is identifiable from afar (imagine, e.g. a blue colored house in a neighborhood, where all other houses are painted white). Furthermore, the hypotheses presented indicate a multiple mediation for *prototypicality*. On the one hand, it is mediated by *visual salience*, which is reasonable based on the fact that mental images of objects may well guide our visual attention on the pre-attentive level (cf. section 2.3.2). On the other hand, *prototypicality* is supposed to have a positive influence on *cognitive salience* because prototypical objects may eventually be conceptualized more easily. This presumably has, in turn, a positive effect on the value the object has for use in route instructions, i.e. on

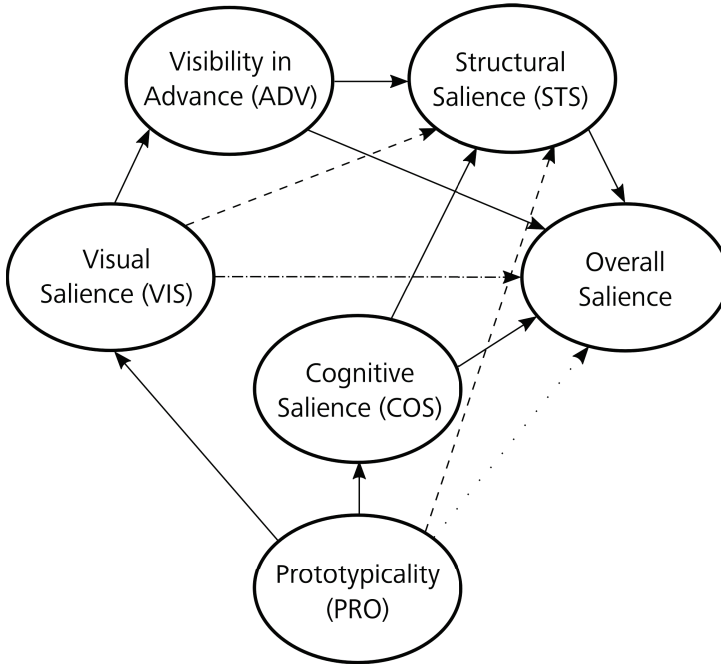


Figure 33.: The structural relationships of *model II*. The dotted line $PRO \rightarrow OVSAL$ reflects the full mediation via VIS . The paths $VIS \rightarrow STS$ and $PRO \rightarrow STS$ are added in order not to inflate unexplained variance. Finally, the path $VIS \rightarrow OVSAL$ was dashed and dotted in order to indicate that a partial mediation of this effect is hypothesized. The figure was drawn using Inkscape (cf. The Inkscape Team, 2016).

structural saliency. As it is common not to model direct paths in mediator analysis (cf. X. Zhao et al., 2010, pp. 204–205), I want to stress that this is done purposefully in the hypotheses H_1 to H_5 . This study is the first one proposing a larger model of *overall saliency*. Based on prior empirical evidence full mediation cannot be assumed. It is important to note, moreover, that these hypotheses are motivated by the aim of establishing a causal chain, which is a major difference to existing models. Raubal and Winter (2002) propose different weights for visual, semantic and structural attraction based on its significance. This means, they do not account for any kind of impact that measures may have on one another. Similarly, the Bayesian network presented by Caduff (2007)

does not include any connections between high-level components such as *visual salience* or *cognitive salience* (cf. section 2.3.3).

Given these hypotheses, *prototypicality* is the only exogenous LV remaining (compared to the first model proposed). This means, *structural salience*, *visual salience*, *cognitive salience* and *visibility in advance* are now endogenous LVs, whereas *overall salience* is endogenous in both models (cf. fig. 33). Before the results are reported, it is important to note that a number of assumptions regarding correctness apply to three-path mediated models.

“As outlined by James, Mulaik, and Brett (1982), correct specification includes specification of the causal order of the variables, specification of the causal direction (no reciprocal paths are estimated), the assumption that the model is self-contained (i.e., there are no omitted variables; this includes the assumption that interactions need not be included), the assumption that there are no moderator effects, and the assumption that the model is stable” (Taylor et al., 2008, p. 265).

The Results

According to Taylor et al. (2008) as well as Preacher and Hayes (2008) many different approaches to mediation analysis have been proposed since the mid-80s. The methods proposed to analyze multiple mediation can be divided into three different classes. Basically, these classes as well as the methods can be seen as extrapolation of the simple mediation case.

CAUSAL STEPS Using this approach means to analyze paths involving the mediators one by one. This approach was suggested by Baron and Kenny (1986). According to Preacher and Hayes (2008, p. 880) this mode of analysis has become fairly popular ever since. Mediation is said to be found, when both paths building a specific indirect effect are significant. However, “[i]t relies on a set of tests of individual [...] paths rather than testing the specific indirect effects, and yields no point estimate or SE of the mediation effect” (Preacher & Hayes, 2008, p. 882). Therefore, other approaches are more suitable for analysis of multiple mediators (cf. section 5.5.2).

PRODUCT OF COEFFICIENT APPROACH This approach is also called Sobel's z -test (cf. Sobel, 1982) in the case of simple mediation, where z is calculated by relating the indirect path to its standard error. For multiple mediator scenarios a method suitable to measure the standard error of the total indirect effect was developed. As the total indirect effect equals the sum of all specific indirect effects, the variance of the total indirect effect can be estimated. Preacher and Hayes (2008, p. 882) provide an example for three mediators based on methods proposed by Bollen (1987) which can be easily generalized to more than three mediators. Assuming normality of the total indirect effect, its first-order standard error is given as square root of the variance. This assumption is a major drawback, particularly as multivariate normality is required for all constituents of indirect effects. Moreover, "the sampling distributions of the total and specific indirect effects are [required to be] normal" (Preacher & Hayes, 2008, p. 883). These assumptions, however, are very rarely met in real world scenarios.

BOOTSTRAPPING is suitable to provide CIs based on resampling and therefore overcomes the requirement of multivariate normality. This method is to be preferred in

- scenarios involving simple mediation (cf. Preacher & Hayes, 2004),
- multiple mediation models (cf. Preacher & Hayes, 2008),
- three-path mediation contexts (cf. Taylor et al., 2008; Williams & MacKinnon, 2008).

Despite the popularity of the causal steps approach, testing multiple mediators simultaneously yields several advantages (cf. Preacher & Hayes, 2008, p. 881):

1. "[T]esting the total indirect effect of X on Y is analogous to conducting a regression analysis with several predictors, with the aim of determining whether an overall effect exists".
2. "[D]etermine to what extent specific M variables mediate the $X \rightarrow Y$ effect, conditional on the presence of other mediators in the model".

3. “[D]etermine the relative magnitudes of the specific indirect effects associated with all mediators”.
4. “[T]he likelihood of parameter bias due to omitted variables is reduced [. . .] (Judd & Kenny, 1981)”.

Given these reasons it seems to be advisable to conduct a single multiple mediator analysis. In multiple mediation contexts this requires two steps (cf. Preacher & Hayes, 2008, p. 882):

1. Assess the total indirect effect, i.e. the sum of all specific indirect effects; this ensures that the mediators transmit the effect L_1 has on L_2 .
2. Examine specific indirect effects separately in order to gain an insight into the role a mediator plays individually. This does not imply that a significant total direct effect is a prerequisite for analyzing specific indirect effects.

Based on these explanations, the advice cited above, and because PLS-PM already makes extensive use of the approach bootstrapping is used to assess the hypotheses presented. Particular attention must be paid to the way the CIs are constructed. Whereas significance of path coefficients is assessed using the BCa-CIs in *model 1*, the way CIs are constructed needs to be discussed for multiple mediation and three-path mediated effects.

MULTIPLE MEDIATION CONTEXT

According to X. Zhao et al. (2010, p. 202) the discussion about the use of bootstrapping started with Preacher and Hayes (2004). This paper provides SPSS and SAS routines for Sobel’s z-test as well as bootstrapping approaches in simple mediation contexts. They propose the percentile bootstrap to build CIs but do not give any formal justification for it. MacKinnon, Lockwood, and Williams (2004) stress that the BCa-CIs have the highest power of all methods tested and its CIs is most correct. However, type I error excess can happen with this method. Four years later, Preacher and Hayes (2008) compares strategies for multiple mediator analysis. Again, routines for SPSS and SAS are provided. They recommend “to

use bootstrapping [...] whenever possible" (Preacher & Hayes, 2008, p. 886) based on the results achieved in a real world example application. This is in-line with the recommendation of Williams and MacKinnon (2008), who recommend bias-corrected bootstrap techniques based on an appropriate number of bootstrap samples. Furthermore, they emphasize that with larger sample sizes, the difference in methods will be negligible.

THREE-PATH MEDIATION

In their simulation study based on a model involving multiple mediation and three-path effects Williams and MacKinnon (2008) regularly found a bias in CIs. However in the case of true effects greater than zero, the results obtained using a bias-corrected bootstrap were most balanced. Once again, one must be aware that the results achieved were rather weak due to the small bootstrap sample size of $N = 1000$. As a consequence, Taylor et al. (2008) investigate three-path mediated effects in detail. By means of a Monte Carlo simulation they analyze six different methods, among them percentile and bias-corrected bootstrapping as well as methods based on multivariate normality assumptions. Based on type I error, power and coverage, both bootstrapping methods were the only ones with desirable properties which are suitable to construct a CI. Although bias-corrected bootstraps consistently showed the highest power out of all methods, their type I error rate was above the nominal level under certain conditions. Therefore, Taylor et al. (2008) recommend using the percentile bootstrap to test three-path mediated effects because this method only has slightly less power but is able to keep type I errors below nominal level in all cases. This means, Taylor et al. (2008) give a more conservative advice than MacKinnon et al. (2004) do for multiple mediation scenarios.

Based on this discussion, I will base the report of CIs on the percentile bootstrap for all indirect effects analyzed, i.e. for two- and three-path mediated effects. This is reasonable as I regard the control of type I error to be more important than a slight increase in power. This means, I adhere to the recommendation of Taylor et al. (2008) given with respect to the three-path mediated effect. A second argument

in favor of using the percentile bootstrap is sample size. The whole analysis is based on a sample with an a priori calculated sample size. Furthermore, bootstrapping results are based on $K = 5000$ resamples. Based on Williams and MacKinnon (2008) both aspects taken together should render the differences between the methods negligible. Overall, these aspects provide reasonable grounds to use a simple percentile bootstrap. Having said this, the parameters of the PLS-PM algorithm are not altered compared to the first analysis (cf. section 5.4). This means, path weighting scheme was used, initial weights were set to +1, a maximum of 300 iterations until convergence was applied, the stop criterion was set to 10^{-5} , measured variable values were z-standardized, $K = 5000$ bootstrap samples were drawn, and the *no sign-change* option was applied.

One must keep in mind that the hypotheses described have an impact on the structural level only. As described earlier (cf. section 3.3.2) neither outer weights nor outer loadings are affected by these changes. This means, the reflective measurement model parts need not be reassessed because none of the quality measures takes structural relationships into account. Similarly, the formative causes of the LOCs used to estimate *visual salience* as a HOC need not be reassessed in terms of measurement quality.

Having said this, the PLS-PM results of the new model can be assessed (cf. fig. 34). Multicollinearity between antecedents of LVs is not a threat to the validity of the inner model ($1.00 \leq VIF \leq 3.46$).

Endogenous construct	R^2	Q^2
<i>overall salience</i>	0.709 (adj. 0.705)	0.611
<i>structural salience</i>	0.680	0.504
<i>visibility in advance</i>	0.437	0.325
<i>visual salience</i>	0.090	NA
<i>cognitive salience</i>	0.059	0.017

Table 28.: Based on the R^2 values the model shows very reasonable predictive validity ($R^2 = 0.709$) with respect to the target construct *overall salience*. Furthermore, it has high predictive relevance ($Q^2 = 0.611$). Please note: Q^2 values were estimated using omission distance $D = 7$ during the blindfolding procedure. As *visual salience* is measured formatively, Q^2 cannot be applied.

5 EMPIRICAL RESULTS

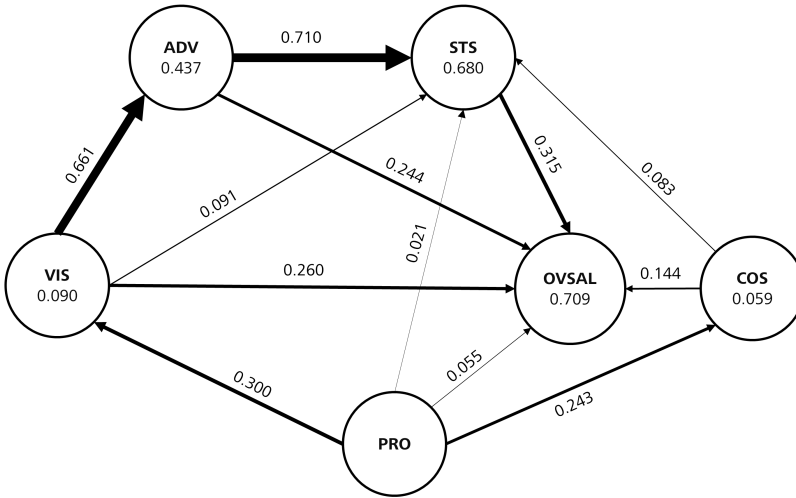


Figure 34.: The structural model results for model II. The LVs show R^2 values achieved, the annotations to lines are the path coefficients estimated. The figure was created using Inkscape (cf. The Inkscape Team, 2016).

²⁸Table 28 and table 29 show the high predictive validity and relevance the model has with respect to the target construct *overall salience*. As a consequence, the predictive validity of both models analyzed is very similar (cf. their adj. R^2 values). It is important to note that high predictive validity and relevance are also revealed for *structural salience* ($R^2 = 0.680$, $Q^2 = 0.504$) in the current model. This is of particular interest, when path coefficients (cf. table 31) are considered. Four LVs point to *structural salience*, but the only significant direct effect is $ADV \rightarrow STS$. This effect, however, is the largest (0.710, $p < 0.001$) direct effect found. This means, *visibility in advance* accounts for a very large proportion of variance in *structural salience*. Table 30 further stresses this aspect by re-

²⁸ Except for effect sizes f^2 and q^2 the way results are presented is based on Klärner et al. (2013).

	$R^2(OVSAL)$	$Q^2(OVSAL)$	f^2	q^2
complete	0.721	0.611	NA	NA
w/a VIS	0.682	0.587	0.140	0.084
w/a ADV	0.707	0.607	0.050	0.031
w/a STS	0.690	0.593	0.111	0.068
w/a PRO	0.718	0.617	0.010	0.005
w/a COS	0.709	0.610	0.043	0.024

Table 29.: Predictive accuracy and predictive relevance of LVs pointing to *overall salience*. The figures reveal the major impact *visual salience* and *structural salience* have.

	$R^2(STS)$	$Q^2(STS)$	f^2	q^2
complete	0.680	0.504	NA	NA
w/a VIS	0.680	0.504	0.003	0.000
w/a ADV	0.429	0.313	0.790	0.377
w/a PRO	0.681	0.502	0.000	-0.002
w/a COS	0.675	0.497	0.019	0.008

Table 30.: Predictive accuracy and predictive relevance of LVs pointing to *structural salience*. The figures indicate the major importance *visibility in advance* has.

vealing high effect sizes for f^2 and q^2 . Similarly, *visual salience* has a strong effect on *visibility in advance* providing evidence for H_6 , which is the second largest effect found, overall. These path coefficients are twice as large as the third largest path coefficient found ($STS \rightarrow SAL$: 0.315, $p < 0.001$). Coming back to predictive validity and relevance values, those values achieved for *visibility in advance* are still reasonable ($R^2 = 0.437$, $Q^2 = 0.325$), but also clearly smaller than those of *overall salience* and *structural salience*. This is a clear indication of the positive impact *visual salience* has on *visibility in advance*, because *visual salience* is the only LV with an incoming connection at *visibility in advance*. At the same time this result fosters the notion that *visibility in advance* and *visual salience* are related concepts but not equal. The meaning of the small values for *visual salience* ($R^2 = 0.090$) and *cognitive salience* ($R^2 = 0.059$, $Q^2 = 0.017$) can be explained by the arrow heads pointing towards them, i.e. *prototypicality* is almost not suitable to explain variance in these LVs.

effect	path coeff.	BCa 99% – CI
<i>ADV</i> → <i>STS</i>	0.710***	[0.607; 0.792]
<i>VIS</i> → <i>ADV</i>	0.661***	[0.592; 0.742]
<i>STS</i> → <i>OVSAL</i>	0.315***	[0.178; 0.4449]
<i>PRO</i> → <i>VIS</i>	0.300***	[0.192; 0.423]
<i>VIS</i> → <i>OVSAL</i>	0.260***	[0.152; 0.366]
<i>ADV</i> → <i>OVSAL</i>	0.244***	[0.093; 0.397]
<i>PRO</i> → <i>COS</i>	0.243***	[0.112; 0.380]
<i>COS</i> → <i>OVSAL</i>	0.144***	[0.062; 0.234]
<i>VIS</i> → <i>STS</i>	0.091*	[-0.018; 0.220]
<i>COS</i> → <i>STS</i>	0.083*	[-0.005; 0.177]
<i>PRO</i> → <i>OVSAL</i>	0.055n.s.	[-0.027; 0.133]
<i>PRO</i> → <i>STS</i>	0.021n.s.	[-0.054; 0.094]

Table 31.: The simple path coefficients found in the complex model. *** indicates $p < 0.001$ and * means $p < 0.05$ ($K = 5000$ resamples). BCa-CIs are used because direct effects are reported.

Table 31 shows that H_1 to H_4 hold, indicating that *visual salience* (0.260, $p < 0.001$), *visibility in advance* (0.244, $p < 0.001$), *structural salience* (0.315, $p < 0.001$) and *cognitive salience* (0.144, $p < 0.001$) have a highly significant positive effect on *overall salience*. The direct effects *structural salience*, *visual salience* and *visibility in advance* have on the key target construct *overall salience* are equal and medium sized. Finally, *cognitive salience* has a highly significant but rather small effect on *overall salience*. It is important to note that the relationships *VIS* → *STS* and *PRO* → *STS* were only included in the model for the purpose of stabilizing the results (cf. Taylor et al., 2008). Consequently, they need not be further examined. The insignificant direct effect *prototypicality* has on *overall salience* (0.055, $p > 0.05$) leads to the analysis of the mediators hypothesized starting with the analysis whether the effect of *prototypicality* on *overall salience* is fully mediated. Regarding the three-path mediated effect *PRO* → *COS* → *STS* → *OVSAL* (i.e. H_8 and H_9), no significant effect was found (cf. section 5.5.2). However, evidence for H_{10} was provided, i.e. *visual salience* mediates the effect *prototypicality* has on *overall salience*. Although significant, the effect is very small (0.078, $p < 0.01$). Finally,

a partial mediation of the effect *visual salience* has on *overall salience* is revealed, as the medium sized effect $VIS \rightarrow OVSAL$ is significant (0.260, $p < 0.001$) as well as the proposed three-path mediated effect (0.148, $p < 0.001$). The amount of variance $VIS \rightarrow ADV \rightarrow STS \rightarrow OVSAL$ accounts for is 25%. This interpretation as partial mediation is based on Hair Jr. et al. (2014, p. 225). They state that partial mediation is present, if an indirect effect accounts for 20% to 80% of the total effect. This result reinforces the importance of *visual salience* per se, but leaves room for other mediators not hypothesized, as well.

effect	specific indir. effect	Percent. 99%-CI
$VIS \rightarrow ADV * ADV \rightarrow STS * STS \rightarrow OVSAL$	0.148**	[0.077; 0.229]
$PRO \rightarrow COS * COS \rightarrow STS * STS \rightarrow OVSAL$	0.006n.s.	[0; 0.017]
$PRO \rightarrow VIS * VIS \rightarrow OVSAL$	0.078**	[0.037; 0.129]

Table 32.: Assessment of the proposed indirect effects. ** denotes $p < 0.01$ ($K = 5000$ resamples).

Reassessing the Influence of LOCs

As *visual salience* as a HOC is now part of a different, complex nomological net of LVs a reassessment of the influence the LOCs have on *visual salience* is appropriate. As noted earlier (cf. section 5.5.1), I will use the LOC-scores, i.e. outer weights will be reported instead of path coefficients. Section 5.5.2 reveals no difference with respect to Surface as its path coefficient remains insignificant.

LOC	OW	Std. Err	99%-CI
Geometry	0.491***	0.066	[0.317; 0.646]
Secondary	0.372***	0.073	[0.179; 0.545]
Color	0.316***	0.062	[0.150; 0.466]
Surface	0.096	0.066	[-0.067; 0.265]

Table 33.: The outer weights (OW) of LOC-LV-scores used as MVs, the standard error of estimated weights (Std. Err.) and the 99%-BCa-CI. *** indicates $p < 0.001$ ($K=5000$ resamples). LOCs with highly significant outer weights are bold-faced.

Furthermore, Geometry is the most important among all LOCs, showing a highly significant outer weight close to 0.5. A difference to the earlier assessment with respect to sizes of outer weights occurs for Secondary and Color. According to the new model both are roughly equally important, with Secondary having a slightly higher outer weight. This reinforces the surprising result that Color is far less important as one might expect based on everyday experience.

5.6 CHAPTER SUMMARY

This chapter details the empirical results of the main study of this thesis. Several important insights were made analyzing different models using PLS-PM methodology. All theoretical concepts needed for the statistical analysis were introduced before their usage was described.

PARTICIPANTS, OBJECTS, ROUTES

Data acquisition for the main experiment required $n = 119$ trials overall and was spread across 48 days. On average, trials lasted $\bar{x} = 60min$. The beginning of trials were scheduled between 09:30 am and 3 pm in order to allow for comparable daylight conditions. Of the 111 usable trials, 61% of the participants were females and 16% were non-students. The majority of the participants was in their mid-twenties ($\bar{x} = 25$ years) with an age range of 18 to 65 years. Objects rated spanned a variety of different categories, and individual buildings only account for roughly $\frac{2}{3}$ of all objects. This means, 120 of the rated objects were non-buildings. Routes were 1.50 km long, a median number of six objects had to be rated on each route, and routes were randomly assigned to participants.

NO INFLUENCE OF SNOWY WEATHER

Due to the long-lasting experiment phase, different weather conditions were unavoidable. Vision constraints were controlled by cancellation of scheduled trials in case of heavy fog/snowfall/rain. However, on four days the city was covered by snow. It seems reasonable to assume that resulting differences with respect to visual occurrence may have had a particular impact on *visual salience* rat-

ings. Therefore, the ratings of these days were examined by means of a bootstrapping based analysis. No severe differences between snow cover ratings and other ratings was revealed, i.e. all ratings were used.

APPROPRIATE REFLECTIVE MEASUREMENT MODEL

PLS-PM fit starts with the assessment of the reflective measurement model parts. Appropriateness of measurement was examined based on internal consistency, convergent and discriminant validity (cross loadings and Fornell-Larcker criterion). Based on these measures, a very reasonable fit including the fact that all outer loadings were significant, was revealed for indicators used to measure *overall salience*, *structural salience* and *visibility in advance*. Regarding *cognitive salience* and *prototypicality* some of the indicators had to be removed or showed only medium sized outer loadings. The latter of these were further analyzed with respect to their impact on AVE.

APPROPRIATE FORMATIVE MEASUREMENT OF VISUAL SALIENCE

Assessing the appropriateness of formative measurement for *visual salience* comprised three steps. Firstly, a convergent validity analysis justified formative measurement per se. Secondly, multicollinearity between causes was examined by means of VIF. As all VIF-values were well beyond 5, correlations between causes are not a threat to the results achieved. Finally, the outer weights and loadings of causes were assessed. Among the 15 causes used, the model estimation suggests to exclude *width*, *height*, *length* and *area* from the model. All of them show a non-significant outer weight combined with a reasonable correlation with *size*, which shows both, highly significant outer weight and loading. This correlation ensures that the meaning of the construct is not altered by excluding these four causes. Taken as a whole, the large number of causes used to measure *visual salience* fosters the probability of attaining insignificant results because of the theoretical threshold of outer weights in the case of completely uncorrelated variables. Therefore, it is advisable to restructure this part of the measurement model by means of a higher-order analysis for visual dimensions.

STRUCTURAL RELATIONSHIPS OF MODEL I

The structural relationships of the first model make extensive use of the hypothesis of independence between LVs. Assessing these relationships revealed a reasonable predictive validity ($R^2 = 0.729$). All LVs except *prototypicality* show a significant relationship with *overall salience*. *Structural salience* and *visual salience* are rendered to be most and equally important (path coefficients of 0.312***), followed by *visibility in advance* (0.224***) and *cognitive salience* (0.119***).

HIGHER-ORDER ANALYSIS OR REGROUPING VISUAL DIMENSIONS

Four LOCs were built by grouping related causal-formative indicators of *visual salience*. Choosing aptonyms, the constructs were called *Color*, *Geometry*, *Surface*, and *Secondary*, respectively. An analysis of the structural relationships between the LOCs and their HOC revealed highly significant relationships for all LOCs except *Surface*. According to these results, *Geometry* has the largest effect on *visual salience*. *Model II* results suggest that *Secondary* is even more important than *Color* is.

MEDIATION INSTEAD OF INDEPENDENCE—MODEL II

Following a common approach in the analysis of SEMs, a second model was compiled, modeling the impact LVs have on each other, i.e. independence of LV was no longer assumed. This model uses multiple mediation and three-path mediation. Using this model, *prototypicality* is the only exogenous construct remaining. The model shows high predictive validity and relevance for the key target construct *overall salience*. The effect of *prototypicality* on *overall salience* is fully mediated: *visual salience* is confirmed to be a mediator of *prototypicality*, but the second mediation hypothesized for *prototypicality* via *cognitive salience* and *structural salience* is rendered insignificant. The impact *visual salience* has on *overall salience* is only partially mediated by the chain *visual salience* → *visibility in advance* → *structural salience* → *overall salience*. Further mediators, which are not hypothesized, are consequently likely to exist. Finally, *visibility in advance* is shown to have a high impact on *structural salience*.

6

DISCUSSION

6.1 INTRODUCTION

This chapter provides a detailed discussion of the findings reported in previous chapters. The first part is dedicated to methodological discussions on PLS-PM and their meaning with respect to how the results should be interpreted. The second part of the chapter deals with the discussion of measurement model components and the structural model. At the end of the second part, the predictive capability will be discussed leading to the question of whether constructs might be missing from the current models. Finally, general limitations of the study are discussed.

6.2 ONGOING METHODOLOGICAL DISCUSSION ABOUT PLS-PM

As mentioned above (cf. section 5.4), the analysis reported in this thesis is based on up-to-date textbooks on SEM in general and PLS-PM in particular. Although Hair Jr. et al. (2014) published their textbook less than two years ago, an in-depth methodological discussion has since taken place (parts of it have started even earlier). There are three facets which are important for this thesis: formative measurement in general, the estimation of formative models using PLS-PM and PLS-PM in general. I briefly address each of these aspects and discuss the importance of these points for the results of this thesis.

6.2.1 *Formative Measurement in General*

Starting in late 2014, a series of articles dealing with formative measurement in general has been published. This discussion is important for the results reported because *visual salience* is modeled formatively.

The series started with Bainter and Bollen (2014). Several commentaries on the issues raised by Bainter and Bollen were published in the same issue of the journal *Measurement: Interdisciplinary Research and Perspectives* (vol. 12 (4), 2014). Both authors responded to those comments in Bainter and Bollen (2015). Basically, Bainter and Bollen (2014) have two goals. They provide empirical evidence that coherent weights for causal-formative indicators can be achieved across different models, if those models are specified correctly. Secondly, they question existing evidence suggesting that causal-formative indicators are *inherently* prone to interpretational confounding, which is defined as:

“the assignment of empirical meaning to an unobserved variable which is other than the meaning assigned to it by an individual a priori to estimating unknown parameters. Inferences based on the unobserved variable then become ambiguous and need not be consistent across separate models” (Burt, 1976, p. 4).

Detailing the whole discussion is beyond the scope of this thesis. Instead, I will give a short summary of the commonalities and differences regarding the original article and the comments on it based on Bainter and Bollen (2015). Four areas of agreement were identified:

1. Causal indicators yield stable results across correctly specified models (cf. Bainter & Bollen, 2015, p. 64).
2. Reflective measurement is not appropriate for particular MVs, i.e. different types of indicators must be distinguished (cf. Bainter & Bollen, 2015, p. 65).
3. The decision whether reflective measurement or formative measurement is employed must not be based on empirical but on theoretical considerations (cf. Bainter & Bollen, 2015, p. 65).
4. Theory on causal-formative scale invention is less developed than its reflective counterpart is (cf. Bainter & Bollen, 2015, p. 65).

Given these agreements, there are several aspects, which are disputed by Bainter and Bollen and the commentators.

1. According to Bainter and Bollen (2015) a construct has substantive meaning apart from any kind of measurement at all, whereas some commentators equate identification and substantive meaning.
2. Some authors give arguments, why they think causal indicators are per se very subjective. In their rejoinder, Bainter and Bollen (2015, p. 69) reassure that there is no inherent mode of any type of measurement for any LV at all, but that mode of measurement is always a matter of viewpoint. Differing coefficients across studies give, therefore, raise to question the particular causal indicators but not causal-formative measurement in general. With respect to misspecification, causal-formative as well as reflective indicators, are prone to cases, where a whole model is wrongly specified because the meaning of the LV is different across studies.
3. Bainter and Bollen (2015) make an important clarification regarding the requirement to include all important causal-formative indicators. In contrast to standard textbook texts (cf. e.g. Hair Jr. et al., 2014, p. 43), completeness is not the most important part, but the assumption, that the disturbance resulting from neglecting causal indicators is independent from those causal-formative indicators included in the model.
4. While all participants in the discussion stress the major importance of theoretical justifications for models, some claim that formative measurement models are more likely to provoke bad scientific practices, i.e. changing the model based on estimations achieved. Bainter and Bollen (2015) disagree on this potentially enhanced danger when formative measurement is applied.

In my opinion, in the context of this discussion, convincing arguments are given by Bainter and Bollen (2015). Hence, the discussion is not a threat to the use of formative measurement in general and in this doctoral work in particular. It is interesting to note, though, that the discussion presented does not touch the difficulties arising from common misunderstandings in formative measurement. The whole aforementioned debate in the journal *Measurement: Interdisciplinary Research and Perspectives* (vol. 12 (4), 2014) does not make clear what types of formative

measurement the discussion is based on. In late 2015, two of the most influential proponents of formative measurement, Kenneth Bollen and Adamantios Diamantopoulos, publish a journal article aiming to clarify important misunderstandings about formative measurement. These stem from the fact that different types of formative measurement are often not distinguished. As a consequence, the applicability of formative measurement in general is questioned increasingly often, which is the reason why a general clarification becomes necessary (cf. Bollen & Diamantopoulos, 2015). Indeed, standard textbooks on PLS-PM like Hair Jr. et al. (2014) or SEM in general like Weiber and Mühlhaus (2010) do not make any of the distinctions Bollen and Diamantopoulos (2015) make. This is, in my opinion, a severe limitation of these textbooks. According to Bollen and Diamantopoulos (2015) five different ways of specifying formative measurement models exist. These are built along four different conditions (cf. Bollen & Diamantopoulos, 2015, pp. 3–4).

1. **Conceptual Unity** for the indicators must be given, i.e. all of them must “correspond to the concept’s meaning” (Bollen & Diamantopoulos, 2015, p. 3). This meaning must be derived based on theory before indicators used to measure the LV are sought for.
2. **Indicators** must be measures of a LV, which is “a random (or nonrandom) variable for which there is no sample realization for at least some observations in a given sample” (Bollen, 2002, p. 612).
3. **Error** is present in the LV and modeled separately.
4. **Estimation** of coefficients is done based on empirical data.

According to Bollen and Diamantopoulos (2015) the conditions one and two must be fulfilled whereas the conditions three and four can be met in order to have causal-formative indicators. Based on these assumptions five different specifications are derived (cf. Bollen & Diamantopoulos, 2015, p. 4):

$$\text{A } \eta = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_n x_n + \zeta$$

$$\text{B } \eta = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_n x_n$$

$$\text{C } \eta = \gamma_1^* x_1 + \gamma_2^* x_2 + \cdots + \gamma_n^* x_n$$

$$D \quad C = w_1x_1 + w_2x_2 + \cdots + w_nx_n$$

$$E \quad C = w_1^*x_1 + w_2^*x_2 + \cdots + w_n^*x_n$$

While specification **A** is the general case, **B** denotes a special case, where the error term equals zero and **C** is a special case of **A** with no error and prespecified weights. Specifications **D** and **E** are not regarded to be causal-formative indicators but are composite indicators, i.e. construct **C** is a linear composite where the condition of conceptual unity is not given (cf. Bollen & Diamantopoulos, 2015, p. 4). However, if models involving formative measurement are estimated using PLS methods, specification **B** and **D** become indistinguishable. This is because **B** is a special case of **A** and **A** cannot be estimated using PLS-PM because construct level errors cannot be modeled in variance-based estimations (cf. Bollen & Diamantopoulos, 2015, p. 4). Based on these clarifications, Bollen and Diamantopoulos (2015) discuss seven commonly occurring criticisms against formative measurement. I quote these below from Bollen and Diamantopoulos (2015) and provide their main argument with respect to each.

1. *“A construct measured with formative indicators does not exist independently of its indicators”* (Bollen & Diamantopoulos, 2015, p. 5).

This criticism is answered from three different perspectives. First, it stems from the obvious confounding of composite-formative and causal-formative indicators. Second, it ignores the fact that a construct and its meaning must be given before any indicators are found to measure it. Third, there is a risk in changing the empirical meaning of a construct using a set of specific indicators, but this risk is not exclusive to causal-formative indicators (cf. Bollen & Diamantopoulos, 2015, p. 7).

2. *“Formative indicators are causes rather than measures”* (Bollen & Diamantopoulos, 2015, p. 5).

One flaw of this criticism is to mix up different uses of formative as a term. Bollen and Diamantopoulos (2015) agree with this criticism, if it refers to specifications **D** and **E**. In these cases the composite variable **C** is nothing else than the sum of indicators each of which is weighted (cf. Bollen & Diamantopoulos, 2015,

- p. 7). However, the criticisms are put forward to specifications **A** and **B** as well. Bollen and Diamantopoulos (2015, p. 7) provide convincing arguments that these criticisms are based on an understanding of measurement suitable to reflective measurement only.
3. *“Formative indicators imply multiple dimensions to a construct and this is a liability”* (Bollen & Diamantopoulos, 2015, p. 5).
This criticism against causal-formative indicators is based on the assumption that these are multidimensional per se. This is doubted by Bollen and Diamantopoulos based on the argument that the dimensionality of a construct is solely a matter of theoretical definition. If it is defined as unidimensional construct, its indicators are tied to a single dimension, no matter whether they are causal-formative or reflective (cf. Bollen & Diamantopoulos, 2015, p. 9).
 4. *“Formative indicators are assumed to be error-free, which is unrealistic”* (Bollen & Diamantopoulos, 2015, p. 5).
Error may occur in any type of indicator, i.e. regardless the type of measurement in which they are used. If the error is non-negligible it must be modeled or obtaining inaccurate estimations must be accepted. This aspect applies to reflective and formative indicators, alike (cf. Bollen & Diamantopoulos, 2015, p. 10).
 5. *“Formative indicators are inherently subject to interpretational confounding”* (Bollen & Diamantopoulos, 2015, p. 5).
According to Bollen and Diamantopoulos this argument stems from two different problems. First, CB-SEM proponents treat covariance as the cause of the model’s parameters instead of acknowledging that the parameters of the model yield the covariance matrix. Second empirical evidence suggesting that the outcome of causal-formative indicators differs per se across models suffers from model misspecifications itself (cf. Bollen and Diamantopoulos (2015, p. 12) who refer to Bainter and Bollen (2014) for an in-depth analysis of these misspecifications).
 6. *“Formative indicators fail proportionality constraints”* (Bollen & Diamantopoulos, 2015, p. 5). By means of a simple example, Bollen

and Diamantopoulos (2015, pp. 12–13) provide an argument that assessing proportionality constraints can be used as a check of causal-formative indicator validity rather than abandon these in general.

7. “*The coefficients for formative indicators should be set in advance*” (Bollen & Diamantopoulos, 2015, p. 6). As there is a difference between causal-formative indicators and composite-formative indicators, there is no need to abandon estimations of causal-formative indicators in general. This means that there can be cases, when setting the coefficients a priori is reasonable, but this is not the general case (cf. Bollen & Diamantopoulos, 2015, p. 13).

I find the clarifications provided by Bollen and Diamantopoulos (2015) including the counterarguments given as a reaction to the criticisms are convincing. Consequently, there is neither reason to abandon formative measurement in general nor in the case of *visual salience* for this doctoral work. Indeed, the questions were phrased by referring to a certain degree a feature contributes to *visual salience*. They show conceptual unity, consequently, which in turn means that each of them is a causal-formative indicator for *visual salience* as classified by Bollen and Diamantopoulos (2015).

6.2.2 PLS-PM and Formative Measurement

In 2014 an important criticism about PLS-PM and formative measurement was thoroughly examined. The criticism specifically deals with a particular class of formative constructs, i.e. those which are exogenous to others, but are endogenous itself and its predictor construct is measured reflectively. This is important for the results of *model II*, where *visual salience* is an endogenous formative construct of this kind due to its structural relationships with *visibility in advance*, *overall salience*, and *prototypicality*. The discussion started with a research note issued by Aguirre-Urreta and Marakas (2014b). Based on theoretical work about the PLS-PM algorithm and a Monte Carlo simulation they claim

to reveal an important aspect: “[PLS-PM] will not accurately capture the relationship between the two constructs, because of the lack of shared variance between the indicators of each that is due to that relationship” (cf. Aguirre-Urreta & Marakas, 2014b, p. 776). Rigdon et al. (2014) is a rebuttal to Aguirre-Urreta and Marakas (2014b), which provides arguments for three major flaws in the research design. These are (cf. Rigdon et al., 2014, pp. 781–783):

1. the use of CB-SEM to create data which is to be analyzed using PLS-SEM,
2. the use of a PLS path model which does not correspond to the proposed population model,
3. the whole model setup, which does not represent formative measurement because the equations presented suggest that “the three observed variables and the predictor construct *jointly* [sic] predict η_1 ” (Rigdon et al., 2014, p. 781).

The comment on Rigdon et al. (2014) by Aguirre-Urreta and Marakas (2014a) deals with all concerns separately. Regarding the difference between population and estimation model, Aguirre-Urreta and Marakas (2014a) stress that this difference was the important point they tried to make. With respect to the criticism regarding the data generation process, they justify their procedure by stressing that this is the way their target audience uses PLS-PM. From their point of view the Information Systems community does estimate population factor models by using PLS-PM. This is the reason why they use a factor model to generate their data. Finally, the objection against the model equation of the whole analysis is rebutted by references to publications with a similar model setup. Overall, while I can only give some highlights of this discussion, I think the most important problem in this scientific exchange is a matter of misunderstandings. This becomes particularly evident in case of the data generation process. Rigdon et al. (2014) provide arguments at the theoretical level which are based on the mathematical differences between CB-SEM and PLS-PM. In contrast, Aguirre-Urreta and Marakas (2014b) refer to the approach which is commonly used in practice. In the light of these misunderstandings, I tend to follow the theoretical

arguments given by Rigdon et al. (2014). As a consequence, I do not regard the evidence provided in Aguirre-Urreta and Marakas (2014b) as a threat to the validity of *model II* results.

6.2.3 PLS-PM in General

Three Lines of Discussion

The discussion about PLS Path Modeling and its usefulness in general is mainly happening in the broader field of management related research, a field PLS-PM has been a widely used method in recent years (cf. Rönkkö, 2014, pp. 168–169). The whole discussion is focused on the central question of whether PLS-PM is an equally suitable method for confirmatory research as CB-SEM²⁹. I give a short overview of the current lines of discussion which are particularly important because they lead to important changes in PLS-PM model assessment.

DIVERGING VARIANCE AND COVARIANCE-BASED SEM Although it has been a topic of discussion for more than thirty years (cf. Sarstedt, Ringle, & Hair, 2014), the discussion about differences and commonalities between PLS-PM and CB-SEM is particularly fostered by Rigdon (2012). From his point of view, the abilities of PLS-PM as a composite method must be stressed. “PLS path modeling gains nothing from its ‘factor-like’ pretensions, but has much to gain from leaving this legacy thinking behind” (Rigdon, 2012, p. 353). This is particularly true with respect to the development of measure validation procedures, which are independent from the factor model at all. Beyond this general claim, Rigdon (2012) provides an important clarification concerning constructs as theoretical concepts and their role in PLS-PM: Any conceptual definition is always applied to a focal concept. This focal concept, in turn, cannot be used in a SEM. Instead, a proxy for the focal concept is used as a stand-in. The MVs used are connected to the proxy via mathematical operations. A validity gap between the focal concept and its proxy in the model cannot be avoided—regardless whether formative or reflex-

²⁹ In fact, this question has been discussed since the days of H. O. A. Wold (1982) but has now become virulent, again.

ive measurement is applied.

A whole issue of the journal *Long Range Planning* (vol. 47 (3), 2014) was dedicated to comments and second thoughts on the ideas Rigdon presented. Together with methodological developments described below, this discussion leads to a new view on PLS-PM presented by Henseler, Hubona, and Ray (2016) (cf. section 6.2.4).

OPPONENTS OF PLS-PM The most fundamental critique as well as thorough empirical analysis of the strength and weaknesses of PLS-PM comes from Mikko Rönkkö and his colleagues. In a series of journal articles, they aim to make several points thereby trying to abandon PLS-PM as a whole: Rönkkö and Evermann (2013) deal with several aspects of PLS-PM, which have long been criticized. They propose that it is not a structural equation modeling method at all, cannot be used to develop measurement models, does not provide results which are more reliable than simple sum scores, is not suitable to do exploratory research, has no sample size requirements and does not allow for statistical tests of the significance of path coefficients estimated. Rönkkö (2014) goes on with a simulation study investigating the so-called problem of capitalization on chance correlations of error terms³⁰. Thereby, they abandon mixed or purely formative models at all. Finally, Rönkkö, McIntosh, and Antonakis (2015) reinforce the problem of capitalization on chance and stress the lack of formal tests of model fit as major weakness. Therefore, they question the usefulness of PLS-PM in psychological research (and beyond). In particular, they question the current practice of applying sign changes in bootstrapping analyses by means of a simulation study.

PROponents OF PLS-PM The proponents of PLS-PM as a means of estimating SEMs have reacted on several occasions to prior and current criticisms. In a rejoinder to Rönkkö and Evermann (2013), Henseler et al. (2014) reveal that their results are based on the assumption “*that the common factor model is indeed correct [sic]*” (Henseler et al., 2014, p. 2). Based on this notion, they can dispel all of the claims made there against

³⁰ These are “correlations that do not exist in the population, but are non-zero in a sample because of sample variability” (Rönkkö, 2014, p. 166).

PLS-PM. Secondly, the publishing of a textbook on PLS-PM by Hair Jr. et al. (2014) can be viewed as a method to overcome mistakes made in applications of PLS-PM. Thirdly, and more generally speaking, as a result from the ongoing discussion several important methodological inventions and developments for PLS-PM have been developed recently (cf. Henseler et al., 2016, p. 3). These focus particularly on estimation of common factor models.

1. The so called Heterotrait-Monotrait Ratio of Correlations (HTMT) (Henseler, Ringle, & Sarstedt, 2015) was developed to assess discriminant validity of reflective LVs. Essentially, this is an upper boundary on correlations of factors (cf. Henseler et al., 2016, p. 11). It is defined in Henseler et al. (2015, p. 121)

“as the average of the heterotrait-heteromethod correlations (i.e., the correlations of indicators across constructs measuring different phenomena), relative to the average of the monotrait-heteromethod correlations (i.e., the correlations of indicators within the same construct). Since there are two monotrait-heteromethod submatrices, we take the geometric mean of their average correlations”.

The corresponding formula is given in eq. (23), where K_x denotes the indicators associated with construct ξ_x .

$$HTMT_{ij} = \frac{\frac{1}{K_i K_j} \sum_{g=1}^{K_i} \sum_{h=1}^{K_j} r_{ig,jh}}{\left(\frac{2}{K_i(K_i-1)} * \sum_{g=1}^{K_i-1} \sum_{h=g+1}^{K_i} r_{ig,ih} \frac{2}{K_j(K_j-1)} * \sum_{g=1}^{K_j-1} \sum_{h=g+1}^{K_j} r_{jg,jh} \right)^{\frac{1}{2}}} \quad (23)$$

2. For many years, a major argument against the use of PLS-PM brought forward by the proponents of CB-SEM has been the absence of a global goodness of fit criterion for PLS-PM. This criticism has been overcome by Dijkstra and Henseler (2015a), i.e. by the development of PLSc. It is calculated based on the bootstrapped difference between the sample-based correlation matrix and the model implied true correlation matrix. Different methods to cal-

culate the discrepancy of matrices are proposed in Dijkstra and Henseler (2015a, p. 20).

3. ρ_A was developed to find a measure for the reliability of construct scores which is based on construct weights and, therefore, does not make use of the assumption of accurate indicator loadings³¹. Equation (24) gives the formula used to calculate ρ_A , where " \hat{w} is the estimated weight vector of the latent variable [...] and S is the empirical covariance matrix of the latent variable's indicators" (cf. Dijkstra & Henseler, 2015b, p. 300).

$$\rho_A = (\hat{w}'\hat{w})^2 * \frac{\hat{w}'(S - \text{diag}(S))\hat{w}}{\hat{w}'(\hat{w}\hat{w}' - \text{diag}(\hat{w}\hat{w}'))\hat{w}} \quad (24)$$

4. Consistent Partial Least Squares Path Modeling (PLSc) (cf. Dijkstra & Henseler, 2015a, 2015b) can be regarded as the most important out of all recent developments. It counterbalances a key fact of the PLS-PM algorithm which is known as consistency at large (cf. Fornell & Bookstein, 1982). This means, both, the number of observations and the number of indicators must converge to infinity in order to achieve consistent path coefficients and outer loadings in common factor models (cf. section 3.3.1). In all other cases, the algorithm tends to overestimate outer loadings and underestimate path coefficients (cf. Dijkstra & Henseler, 2015b, p. 298). PLSc overcomes both issues. Its calculation is based on ρ_A as defined above and comprises four steps:

- a) Calculate scores for LVs using the traditional algorithm (cf. section 3.3.2).
- b) Use the correlations of inconsistent LVs to calculate ρ_A for each reflectively measured LV. If formative LVs are present, ρ_A is set to one by definition.
- c) "Obtain consistent construct correlations r_{ij} by means of the classical correction for attenuation" (Dijkstra & Henseler,

³¹ This assumption does not hold due to the consistency at large of PLS-PM (cf. section 3.3.1).

2015b, p. 301) based on both, inconsistent correlations and the ρ_A -values:

$$r_{ij} = \frac{r_{ij}^*}{\sqrt{\rho_A(\tilde{\xi}_i)\rho_A(\tilde{\xi}_j)}} \quad (25)$$

- d) Estimate path coefficients using either OLS regression (recursive model case) or routines more adequate for non-recursive models.

Overall, I view PLS-PM to be a suitable method to answer the research questions of this thesis. The arguments provided by the proponents, the fact that major criticisms could be rejoined and methodological inventions were made to overcome these issues are convincing to me. For my own analysis, I draw several conclusions from this discussion.

6.2.4 *Specific Consequences for My Analysis*

In sum, the methodological discussion leads to an updated view on PLS-PM presented in Henseler et al. (2016), who also incorporate updated rules to conduct PLS-PM analyses. It is important to note, that the methodological issues and advancements described are very recent developments. Although my analysis was based on the up-to-date textbook (Hair Jr. et al., 2014), these innovations are not acknowledged there. Nevertheless, based on these developments as well as on suggestions by Henseler et al. (2016) I take several actions and precautions in order to report valid results throughout this thesis.

1. The argument against the application of sign change corrections given by Rönkkö et al. (2015, p. 80) is convincing. Literature on bootstrapping analysis in general does not propose corrections of this kind. Even more importantly, in applying corrections the most basic principle of bootstrapping, i.e. to repeat the same calculations as done for the original sample, is violated. As a consequence, the advice given by Hair Jr. et al. (2014) is outdated with respect to construct level or individual sign changes (cf. Henseler et al., 2016, p. 15). This means, the analyses reported in section 5.4.3 reflect the

evidence provided by Rönkkö et al. (2015) and do not apply sign change corrections for bootstrapping analyses.

2. I have double checked the discriminant validity of the reflective LVs based on the HTMT and will report these results when discussing the measurement models below (cf. section 6.3.1).
3. I have reassessed both models based using PLSc. I discuss these results below, based on ρ_A , AVE and outer loadings for the measurement model parts (cf. section 6.3.1) and path coefficients for the structural model parts (cf. section 6.4.1).
4. I will not assess the overall fit of the models based on PLSc. There are two reasons for this decision. First, the fit index proposed by Dijkstra and Henseler (2015a) was developed for common factor models. As PLSc is a very new approach, there are currently (April 2016) no studies, which investigate the usefulness of the fit index in models with both, formatively and reflectively measured LVs. Second, Dijkstra and Henseler (2015a, p. 20) show that the fit index tends to require very large sample sizes. If a sample as large as $n = 1200$ is used, the desired results are achieved. However, when using 300 simulated observations, the

“nominal rejection probabilities [...] are clearly too small. The test is more cautious than desired. Apparently, when the sample does not fit too well, the bootstrapped transformed sample tends to be worse” (Dijkstra & Henseler, 2015a, p. 20).

As the sample size of this doctoral work is $n = 360$ and I employ a model containing both, composite and common factors, I will not investigate overall model fit based on these considerations.

6.3 THE MEASUREMENT MODELS

This section has two parts. I will discuss the results of reflectively measured LVs first. This includes the reassessment of these parts based on PLSc. Subsequently, a particular focus is given on *visual salience* which

is necessary due to its formative nature. It is important to note, that the corrections for consistency do not apply to formative constructs, i.e. there is no need to reassess *visual salience* using PLS.

6.3.1 The Reflectively Measured Part

PLSc Reassessment

Regarding measurement model results applying PLSc does not yield major differences compared to traditional PLS-PM results.

1. In accordance with non-consistent PLS-PM estimation p_sim must be removed ($\lambda = -0.306$) as well as c_cus ($\lambda = -0.005$). Weak outer loadings are also found for c_per ($\lambda = 0.320$), c_pus ($\lambda = 0.317$) and both are removed. The removal of c_his ($\lambda = 0.663$) does not yield the desirable result in terms of AVE and will be retained, whereas removing c_eas ($\lambda = 0.583$) does. However, the latter will be retained for the reasons given above (cf. section 5.4.2). As the deletion of a_vis ($\lambda = 0.631$) did not increase AVE significantly, the indicator was retained. Overall, the same measurement models for PLSc and PLS-PM result with respect to the indicators used.
2. The ρ_A values (cf. table 34) do not show any contradictions to the assessment based on non-consistent PLS-PM. The threshold of $\rho_A > 0.7$ (cf. Henseler et al., 2016, p. 12) is met for all constructs except *cognitive salience*. However, a threat to reliability is not given because the indicator c_eas was retained based on theoretical considerations, thereby weakening the empirical results regarding *cognitive salience* measurement.

	ADV	COS	PRO	OVSAL	STS
ρ_A	0.913	0.692	0.866	0.923	0.901

Table 34.: The ρ_A values achieved for the reflectively measured constructs.

3. In contrast to Hair Jr. et al. (2014), Henseler et al. (2016) stress that AVE-values are sensible for PLS_c, only. However, comparing the AVE values reported for traditional PLS-PM with those of PLS_c does not reveal major differences with respect to common thresholds. Table 35 shows the values based on the measurement model setup without those indicators having weak outer loadings. The most important difference occurs for *cognitive salience* where the AVE value drops below the commonly used threshold. However, this is not surprising as *c_eas* is retained for theoretical reasons.

Method	OVSAL	PRO	COS	STS	ADV
PLS	0.900	0.900	0.600	0.800	0.800
PLS _c	0.800	0.800	0.400	0.700	0.700

Table 35.: A comparison of AVE values for the reflectively measured LVs.

4. According to Henseler et al. (2015, pp. 121–122) two ways of assessing HTMT exist: comparing the model values with thresholds or performing a bootstrapping analysis to determine whether the values are significantly less than one. Table 36 provides the results for the bootstrapping analysis for the model based on assumptions of independence. All values achieved are significantly lower than one at a significance level of $\alpha = 0.01$. However, despite the significant difference to one, the HTMT-values for *ADV* and *SAL*, for *ADV* and *STS* and for *STS* and *SAL* are large. This suggests that these constructs are interrelated—a fact further examined by means of the mediation analysis reported (cf. section 5.5.2). This result is an empirical hint that participants tend not to view different subdimensions of *salience* as separate from each other.

Applying PLS_c did not yield major differences compared to PLS-PM with respect to reflective measurement models. As a consequence, the discussion of the measurement model results will not make a difference between non-consistent PLS-PM and PLS_c estimations.

	COS	PRO	SAL	STS
ADV	0.547***	0.373***	0.815***	0.881***
COS		0.293***	0.694***	0.566***
PRO			0.394***	0.346***
SAL				0.831***

Table 36.: The bootstrapping results for HTMT-values of reflective constructs. *** indicates $p < 0.001$. A significant result means that the HTMT-value is significantly smaller than one.

A Well Fitting Reflective Measurement Model

Overall, the reflective measurement models proposed for *overall salience*, *structural salience*, and *visibility in advance* show reasonable fit. First of all, the excellent fit of the indicators proposed for LV *overall salience* is an important finding. This means, these indicators can be used to assess an object's *salience* without taking any of the subdimensions into account. Designing LM selection studies based on these simple questions can enhance empirical designs, where up until now object selection is often not justified at all (cf. section 7.2.4). However, a closer look at those cases, where indicators did not yield desirable loadings is advisable. One of the three indicators of *prototypicality* was rendered insignificant. As stated earlier p_{sim} required participants to answer the question "How often do you encounter similar objects?". The insignificant result may stem from differing interpretations by participants. Imagine a building where a shop of a large company is located. Answering this question, one may either refer to the building itself or to buildings hosting shops of this company. While the first interpretation represents the meaning intended, the second is also a possible variation. Despite the insignificant result for this indicator, the two remaining items show highly significant outer loadings. Therefore, the lower bound for the number of indicators in reflective measurement models is still met (cf. Kenny, 1979, pp. 178–179). As a consequence, the effect *prototypicality* has on other constructs can still be analyzed. Nevertheless, *prototypicality* turns out to be a LV which is difficult to measure—a finding in line with Raubal and Winter

(e.g. 2002), who, therefore, neglect *prototypicality* in their model, completely.

With respect to the measurement of *cognitive salience* only two (c_his and c_wor) out of six indicators achieved desirably high outer loadings. These reveal a major importance of historical aspects, which is a clear difference to the empirical result of Schroder et al. (2011). They found historical and cultural significance to be of least importance. Of the four indicators with weak outer loadings, c_eas will be kept. This has, of course, an impact on measurement model fit achieved for *cognitive salience*, which was clearly revealed by PLS_c results. However, I will accept the lowered model fit based on the aforementioned theoretical considerations (cf. section 5.4.2). Some of the difficulties arising in the measurement of *cognitive salience* can be explained by the behavior of the participants. Very weak indicator loadings were found for c_cus and c_pus based on PLS_c assessment. This result was to be expected based on observations during experiment trials (cf. section 5.4.2). Finally, the low outer loading of c_per may be a first empirical hint for missing constructs dealing with familiarity and/or emotions (cf. section 6.4.2).

6.3.2 Interpreting the Results for Visual Salience Measurement

One important result of this doctoral work is the fact, that empirical evidence was found that formative measurement for *visual salience* is adequate. The analysis of the non-hierarchical model for *visual salience* revealed an important interpretation of *size* by participants: Obviously, they interpret *size* to subsume *area*, *length*, *width* and *height* as all of these are highly correlated with *size*. This evidence reinforces the importance of the approach taken in this thesis, i.e. considering prior empirical results simultaneously to build a measurement model. The HOC-LOC-analysis performed further stresses the importance of *size*³². I will base the discussion on *model II* results (cf. section 5.5.2) as the PLS_c results sug-

32 NB: A reassessment of the HOC-LOC-analysis for *visual salience* using PLS_c is not necessary, because the LOCs are all formative predictors of *visual salience*. As a consequence, no differences between PLS-PM and PLS_c results are to be expected for path coefficients of LOCs.

gest that *model I* does not hold (cf. section 6.4.1). *Geometry* as a LOC has by far the most important effect on *visual salience*. This result is somewhat counter-intuitive as one would have expected *Color* to be more important than *Geometry*. Moreover, even *Secondary* is slightly more important than *Color* turns out to be (cf. section 5.5.2). Although surprising on the first sight, these results are in line with earlier, qualitative findings. For example, Schroder et al. (2011) reveal *color* to be only fourth frequently named after *name*, *size*, and *age* which is in line with the results of the HOC-LOC analysis. Based on a small scale lab-based study, Winter et al. (2005, p. 139) report the following weights for façades comparing day- and nighttime:

	Area	Shape	Color	Visibility	Marks
Day	0.11	0.15	0.37	0.26	0.12
Night	0.26	0.0	0.21	0.23	0.30

Table 37.: The importance of different dimensions according to Winter et al. (2005, p. 139). The figures are weights based on a robust regression.

These results suggest a rather different picture as color shows the largest regression weight in the daytime condition. However, the difference to the results of this doctoral work may stem from the fact that an on-line study based on 360° panoramic images was used by Winter et al. (2005). This may be an indication that results cannot be transferred between different experimental modalities: The illumination conditions captured on photos may yield a different impression than an observer has when viewing the object in-situ. Overall, doing a HOC-LOC analysis was an important step. It aggregates causal-formative indicators to LOCs, which allows to judge the importance of future indicator candidates a priori.

6.4 THE STRUCTURAL MODEL RESULTS

6.4.1 PLS_c results

The same setup was applied for PLS_c as for the non-consistent PLS-PM analysis, i.e. $\alpha = 0.01$ and $K = 5000$ bootstrapping resamples were used to assess the significance of structural relationships.

Model I

Table 38 contrasts the results for path coefficients assessed for *model I* which was based on assumptions of independence. While the contribution of *prototypicality* is not significant in both cases, major differences between PLS-PM and PLS_c estimations are found. Although the values estimated are equal or higher than those of PLS-PM estimations neither *visibility in advance* nor *cognitive salience* are rendered significant in case of PLS_c.

LV	PLS path coeff.	PLS _c path coeff.
VIS	0.312***	0.243***
COS	0.119***	0.202
PRO	0.049	0.039
STS	0.312***	0.356**
ADV	0.224***	0.201

Table 38.: Each row shows the path coefficient between the current LV and *overall salience*. *** indicates $p < 0.001$, ** indicates $p < 0.01$ ($K = 5000$ resamples).

From my perspective this is a clear hint that the independence assumptions made in the model do not hold according to PLS_c results. As all subdimensions of *overall salience* are derived from the literature, it is not reasonable that only two significantly contribute to *overall salience*. This means, the empirical data does not suggest independent subdimensions according to PLS_c results. As a consequence, I will not further discuss the results of *model I*.

Model II

In case of the model based on hypotheses of mediation, the reassessment is particularly important, as Dijkstra and Henseler (cf. 2015a, p. 298) point to the fact that there is evidence for overestimated path coefficients for this type of model. Table 39 contrasts the results for the hypotheses examined (cf. section 5.5.2).

	Relationship	PLS-PM	PLSc
H_1	$VIS \xrightarrow{?+} OVSAL$	holds	holds
H_2	$COS \xrightarrow{?+} OVSAL$	holds	holds
H_3	$STS \xrightarrow{?+} OVSAL$	holds	holds
H_4	$ADV \xrightarrow{?+} OVSAL$	holds	does not hold
H_5	$PRO \xrightarrow{?+} OVSAL$	does not hold	does not hold
H_6	$VIS \xrightarrow{?+} ADV$	holds	holds
H_7	$ADV \xrightarrow{?+} STS$	holds	holds
H_8	$PRO \xrightarrow{?+} COS$	holds	holds
H_9	$COS \xrightarrow{?+} STS$	holds ($\alpha = 0.05$)	holds ($\alpha = 0.05$)
H_{10}	$PRO \xrightarrow{? \text{ mediated by } VIS} OVSAL$	holds	holds

Table 39.: A comparison of PLS-PM and PLSc results for the hypotheses *model II* was built on (cf. section 5.5.2). ?+ means a positive relationship is investigated.

The interpretation regarding the hypotheses is based on the numerical results presented in table 40, table 41 and table 42. In general, the PLSc results are in-line with the interpretations given before. The three-path-mediated effect of *visual salience* on *overall salience* via *visibility in advance* and *structural salience* is highlighted by the PLSc assessment. There are two reasons for this interpretation. First H_4 is no longer supported, i.e. *visibility in advance* does not show a significant direct effect on *overall salience*. Furthermore, the amount of variance explained by the specific indirect effect is enlarged. It accounts for approx. 40% of the variance in case of PLSc estimation as compared to approx. 25% in the non-consistent case. The specific indirect effect of *prototypicality* on *overall*

salience as mediated by *visual salience* is significant but marginal. Similarly, although the assessment with PLS_c yields a significant indirect effect of *prototypicality* via *cognitive salience* and *structural salience*, its size is still very small and practically irrelevant.

Effect	PLS-PM path coeff.	PLSc path coeff.
ADV → STS	0.710***	0.841***
VIS → ADV	0.661***	0.703***
STS → OVSAL	0.315***	0.371**
PRO → VIS	0.300***	0.007 n.s.
VIS → OVSAL	0.260***	0.199***
ADV → OVSAL	0.244***	0.213n.s.
PRO → COS	0.243***	0.317***
COS → OVSAL	0.144***	0.211**
VIS → STS	0.091*	-0.022n.s.
COS → STS	0.083*	0.123*
PRO → OVSAL	0.055n.s.	0.049n.s.
PRO → STS	0.021n.s.	0.004n.s.

Table 40.: Comparing PLS_c and PLS-PM results for *model II*. *** indicates $p < 0.001$, ** indicates $p < 0.01$ and * means $p < 0.05$ ($K = 5000$ resamples).

Method		Total effect	Direct effect	Total indirect effect
PLS	VIS on OVSAL	0.598***	0.260***	0.338***
PLSc	VIS on OVSAL	0.560***	0.199***	0.361***

Table 41.: Bootstrapping results concerning significance of total, direct and total indirect effects for *visual salience*. *** indicates $p < 0.001$ ($K = 5000$ resamples).

6.4.2 Discussing the Results

The Importance of Visual Salience—and Its Mediation

This doctoral work found a large direct effect of *visual salience* on *overall salience*. On the one hand, this finding is in line with other studies in the broader field of research on *salience*. For example, Davies and Peebles

Effect	PLS spec. indir.	PLSc spec. indir.
$VIS \rightarrow ADV * ADV \rightarrow STS * STS \rightarrow OVSAL$	0.148**	0.219**
$PRO \rightarrow COS * COS \rightarrow STS * STS \rightarrow OVSAL$	0.006n.s.	0.014**
$PRO \rightarrow VIS * VIS \rightarrow OVSAL$	0.078**	0.001**

Table 42.: PLS-PM and PLSc assessment of the proposed specific indirect effects. ** denotes $p < 0.01$ ($K = 5000$ resamples) based on a simple percentile CI.

(2010) study the importance of *visual salience* for the strategies used to orient oneself in a real-world spatial environment using different kinds of maps. They provide evidence for the high distractive impact visually salient objects have on the orientation of map viewers: “[V]isible salient landmarks bias people away from using optimal geometry-matching strategies” (Davies & Peebles, 2010, p. 135). However, there is research in the broader field of psychology which challenges the idea of visual cues to be most important in all cases. Hamburger and Röser (2014) analyze the influence of landmark modality (visual, acoustic, etc.), familiarity and famousness of buildings on recognition time and wayfinding performance. The findings suggest that “[o]ther modalities besides vision may [...] be suitable for generating a spatial image which allows for successful wayfinding independent of any visual cues” (Hamburger & Röser, 2014, p. 212). This is in line with the important indirect effect found for *visual salience*. Parts of the total effect *visual salience* has on *overall salience* was found to be mediated by the three-path-mediation $VIS \rightarrow ADV \rightarrow STS \rightarrow OVSAL$. The path coefficient connecting *visual salience* and *visibility in advance* is the second largest of all coefficients and highly significant. This is also in line with the findings by Miller and Carlson (2011) who report on the strong influence *visual salience* has on object recognition based on a route learning and description study in a VR environment.

The highest path coefficient was found for the impact of *visibility in advance* on *structural salience*. Further examining the predictive capabilities of the model with respect to *structural salience*, *visibility in advance* shows a very large effect size for both relevance and accuracy (cf. table 43). Beyond *visibility in advance*, only *cognitive salience* has an effect on *structural salience* which is small in size. *Structural salience* itself shows the largest

	$R^2(STS)$	$Q^2(STS)$	f^2	q^2
complete	0.821	0.501	NA	NA
w/a VIS	0.823	0.501	-0.011	0.000
w/a ADV	0.505	0.313	1.765	0.377
w/a PRO	0.821	0.502	0.000	-0.002
w/a COS	0.809	0.497	0.067	0.008

Table 43.: Predictive accuracy and predictive relevance of LVs with a structural relationship to *structural salience* based on PLS estimations. The figures stress the major impact *visibility in advance* has.

f^2 effect on *overall salience*. This is, again, similar to findings in the field of psychology, where LMs located at intersections and their resulting *structural salience* have drawn particular interest in recent years. In his dissertation Röser (2015) analyzes the *salience* of objects located at four-way intersections. He finds that participants prefer those “landmarks that were located in the direction of turn” (cf. Röser, 2015, p. 146). In GIScience research, *structural salience* was recently examined by Quesnot and Roche (2015). They analyze the influence familiarity and gender have on the preference of LMs by means of an online study. Participants were asked to choose their favorite LM position at cross intersections. The findings suggest that *structural salience* affects females more than males.

In general, the experiment reported in this doctoral work provides empirical evidence that the subdimensions of *overall salience* are not equally important. From the beginning of *salience* theory, weights for the different subdimensions have been incorporated (cf. Raubal & Winter, 2002). While this doctoral work derives weights for each of the subdimensions of *salience*, earlier empirical studies trying to estimate weights are rarely found or do not take all subdimensions simultaneously into account. This shortcoming is overcome by the current study by means of SEM based on data gathered in-situ (as compared to online studies (cf. Winter et al., 2005) or those conducted in VR-environments (cf. Röser, 2015)). As mentioned earlier, Winter et al. (2005) derive robust regression weights for particular variables, such as color, shape area. They find different degrees of importance at daytime as compared to nighttime. The most

recent example to calculate weights for a single subdimension is the study by Röser (2015). Based on his experimental VR-results, he devises the following formula eq. (26), where $w_v + w_u = 1$ and all quantities are scaled to the interval $[0; 1]$:

$$s_t = w_u(d \times o \times v_{vis} \times s_u) + (w_v s_v) \quad (26)$$

According to eq. (26) the *salience* of an object equals the sum of the weighted *visual salience* (s_v) and the product of the distance (d) an observer has to the object times its *structural salience* (s_u) times its orientation towards the observer (o) times the visible part of the LM from the current position of the observer (v_{vis}). Thus, Röser (2015) extends the model presented in Raubal and Winter (2002) by introducing *viewpoint-based salience*. The weighting factors Röser (2015, p. 125) derived emphasize the importance of *structural salience* for the *salience* of a LM—a finding in line with the results of this doctoral work particularly when predictive capabilities are taken into account.

Predictive Capabilities of Models

While the Q^2 values remain unchanged, the R^2 values are enlarged when PLS_c assessment is used (cf. table 44)³³.

Mode	OVSAL	ADV	COS	STS
PLS	0.709	0.437	0.059	0.680
PLS _c	0.813	0.506	0.136	0.821

Table 44.: A comparison of R^2 values achieved with PLS-PM and PLS_c estimations.

Due to the method's novelty, sources providing insights on how to interpret these differences with respect to R^2 are rarely found. According to Becker (2015) these differences may result from large corrections

³³ There is no need to refer to R^2_{adj} values in comparing the results because the estimations are based on the same model.

	$R^2(OVSAL)$	$Q^2(OVSAL)$	f^2	q^2
complete	0.813	0.619	NA	NA
w/a VIS	0.801	0.587	0.064	0.084
w/a ADV	0.806	0.607	0.037	0.031
w/a STS	0.791	0.593	0.118	0.068
w/a PRO	0.811	0.617	0.011	0.005
w/a COS	0.788	0.610	0.134	0.024

Table 45.: Predictive accuracy and predictive relevance of LVs with a structural relationship to *overall salience* based on PLS_c estimations. The figures reveal the major impact *visual salience* and *structural salience* have.

needed; however, the ρ_A values provide no further indication³⁴. Table 45 gives an overview of the PLS_c based effect sizes achieved for predictive accuracy and relevance the subdimensions have on *overall salience*.

These results reveal two important aspects. First, medium sized f^2 effect sizes, which are still rather low, are shown for all subdimensions of *overall salience* except *prototypicality*. *Prototypicality* is of very small importance to *overall salience*, with respect to predictive capabilities, direct effects and its specific indirect effects. As it is the only LV which has a relationship with *cognitive salience* but $R^2(COS)$ is very low, it does not explain much variance in *cognitive salience*, as well. The slight effect of *prototypicality* may stem from its fundamental measurement model, where only two indicators show desirable indicator loadings. Second, of those effect sizes, which are medium sized, *structural salience* and *cognitive salience* show the largest values. In particular, *cognitive salience* is of importance for *structural salience* (cf. table 43) and *overall salience* (cf. table 45) alike. On the other hand, although the Q^2 values are large, the effect sizes of predictive relevance are rather low, either. This means that the subdimensions have a small but important effect. Combined with the results of the analysis of mediation, this result fosters the notion of incomplete theoretical conceptions of *salience*. Important subdimensions other than those proposed in common theories may be missing. One candidate dimension is *emotional salience*, which has recently gained im-

³⁴ As noted earlier, a ρ_A value below thresholds is acceptable for *cognitive salience* due to the fact that *c_eas* was retained purposefully.

portance particularly in psychological research. Niu et al. (2012) present evidence for the importance of *emotional salience* at a very early stage in the perception of pictures. Indeed, they show that “[e]motional salience can override visual salience and can determine attention allocation in complex scenes. These findings are consistent with the hypothesis that cognitive/affective factors play a dominant role in active gaze control” (Niu et al., 2012, p. 13:1). By means of a lab-based VR study Balaban, Röser, and Hamburger (2014) find evidence that positive and negative emotions have an influence on recollection and wayfinding. However, the direction of influence remains undecided. Wayfinding performance is enhanced by those LMs with which negative emotions are associated, whereas positive emotions foster route learning. Another dimension worth investigating is *familiarity*. Quesnot and Roche (2015) reveal *visual salience*, *structural salience* and *semantic salience* to have an impact on all participants, but those who are familiar with the study area prefer objects which have a meaning for them. Consequently, *familiarity* may be an important aspect, which is missing in the models of this thesis, too. This facet, however, may be hard to distinguish from *emotional salience* or may at least have an impact on it. Imagine the object to be rated is a person’s school house. This object is certainly familiar to her/him, but it is also likely to evoke emotional affect due to this familiarity. Further analysis of the dimensions of *emotional salience*, however, is necessary to substantiate this claim. In particular, future experiments (cf. section 7.3.3) should focus on deriving questions useful for measuring *emotional salience* in real-world navigation contexts—and test whether it can be empirically distinguished from *familiarity* (cf. section 7.3.3).

6.5 GENERAL LIMITATIONS

Some general limitations apply to the work presented. However, these may lay the foundation of future work.

AN OPERATIONAL DEFINITION OF SALIENCE

Chapter 2 is dedicated to *salience*, which is the central concept of this doctoral work. The studies reviewed suggest specific measurements to capture *salience*, i.e. they are operational. I have tried to generalize these insights, aiming to find the common theoret-

ical ground upon which the various disciplines base their operationalization of *salience*. As the focus of the whole doctoral work is on measuring the LVs involved, an operational definition was given based on these insights (cf. section 2.4)—without finding a theoretical definition before. This may be regarded a weakness. However, this thesis can contribute to the future development of a theoretical definition through the analysis of a measurement model and two different structural models. This includes an examination of the way different sub-constructs of *salience* influence each other. Given its aims, drawing on an operational definition is not a weakness in terms of theoretical grounding of the thesis, but a strength.

OUTDOOR ONLY

The SEMs proposed and analyzed throughout this doctoral work are based on models for outdoor environments and the experiment was conducted outdoors, consequently. It remains an open research question (cf. section 7.3.4), whether the weights learned for models throughout this doctoral work will hold for indoor environments as well.

HISTORICAL IMPORTANCE OF REGENSBURG

Clearly, being part of the UNESCO Cultural World Heritage the Old Town of *Regensburg* is an environment which has distinctive features due to its historical importance and architectural style. This may have had an impact on the weights assessed. However, the structural and measurement models proposed are applicable to other urban environments (e.g. suburbs), too. In order to counterbalance a potential bias of historical importance two important provisions were made (cf. section 4.2.2):

1. The experimental area was extended beyond the UNESCO Cultural World Heritage District (cf. UNESCO World Heritage Convention (2006) and fig. 13).
2. Objects were randomly selected in the experimental area, i.e. participants were not allowed to choose objects themselves as well as the objects were not chosen based on expert judgment.

Future work (cf. section 7.3.4) will reveal, whether classes of urban environments exist showing weights similar to those found for *Regensburg*.

IMPORTANCE OF SEASONS

According to Sarjakoski et al. (2013), seasons are important to judgments of *saliency*. In order to control this confounding variable all trials were conducted during late autumn and winter time. Furthermore, all experiments were conducted during daylight, which is important to note because there is evidence that nighttime has an impact on *visual saliency* (cf. Winter et al., 2005). It is, still, impossible to control all of the side effects differing daylight conditions possibly have on ratings. However, evidence was found that ratings are rather invariant: As reported in section 5.2.4 even considerable snow cover did not yield a more severe impact on data than other weather conditions had. This is also a strong finding against the background that real-world users will contribute LM_s (cf. section 7.2.3) at different times of the day, and in diverse daylight and seasonal conditions.

RESTRICTIONS REGARDING PARTICIPANTS

ACCIDENTAL SAMPLING OF PARTICIPANTS

I chose the objects used in this study (cf. section 4.2.2) randomly. This determined two out of three factors *saliency* results from (cf. section 2.4): the object itself and the environment within which it is embedded in. However, drawing a random sample out of all potential observers is impossible to achieve. The accidental sampling of participants may have had an impact on the estimated weights. This may be all the more true, as 80% of all participants were university students, i.e. in their early twenties and well-educated persons. While this seems to be a trivial insight, it is made to stress the fact that a non-convenient sample of participants cannot be achieved given the research question this thesis aims to answer. This problem may be counterbalanced by the fact that the weights of the model can be re-estimated once

more users with different (educational) backgrounds have contributed to the database of *salience* values for objects.

20/20 VISION AND NO OTHER SPECIAL NEEDS

All participants of this study were required to have 20/20 vision or vision corrected to 20/20 vision. Furthermore, participants were restricted to those not suffering from color-blindness and not having any physical handicaps or mental disabilities. Persons with disabilities form particular groups of users (cf. e.g. Karimi, 2011, pp. 115–118). This means, the needs of the different groups with handicaps may be incompatible to each other. They have special needs which are likely to result in particularly different weights or even dimensions which may not be suitable for persons without handicaps. Hence, the decision to exclude disabled persons was made in order to control for a large amount of variability which would have been introduced by their different needs.

HOMOGENOUS SoD

As noted earlier (cf. section 5.2.1), participants were a homogenous group with respect to SoD ($\bar{x} = 4.10$, $\tilde{x} = 4.16$, $SD = 1.00$ on a 7-point Likert-scale). This homogeneity is likely to be a consequence of accidental sampling. It is reasonable that persons who know about their weaknesses in terms of SoD are not likely to participate in a navigation experiment. On a closer look, however, the observed value indicates an average group of persons in terms of SoD, particularly because the value is similar to those achieved by Montello and Xiao (2011). They use the Santa Barbara Sense of Direction Scale (cf. Hegarty et al., 2002) to assess the universality of SoD as a concept. Based on a large variety of participants from different countries, the values achieved on a 7-point Likert scale are very similar (cf. Montello & Xiao, 2011, p. 271) to those reported in this doctoral work. This suggests that the sample of participants in this thesis is not biased in terms of SoD.

Overall, these limitations are not regarded to be a major threat to the validity of the results of this doctoral work. Indeed, adequate precautions were taken to counterbalance potentially harmful effects.

6.6 CHAPTER SUMMARY

The aims of this chapter were threefold. First, major statistical advancements were reviewed and their consequences on the results were discussed. Second, the empirical results were considered with respect to findings of prior studies. Third, the limitations of the study were detailed. The main findings of this chapter include:

RECENT METHODOLOGICAL ADVANCEMENTS as well as their potential impact on the analysis presented in this study were described. Important clarifications and discussions about formative measurement in general and its impact on traditional PLS-PM estimations were outlined. Neither of those is a threat to the validity of the results reported. General criticisms of PLS-PM as a method lead to the important innovation known as Consistent Partial Least Squares Path Modeling (PLSc), among others. Several precautions, e.g. calculating ρ_A and reassessing the models based on PLSc, were taken, in order to ensure the validity of the achieved results.

CONSISTENT PLS PATH MODELING results of measurement and structural models were calculated and discussed. In general, the results are very similar to those which are based on PLS-PM. However, one important difference was found. The corrections applied by PLSc reveal weak results for *model I*, which is based on assumptions of independence between subdimensions of *overall salience*. Only *structural salience* and *visual salience* had a significant impact on *overall salience*, which is implausible based on earlier empirical and theoretical results. As a consequence, *model I* results were not further discussed.

DIFFERENT IMPORTANCE of different subdimensions of *overall salience* was discussed and found to be in line with previous findings. However, this doctoral work is the first large-scale empirical study which analyzes the difference in importance.

THE HIGH IMPACT OF VISUAL DIMENSIONS found was related to findings of other empirical studies. The results of the LOC-HOC-study are generally in line with the findings of previous work. *Color* is not most important to *visual salience*, but *geometry* and *secondary factors*, e.g. *visual age*, are. Furthermore, *visual salience* has a large direct impact on *overall salience*. However, aligning with the findings in other studies, *structural salience* was also discovered to be of major importance to *overall salience*. In turn, *structural salience* is heavily influenced by *visibility in advance*. *Cognitive salience* is also important for this subdimension.

EMOTIONAL SALIENCE and *familiarity* were identified as possibly missing subdimensions contributing to *overall salience*. This finding is based on the small effect sizes revealed.

STRUCTURAL SALIENCE was very important for the prediction of *overall salience*. This finding is in-line with current studies in GIScience as well as psychological research.

GENERAL LIMITATIONS apply with respect to participants and the historical importance of the in-situ environment of experiments. This study focuses, furthermore, on non-disabled participants and is based on an accidental sample as achieving a random sample of all persons potentially using a PNS is impossible. Perhaps because participants with poor SoD are less willing to participate in navigation experiments, a homogenous group with respect to spatial abilities was tested. However, other studies show that the average SoD of the participants is not uncommon. The impact of the environment was minimized by randomly sampling objects and through the extension of the study area beyond the Cultural World Heritage District. The proposed models are, however, generally applicable to other environments.

*Causarum enim cognitio cognitionem
eventorum facit.*

For the knowledge of causes produces a knowledge of effects. —
MARCUS TULLIUS CICERO, *Topica*, 67

7

CONCLUSIONS AND FUTURE WORK

7.1 INTRODUCTION

This doctoral work is dedicated to the measurement of *salience* of objects which can be included in route instructions. This question has been answered empirically. I started with a thorough review of terminology and existing literature which reports experiments on *salience* estimation (cf. Chapter 2). I tried to overcome major weaknesses of these earlier empirical attempts. In particular, I calculated an appropriate sample size of objects a priori and randomly selected these based on geographical coordinates (cf. Chapter 4). This means, I ensured to include both, buildings and non-buildings in the sample. Using SEMs as principal method of statistical modeling and PLS-PM as the method of statistical analysis the dimensions found earlier were assessed simultaneously (cf. Chapter 3). By taking existing evidence systematically into account, I am able to contribute to both, theories on *salience* and its measurement (cf. Chapter 6). This final chapter of my thesis has two goals. First of all, it summarizes major achievements with respect to both, theoretical and empirical insights. Secondly, an outline of several aspects for future research work is given.

7.2 CONTRIBUTIONS

7.2.1 *With Respect to Landmark Identification*

In this doctoral work, I developed measurement models for *overall salience* and its subdimensions based on state-of-the-art empirical studies and theories. Overall the results achieved contribute from three different perspectives to the estimation of *salience*.

1. Measurement models with very good fit were proposed for *structural salience*, *visibility in advance* and *overall salience* (cf. section 5.4.2). This is particularly important, as the measurement model for *overall salience* enables future users of my results to reliably measure this construct without taking subdimensions into account.
2. With respect to the LOCs of *visual salience* geometrical features and secondary aspects (e.g. (visual) age) are more important than color is (cf. section 5.5.1). This is of major importance for selecting objects to explain routes to users: The causal-formative indicators of the LOC *Geometry* can often be derived from existing data sources.
3. Although one of three indicators had to be dismissed for *prototypicality* (cf. section 5.4.2) the remaining indicators, *p_exa* and *p_img*, are a reliable way to measure *prototypicality*. *Cognitive salience* turned out to be difficult to measure: half of the indicators proposed for *cognitive salience* were inappropriate. However, the three remaining indicators suffice to provide adequate measurement according to common rules (cf. section 3.3.3).

Overall, the well fitting measurement models provide a sound basis for further analyses of the structural relationships between constructs.

7.2.2 With Respect to Theory about Salience

In this thesis I assessed two different structural models of *salience*. These models describe how the different subdimensions of *overall salience*, i.e. *visual salience*, *visibility in advance*, *structural salience*, *prototypicality*, and *cognitive salience* are connected. Three major conclusions can be drawn from the analysis.

1. Analyzing *model II* reveals a large impact of *visual salience* and *visibility in advance* on *structural salience* (cf. section 6.4.2). Empirical evidence suggests that visual dimensions have a direct impact on *overall salience* in addition to the way they strongly foster *visibility in advance*. In turn, *visibility in advance* is most important for ease of inclusion in route instructions. *Structural salience* is also positively, but less significantly influenced by *cognitive salience*, which also shows a significant direct effect on *overall salience*.

2. In combination with the results of the analysis of mediational effects, the low effect sizes for predictive relevance and accuracy achieved for *model II* (cf. section 5.5.2) suggest that one or more subdimensions are missing in the model. *Emotional salience*, which has recently gained momentum in cognitive psychology research (cf. section 6.4.2), may be a promising candidate for inclusion. However, this subdimension has the major disadvantage that it is different for all users, i.e. it makes personalization inevitable as emotions are elusive across users.

3. The results of *model I* based on PLS_c (cf. section 6.4.1) show that the assumption of independence does not hold for the subdimensions of *salience*. The poor model fit suggests that participants tend to view subdimensions as influencing each other. This means, the way *salience* is estimated by humans is not comparable to a stream of independent sensors. In contrast, the subdimensions of *salience* are evaluated in a holistic manner, a fact *model II* takes into account.

Based on the contributions to measurement of *salience* and the structural model fit revealed, this doctoral work can contribute to survey-based *salience* estimations by end users.

7.2.3 With Respect to Collaborative Landmark Mining

This doctoral work adds to prior discussions about LMs and *salience* models from two different perspectives. On the one hand, it fulfills the need to empirically establish weights for different subdimensions of *salience* (cf. Sadeghian & Kantardzic, 2008). Of course, there are other studies which try to get an empirical insight into those weights (cf. section 6.4.2). However, these studies are either located in VR environments or draw on very small sample sizes etc. In sharp contrast to these studies, this thesis uses a large-scale in-situ experiment to draw conclusions (cf. chapter 4). On the other hand, the need to establish crowd-sourcing based data sources for geographical data in general and data on LMs in particular has been claimed for some years now (cf. Raubal, Mark, &

Frank, 2013; Richter & Winter, 2011a, 2014; Winter et al., 2010). Hirtle and Raubal (2013, p. 145) define this type of information:

“Volunteered Geographic Information (VGI) describes a large number of related activities in which collections of individuals provide geographic information for common consumption, in contrast to relying on the traditional authorities alone to provide maps and spatial information (Goodchild, 2007). VGI allows individuals to mark information about locations that are of particular interest using either standalone applications or generic platforms”.

Survey ratings of objects are an example of VGI in this sense. I agree with Richter and Winter (2014, pp. 168–169) who stress that, although many challenges remain, user-involvement is the only viable way to get “sufficiently detailed and up-to-date data” about LMs, which is large enough to be used as a basis in real-world navigation scenarios. In my opinion a collaborative in-situ assessment of *salience* based on survey ratings, for which the term Collaborative Landmark Mining (CLM) was coined (cf. Ludwig et al., 2013), can be based on the measurement model results and the structural relationships found in *model II*. I suggest to use five questions as a starting point for CLM, thereby excluding *prototypicality* because of its very small impact.

1. *Overall salience: To what extent does this object draw your attention?*
2. *Visibility in advance: To what extent can one easily refer to this object from afar?*
3. *Structural salience: How easy is it for you to refer to this object in a route description?*
4. *Cognitive salience: To what extent do you regard this object to be worthy of preservation?*

5. *Visual salience: Given its local surroundings, to what extent does this object draw your attention because of its:*
- a) *size*
 - b) *shape*
 - c) *color*
 - d) *location*
 - e) *(visual) age*

The enhancement of applications with these questions would provide a solid measurement for *salience*. For example, this survey could be included in a recently presented system: Drawing on earlier work (cf. Richter & Winter, 2011b) Wolfensberger and Richter (2015) present a mobile application, which can be used to tag OpenStreetMap (OSM) objects as LMs. The application is based on photos, from which users can choose the most suitable LM given the current environment. It would be easy to ask users to provide reasons for their choice using the survey questions. Beyond gaining *salience* data, the rating behavior can be exploited for personalization purposes. For example, although color is across users less important than *Geometry* is, some users may prefer color. This preference could, e.g., be derived from the behavior in answering the survey questions regarding color as compared to e.g. size.

7.2.4 *Regarding Experimental Design in GIScience*

One of the goals of this doctoral work was to provide a critical overview of prior empirical work. Based on this overview (cf. section 3.2.4) the methodological setup derived is different to other studies in GIScience. This holds true with respect to three major empirical design decisions. First, in contrast to other in-situ studies (cf. Ishikawa & Nakamura, 2012; Schroder et al., 2011) the sample of objects was drawn randomly (cf. section 4.2.2) based on geographical coordinates. Using randomly selected objects has the advantage of including objects in the sample which are salient and those which are not. This is important for statistical reasons as variance induced by the objects rated will enhance the stability of the

estimations. Beyond that, the object selection method employed has the advantage of being independent of expert judgment about the *salience* of objects. Up until now many empirical studies on *salience*—regardless of whether they are conducted in-situ or not—lack a proper justification of the way objects and routes were chosen.

A second contribution to experimental design is the measurement model provided for *overall salience*. It shows a very good fit and can enhance future experimental designs, where randomly selecting objects is not viable or possible. The three questions used to measure *overall salience* are sufficient to provide empirical justification as to why an object is regarded to be salient in experimental designs.

Third, using a survey as a systematic data acquisition method is not common in in-situ experiments on LMs, today. While surveys may be generally prone to misunderstandings, this disadvantage is overcome when surveys are used in in-situ settings where the experimenter guides participants along a route. In this case clarifications can be given, if needed, and the convenient method of electronically acquired data in standardized form can be used at the same time.

7.3 FUTURE WORK

These findings pave the way for considerable future work. From a methodological standpoint the use of Bayesian approaches looks promising; from a participant perspective, the impact of different subgroups is interesting. Finally, from an object-based perspective applying the model to datasets gained in different environments appears reasonable.

7.3.1 *Find New Structural Models Using Bayesian Networks*

Two causal models were tested for their predictive capabilities throughout this thesis and both are reasonable from a theoretical perspective. However, other approaches of analysis may prove to be useful to foster the development of theory about *salience*. In particular, combinations with non-frequentist approaches will likely provide further insights. Bayesian Networks may be a particularly useful vehicle to push boundaries in theory on *salience*. For example, they may be used to develop

causality directions based on empirical data (cf. Wu, 2010). Going one step further, it would be interesting to see the results, if neither a measurement model nor a structural model is provided as input as is suggested by Zheng and Pavlou (2010). In a first step, they use Bayesian Networks to find the most plausible set of LVs based on the MVs. In a second step they compare all structural relationships possible to connect the found LVs. This data driven approach may reveal new or strengthen existing approaches to measurement and modeling of *salience*.

7.3.2 Heterogeneity Analysis

The data about SOs provided through the survey can be used to provide a dataset to base route instructions on LMs. However, any PNS which aims to provide personalized route instructions to its users must at least acknowledge the needs of different user groups if not individual users. This is important as wayfinding is known to be influenced by individual abilities and the spatial environment, both of which are interrelated (cf. Farr et al., 2012, p. 5). Different needs of groups of users may be reflected by model weights. This means, one particular object may offer distinct utility for particular groups of people. As a consequence, the analysis of heterogeneity of weights for different user groups is promising. Three main criteria come to mind:

GENDER Although gender-based differences with respect to SoD are well studied (cf. Colluccia and Louse (2004) for a thorough overview and cf. Farr et al. (2012, pp. 22–23) for a condensed view on current research about gender differences in spatial cognition), a potential difference in gender regarding the *salience* of LMs has not been examined until very recently. Quesnot and Roche (2015) find a difference between women and men, i.e. their results indicate that the importance of *structural salience* is a function of gender. Hence, gender-based differences regarding inner weights may occur with respect to the models presented.

AGE may yield different weights for biological reasons. Elderly persons are known to have special demands on PNSs (cf. Goodman, Brewster, & Gray, 2005). Among others, visual abilities of el-

derly people are known to change (cf. Lindenberger & Ghisletta, 2009). Consequently, *visual salience* may become less important the older a person is. On the other hand, spatial capabilities are known to differ in different stages of life (cf. e.g. Jansen-Osmann et al. (2007) for differences between adults and children), including the formation of CMs (cf. Iaria et al., 2009). This means, different age groups may not only have different visual impressions of *salience*, but *structural salience* estimations may be affected, too.

KNOWLEDGE There is empirical evidence that the level of knowledge about the local environment has an impact on the selection of LMs with respect to their *semantic salience* (cf. Quesnot & Roche, 2015). However, as my results show that *cognitive salience* has a significant impact on *overall salience* a deeper understanding of the influence different levels of knowledge have is needed.

All of these aspects require very large empirical effort in terms of the number of participants needed. This holds particularly true for age and knowledge, both of which are not dichotomous variables, i.e. various levels of knowledge and age groups must be examined. Nevertheless, the analyses of groups will be a first step towards personalized route instructions based on LMs and, therefore, seems to be well worth the effort.

7.3.3 *Emotional Salience and Familiarity Experiments*

As mentioned above (cf. section 6.4.2), *emotional salience* and *familiarity* may be two important constructs currently neglected in models on *overall salience*. In planning these experiments as in-situ endeavors a major difference to the subdimensions used in this doctoral work must be carefully considered. The degree of emotional affect towards an object is inevitably dependent on personal experience. This means, while e.g. *visual salience* is rather interpersonal, *emotional salience* and *familiarity* are not, i.e. their importance may hardly be generalizable. Nevertheless, both constructs may be of major significance for the next step in incorporating LMs in mobile PNSs: personalization. This gives rise to manifold issues regarding the acquisition of participants as well as

the local environment any experiments are conducted in. Moreover, randomly sampling objects will become virtually impossible because a single chosen object may not evoke affect at all or only for a person who is not part of the sample etc. Major thoughts about the justification of objects selected will be necessary, consequently.

7.3.4 *Use Different Local Environments*

As explained in section 2.3.3 the local environment has a major impact on *salience*. Although precautions were taken to counterbalance potential bias of the rich local environment (cf. section 6.5), further studies are necessary to analyze the predictive capabilities of the model in more uniform environments. Several aspects may be varied in order to draw valuable conclusions. A neighborhood with a uniform architectural style regardless its historical importance may be used. Different visual cultures may play an important role, because of existing evidence that language and culture play an important role in spatial cognition (cf. Farr et al., 2012, p. 23). For example, e.g. non-European cities could be used as a basis to draw conclusions about the influence visual style has on the model's weights. Finally, the usefulness of the model in indoor environments should be investigated. This may lead to further advancements of *salience* model theory, because different LVs may be suggested or other weights may be found.

7.3.5 *The Model in Different Mental Contexts*

Finally, it seems worthwhile to investigate known problems in (human) route communication based on the *salience* assessments using the proposed survey.

1. It would be useful to gain data about the way humans negotiate misunderstandings in route instructions. This would provide an insight into whether the survey results can help to predict these problems. One reason for any of these misunderstandings can be the fact that several objects are equally salient in the local environment or others are far more salient than the object used

7 CONCLUSIONS AND FUTURE WORK

- is. This should be reflected by *saliency* values derived from the survey.
2. Acquisition of qualitative data should be used to validate the important influence *visual saliency* and *visibility in advance* have on *structural saliency*.
3. As noted earlier (cf. section 3.2.2), differences between memorizability and in-situ *saliency* of objects exist. Therefore, it would be interesting to see whether or not *saliency* scores based on *model II* reflect this difference.

7.4 CHAPTER SUMMARY

The last chapter of this thesis had two goals. The results achieved were reviewed and future lines of research were proposed. This doctoral work has provided empirical evidence, which advances theory on *saliency* as well as how it is measured. The way in which results were achieved yields several lines of future research about the *saliency* of LMs. Future work will include the assessment of models for different local and mental contexts and the application of different data analysis techniques such as e.g. Bayesian Networks.

A

APPENDIX

The appendix to this doctoral work provides several additional pieces of information regarding the surveys used and experiment trials.

A.1 SURVEYS

A.1.1 *Assessing Sense of Direction by Using a German Language Scale*

The German language *Fragebogen Räumliche Strategien* (Survey of Spatial Strategies) comprises 19 questions to self-report sense of direction. The questions are given below in German language in the order of sequence presented to users. The items are taken literally from Münzer and Hölscher (2011). All ratings are given on a seven-point Likert scale, where poles were named “strongly reject” (*lehne stark ab*) and “highly agree” (*stimme stark zu*).

1. *Wenn ich durch eine unbekannte Stadt laufe, dann weiß ich, aus welcher Richtung ich gekommen bin und in welche Richtung ich mich bewege.*
2. *Wenn mich jemand in meiner Stadt nach dem Weg fragt, dann stelle ich mir meine Stadt wie auf einer Karte vor und ermittle daraus den Weg.*
3. *Wenn ich mich durch ein großes Gebäude bewege, dann stelle ich mir dabei eine Art Plan oder Grundriss (Überblicksansicht) vor.*
4. *Ich bin sehr gut darin, von meinem gegenwärtigen Standort aus Richtungen zu anderen Orten anzugeben.*
5. *In der freien Natur (z.B. Wald, Gebirge) kann ich mich an einen Weg sehr gut erinnern, wenn ich ihn einmal gegangen bin.*
6. *Ich kann spontan zeigen, wo Norden, Süden, Osten und Westen liegt.*

A APPENDIX

7. *Ich stelle mir die Umgebung stets wie auf einer „mentalen Karte“ (Überblicksansicht) vor.*
8. *Ich finde stets ohne Probleme zu meinem Ziel.*
9. *Ich verfüge über eine sehr gute Vorstellung von meiner Stadt, wie auf einer Karte.*
10. *In einer unbekanntenen Umgebung finde ich mich gut zurecht.*
11. *In der freien Natur versuche ich, die räumlichen Gegebenheiten aus der Vogelperspektive zu verstehen.*
12. *Ich bin sehr gut darin, mir Wege zu merken und finde auch ohne Mühe den Rückweg.*
13. *In einem großen Gebäude habe ich keine Schwierigkeiten, einen Weg nochmals zu gehen, wenn ich den Weg einmal gegangen bin.*
14. *Mein Orientierungssinn ist sehr gut.*
15. *In meiner Stadt kann ich von einem beliebigen Punkt aus spontan angeben, in welchen Richtungen markante Gebäude oder Bezugspunkte liegen.*
16. *Wenn ich in meiner Stadt unterwegs bin, dann kann ich mir meine Position wie einen Punkt auf meiner „mentalen Karte“ vorstellen.*
17. *In der freien Natur kann ich spontan zeigen, wo Norden, Süden, Osten und Westen liegt.*
18. *In einem großen Gebäude weiß ich spontan, in welcher Richtung der Eingang liegt.*
19. *Wenn ich mich in einer unbekanntenen Stadt bewege, dann bilde ich in meiner Vorstellung eine Art „mentale Karte“.*

A.1.2 *Questions Not Used in the Measurement Model*

The *salience* rating survey contained seven questions which were not used in any of the measurement models.

P_CAT (SINGLE-CHOICE INCLUDING OPTION OTHER) Please choose a category this object belongs to: (*Meiner Meinung nach fällt dieses Objekt in folgende Kategorie:*)

The categories participants had to choose from were (given in the order of sequence they were presented):

- building (Gebäude)
- part-of-building (Gebäudeteil, z.B. Laden, Kaffee, Restaurant usw.)
- parking (Parkplatz)
- street-furniture (Straßenausstattung (z.B. Telefonzelle, Parkbucht usw.)
- park (Park)
- waters (Gewässer)
- monument (Monument/Statue)
- square (Platz)
- other (*Sonstiges*)

C_CST (FREETEXT) Please indicate the current use of this object. (*Wie wird dieses Objekt aktuell genutzt?*)

C_PST (FREETEXT) Please indicate the original use of this object. (*Was ist diese ursprüngliche Funktion des Objekts Ihrer Meinung nach?*)

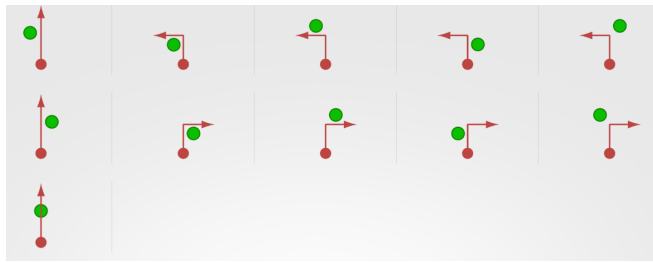
C_HID (YES, NO, NO IDEA) Usually this kind of object has a specific name. (*Objekte dieser Art haben üblicherweise eine feste Bezeichnung.*)

C_KID (YES, NO, NO IDEA) I know the name of this particular object. (*Ich kenne die feste Bezeichnung dieses Objekts.*)

A APPENDIX

C_OID (FREETEXT) Please indicate how you would refer to this object in route instructions. (*Welche Bezeichnung/Benennung würden Sie verwenden, um auf dieses Objekt in einer Routenbeschreibung zu verweisen?*)³⁵

S_PIC (SINGLE-CHOICE) Which of the following pictures is most suitable to your current spatial situation (green dot: current object; red dot: your current position; red arrow: your way on the current route)? (*Welche Darstellung entspricht am besten Ihrer aktuellen Situation (grüner Punkt: Lage des Objekts; roter Punkt: Ihr aktueller Standpunkt; roter Pfeil: Ihre Bewegungsrichtung auf der aktuellen Route)?*) This question must be answered by choosing one of the following pictures, which were adapted from Klippel and Winter (2005, p. 352):



A.1.3 Demographic Data

Section 4.3.3 details reasons for the variables used to measure demographic data. The wording of the questions is given below, where labels refer to those used in section 4.3.3. If no scale is mentioned, the 5-point scale described earlier (cf. section 4.3.1) applies.

SEX (single choice)

Are you male or female? (*Bitte geben Sie Ihr Geschlecht an:*)

YEAR OF BIRTH (freetext)

What is your year of birth? (*In welchem Jahr wurden Sie geboren?*)

³⁵ For unknown technical reasons the application did not save an answer to this question for 20 objects.

PLACE OF BIRTH (freetext)

What is your place of birth? (*In welchem Ort wurden Sie geboren?*)

COUNTRY OF BIRTH (single choice)

What is your country of birth? (*Wo wurden Sie geboren?*)

CURRENT LOCATION THE PERSON LIVES AT (freetext)

Where are you currently living? (*In welchem Ort leben Sie gerade?*)

KNOWLEDGE ABOUT PLACE

How well developed is your knowledge about the Old Town of *Regensburg*? (*Wie gut kennen Sie sich im Altstadtgebiet von Regensburg aus?*)

FREQUENCY OF WALKING IN REGENSBURG

How often do you walk through the Old Town of *Regensburg*? (*Wie häufig sind Sie zu Fuß im Altstadtgebiet von Regensburg unterwegs?*)

FREQUENCY OF USAGE OF MOBILE PNSS

How often do you use pedestrian navigation systems on mobile devices? (*Wie häufig nutzen Sie Fußgängernavigationssysteme auf mobilen Endgeräten?*)

COURSE OF STUDY OR CURRENT JOB POSITION (freetext)

What is your course of study or which profession are you pursuing? (*Welchen Studienabschluss streben Sie an (Fachrichtung/kombination u. Abschlussart) bzw. wie lautet Ihre momentane Berufsbezeichnung?*)

A.2 TRIALS

A.2.1 Trial Statistics Per Day

Table 46 shows the number of trials for each of the 48 days on which experiments were conducted. The average duration for the trials on each day is given additionally. The column indicates the duration of a single trial if only one trial took place on a particular day.

A APPENDIX

date	num trials	mean(dur.)
2014-11-12	1	50
2014-11-13	2	46
2014-11-14	1	47
2014-11-17	1	38
2014-11-18	1	46
2014-11-19	1	58
2014-11-20	2	52
2014-12-03	1	51
2014-12-04	2	61
2014-12-05	3	67
2014-12-08	2	60
2014-12-09	1	64
2014-12-10	4	61
2014-12-11	1	60
2014-12-12	2	52
2014-12-15	4	60
2014-12-16	1	78
2014-12-17	3	62
2014-12-18	3	56
2014-12-19	3	64
2015-01-12	4	56
2015-01-14	2	56
2015-01-15	3	59
2015-01-16	3	69
2015-01-17	3	60
2015-01-19	4	52
2015-01-20	2	43
2015-01-21	3	56
2015-01-22	4	59
2015-01-23	3	57
2015-01-24	1	72
2015-01-26	4	56
2015-01-27	2	58
2015-01-28	3	61
2015-01-29	2	70
2015-01-30	4	68
2015-01-31	1	63

Continued on next page

Table 46 – *Continued from previous page*

date	num trials	mean(dur.)
2015-02-02	3	59
2015-02-03	2	55
2015-02-04	1	86
2015-02-06	3	68
2015-02-09	2	55
2015-02-10	2	61
2015-02-12	1	52
2015-02-14	3	64
2015-02-17	2	83
2015-02-18	3	69
2015-02-19	3	66

Table 46.: The number of trials (**num trials**) conducted per day (format of date: yyyy-mm-dd) and the average duration in minutes (**mean(dur.)**) of the trials per day (days with one trial show the duration of this trial).

A.2.2 Route Statistics

Table 47 shows the starting points of the 55 routes, the number of objects on each route, the length of each route in kilometers and the average duration of experiments on that route.

r	nobj	starting point	len	mean(dur.)
1	7	Dachauplatz (vor Restaurant Scholz)	1.55	60
2	7	Dachauplatz (beim ehemaligen Brunnen)	1.30	59
3	7	Bahnhofsvorplatz (unter der großen Wanduhr)	1.72	58
4	7	Stadtamhof (am Fuß der Steinernen Brücke)	1.94	60
5	6	Café am Peterstor (am Peterstor 1a)	1.26	50
6	7	Dachauplatz (vor Restaurant Scholz)	1.59	58
7	7	Dachauplatz (beim ehemaligen Brunnen)	1.88	67
8	7	vor der Wurstkuchl	1.20	62
9	7	Bismarckplatz (vor der Zufahrt zur Tiefgarage)	1.51	57
10	7	Stadtamhof (am Fuß der Steinernen Brücke)	1.29	52

Continued on next page

Table 47 – *Continued from previous page*

route	nobj	starting point	len	mean(dur.)
11	7	Dachauplatz (beim ehemaligen Brunnen)	1.99	70
12	6	Dachauplatz (vor Restaurant Scholz)	1.64	57
13	6	Bismarckplatz (vor der Zufahrt zur Tiefgarage)	1.18	45
14	6	Fischmarkt (vor den steinernen Fischbänken)	0.94	49
15	7	Bismarckplatz (vor der Zufahrt zur Tiefgarage)	1.55	59
16	7	Restaurant Jalapenos (Schottenstraße 4)	1.34	63
17	6	Arnulfplatz (vor der Arnulf-Apotheke)	0.85	56
18	6	Neupfarrplatz (vor der Sparkasse)	1.36	48
19	7	Ernst-Reuter-Platz (vor dem Europa-Brunnen)	1.67	74
20	7	Restaurant Jalapenos (Schottenstraße 4)	1.73	69
21	7	Ernst-Reuter-Platz (vor dem Europa-Brunnen)	1.03	55
22	7	Café am Peterstor (am Peterstor 1a)	1.28	60
23	7	Bismarckplatz (vor der Zufahrt zur Tiefgarage)	1.07	50
24	6	Altes Rathaus (vor der Touristinformation)	1.45	64
25	8	Bismarckplatz (vor der Zufahrt zur Tiefgarage)	1.61	68
26	7	Dachauplatz (beim ehemaligen Brunnen)	1.75	76
27	6	Neupfarrplatz (vor der Sparkasse)	0.94	40
28	6	Emmeramsplatz (vor dem Schlosseingang)	1.55	51
29	7	Dachauplatz (vor Restaurant Scholz)	1.27	56
30	8	Bahnhofsvorplatz (unter der großen Wanduhr)	1.35	53
31	7	Dachauplatz (beim ehemaligen Brunnen)	1.37	67
32	7	Alter Kornmarkt (an der Litfaßsäule)	1.28	58
33	6	Bahnhofsvorplatz (unter der großen Wanduhr)	1.56	61
34	7	Pustetpassage (neben Bücher Pustet)	1.53	55
35	7	Bahnhofsvorplatz (unter der großen Wanduhr)	1.73	60
36	6	Dachauplatz (beim ehemaligen Brunnen)	1.42	53
37	6	Emmeramsplatz (vor dem Schlosseingang)	1.12	56
38	7	Restaurant Jalapenos (Schottenstraße 4)	2.28	56
39	7	Dachauplatz (vor Restaurant Scholz)	1.11	55
40	7	Dachauplatz (beim ehemaligen Brunnen)	1.67	64
41	7	Arnulfplatz (vor der Arnulf-Apotheke)	1.99	58
42	6	Café am Peterstor (am Peterstor 1a)	0.90	53
43	6	Restaurant Jalapenos (Schottenstraße 4)	1.67	53
44	6	Restaurant Jalapenos (Schottenstraße 4)	1.83	58
45	7	Restaurant Jalapenos (Schottenstraße 4)	2.12	76
46	7	Bahnhofsvorplatz (unter der großen Wanduhr)	2.36	89
47	6	Café am Peterstor (am Peterstor 1a)	1.96	65
48	7	Dachauplatz (beim ehemaligen Brunnen)	1.91	80
49	6	Café am Peterstor (am Peterstor 1a)	1.72	53
50	7	Bismarckplatz (vor der Zufahrt zur Tiefgarage)	1.94	67

Continued on next page

Table 47 – *Continued from previous page*

route	nobj	starting point	len	mean(dur.)
51	6	Ernst-Reuter-Platz (vor dem Europa-Brunnen)	0.74	63
52	7	Ernst-Reuter-Platz (vor dem Europa-Brunnen)	1.71	73
53	7	Pustetpassage (neben Bücher Pustet)	1.23	81
54	7	Dachauplatz (beim ehemaligen Brunnen)	1.32	42
55	7	Fischmarkt (vor den steinernen Fischbänken)	1.09	65

Table 47.: An overview of the routes with respect to different starting points. *nobj* denotes the number of objects to be rated on the route, *starting point* gives the meeting and starting point, *len* denotes the route length in *km*, and the mean duration in minutes is given in column *mean(dur.)*.

A.3 PARTICIPANTS

A.3.1 Explanations Given

Several explanations regarding the topics below were given to participants. As I give these for the sake of completeness, the explanations will not be translated into English:

WELCOME MESSAGE

Schön, dass Sie mitmachen! Ich erkläre am Besten kurz den Ablauf. Ich hätte zwei Fragebögen, die ich Sie bitte, hier vor Ort auszufüllen. Der eine betrifft Sie als Person - warum aber überhaupt persönliche Daten erheben? Weil das meinen Auswertungsspielraum erweitert. Das beginnt bei einfachen Dingen wie der Frage, ob es Unterschiede zwischen Frauen und Männern hinsichtlich der Bewertungen gibt. Und es endet bei komplexeren Fragestellungen, ob es Unterschiede in den Antworten gibt, je nachdem wie gut man sich in Regensburg auskennt. Dann habe ich als zweites einen Fragebogen zu Ihrem Orientierungssinn. Ganz wichtig und das gilt für alle Fragen, die ich Ihnen heute stelle: Es gibt kein richtig und kein falsch; es geht also nicht darum, über Sie die Lanze zu brechen, sondern darum, dass Sie so antworten, wie Sie es für richtig empfinden. Nach diesen Fragebögen gehen wir zusammen eine Route. Diese Route zu gehen wird ca. 50 Minuten dauern. Mehr dazu erkläre ich Ihnen gleich noch [auf Grund häufiger Nachfragen in Christkindlmarkttrouten hinzugefügt: nur zwei Dinge vorneweg: Sie müssen sich die Route nicht merken und ich werde Sie am Ende der Route nicht allein lassen, wenn Sie sich dort nicht auskennen]. Alles klar soweit? [Warten auf Antwort; ggf. Klärungen] Gut, dann melde ich

A APPENDIX

Sie am System an. Möchten Sie einen Eingabestift oder lieber die bloße Hand benutzen? [Person mit vorgegebener email-Adresse angemeldet]

DEMOGRAPHIC DATA SURVEY

[Bei Frage zum Geburtsort: Nachfragen, ob VP auch dort aufgewachsen ist; falls nein durch Komma abgetrennt den Ort nennen lassen, in dem die Person aufwuchs. Bei Frage nach Studienfach gängige Abkürzungen angeboten.] [Sense of direction ausfüllen lassen]

INSTRUCTIONS GIVEN TO PARTICIPANTS BEFORE STARTING TO WALK THE ROUTE

Hier sieht man schon, wie das Ganze ablaufen wird. Wir gehen zusammen eine Route und in der Mitte wird immer ein Bild angezeigt. Auf diesem Bild ist ein Objekt im allerweitesten Sinne rot umrandet. Das Ziel ist also, dass Sie dieses Objekt, während wir auf der Route gehen, erkennen. Wenn Sie das Gefühl haben, dass Sie das Objekt gut in seiner Umgebung sehen und zwar so gut, dass Sie in einem Fragebogen Auskunft zu dem Objekt in seiner Umgebung geben können, dann bleiben wir zusammen stehen, Sie klicken auf Objekt bewerten, beantworten den Fragebogen und, wenn Sie fertig sind, gehen wir zum nächsten Objekt. Ganz wichtig: Dieses Bild dient nur zur Kommunikation, welches Objekt Sie bewerten sollen – um also Missverständnisse zu vermeiden: Wir bleiben dort stehen, von wo aus Sie das Gefühl haben, das Objekt in seiner Umgebung gut genug zu sehen, um Fragen zu diesem Objekt in seiner Umgebung zu beantworten. Diese Position muss keinesfalls mit der Perspektive auf dem Bild übereinstimmen. Das ist übrigens auch ein weiterer Grund, weshalb ich Sie bitte, das Gerät zu tragen: Für mich ist die Information sehr wertvoll, zu wissen von wo aus Sie das Gefühl haben, dass Sie ein Objekt in seiner Umgebung so gut sehen, dass Sie einen Fragebogen zu dem Objekt in seiner Umgebung beantworten können. Deshalb wird beim Klick auf „Objekt bewerten“ auch Ihre aktuelle GPS-Position gespeichert. Da aber GPS im Altstadtgebiet stark fehleranfällig ist, wird dieses Signal durch einen Schrittzähler korrigiert. Und selbstverständlich lasse ich Sie übrigens nie an dem Objekt vorbeigehen. Das Ziel ist aber schon, dass Sie es selbst erkennen. Alles klar? [ggf. Klärungen] Gut, dann gehen wir los.

[Auf dem Weg zum ersten bzw. zwischen zwei Objekten wurde mit VP über verschiedene Dinge geplaudert.]

EXPLANATIONS GIVEN TO PARTICULAR QUESTIONS CONCERNING

CATEGORIZATION OF THE OBJECT

Wenn Ihnen irgendwann keine Kategorie passend erscheinen sollte, dann können Sie beim Klick auf „Sonstiges“ selbst eine Kategorie erfinden.

VISUAL DIMENSIONS

Wenn Sie diese Frage ganz gelesen haben, dann habe ich einen Hinweis dazu. [Warten bis zur Bestätigung] Also: Hier geht es immer um das Objekt in seiner Umgebung und Ihre Aufmerksamkeit wegen jeder dieser Dimensionen. Und zu dieser langen Liste, möchte ich folgendes sagen: Nehmen wir als Beispiel Höhe [nur bei Objekten verwendet, in denen alle Objekte in der Umgebung ähnlich hoch sind; andernfalls auf Breite oder Farbintensität abgestellt]: Angenommen es käme Ihnen ein Objekt in seiner Umgebung besonders niedrig vor, würde aber genau auf Grund dieser geringen Höhe Ihre Aufmerksamkeit auf sich ziehen, dann dürften Sie viele Sterne bei Höhe vergeben – genauso wenn es besonders hoch wäre und deshalb in seiner Umgebung auffiele. Und dies gilt entsprechend für alle Dimensionen.

STRUCTURAL SALIENCE BASED ON THE PICTORIAL REPRESENTATION

[NB: Für das letzte Objekt wurde, wenn es mehrere Alternativen für den weiteren Verlauf gab, ausgelost, welche Richtung Versuchsteilnehmenden mitgeteilt wird]: Dazu muss ich etwas zum Routenverlauf sagen: Das was Sie gerade bewerten ist der grüne Punkt. Der rote Pfeil beschreibt den Verlauf der Route – und welches der Bilder jetzt am besten passt, ist Ihre Entscheidung.

CURRENT AND PRIOR USE OF THE OBJECT RATED

Current use: Wenn Sie diese nicht erkennen können, geben Sie bitte an, wie es auf Sie wirkt.

Prior use: Hier bitte ich Sie, Ihre Bewertung der klaren Ersichtlichkeit zu explizieren: „keine Angabe“ (auch „k.A.“ oder „weiß nicht“ erlaubt), „siehe oben“ oder „trifft nicht“ zu würde bedeuten, dass die aktuelle Nutzung Ihrer Meinung nach gleich der ursprünglichen Funktion ist. Oder aber Sie kennen die Funktion oder Sie raten, auch das wäre hier erlaubt.

NAMING OF OBJECTS

Hierzu habe ich noch einen letzten Hinweis: Da geht es in den ersten beiden Fragen um die feste Bezeichnung - was meine ich damit: [falls keine Kirche auf der Route, falls schon Hauptbahnhof benutzt] Stellen Sie sich vor, wir stünden gerade vor dem Regensburger Dom und Sie würden dieses Gebäude bewerten. Dieses Gebäude hat eine feste Bezeichnung: Hoher Dom St. Peter zu Regensburg. Vor dem können wir uns auch verabreden. Da wissen Sie, wo Sie hinhüben und ich weiß es ebenso. Bei

der Bezeichnung oder Benennung, die Sie verwenden würden, geht es darum: Stellen Sie sich vor, Sie müssen unsere Route beschreiben und müssen nun auf genau dieses Objekt, das Sie gerade bewerten verweisen. Wie würden Sie das machen? Dazu dürfen Sie alles benutzen, von Eigenschaften des Objekts bis hin zur festen Bezeichnung; wichtig ist nur, dass genau dieses Objekt identifizierbar wird.

DISTANCES BETWEEN SUBSEQUENT OBJECTS

Das dauert immer kurz, bis das nächste Objekt angezeigt wird, weil die Daten direkt an den Server geschickt werden. Erst wenn dieser antwortet “Die Daten wurden erfolgreich übertragen”, dann dürfen Sie weiter. Gut, dann gehen wir zum nächsten Objekt. Diese sind übrigens nicht gleichverteilt auf der Route, d.h. manchmal sind die Gehdistanzen zwischen zwei Objekten lang und manchmal sehr kurz. [Die Anzahl von Objekten wurde nur auf Nachfragen mitgeteilt, um vorzubeugen, dass die Personen zu sehr die Einzelobjekte untereinander vergleichen: “es muss noch Luft nach oben sein” soll ja gerade nicht passieren]

Once a trial was finished information about how the winner of the Samsung Galaxy Tab 4 10.1 WiFi (cf. section 4.2.4) will be announced and how results of this research will be communicated to participants were given. Finally, the goals of the study were detailed in case participants were interested.

A.3.2 *Research Diary*

A German language research diary (cf. table 48) was used in order to keep track of the progress of experiments. For each trial, the following information was kept:

- the participant id associated,
- the date and time the trial was scheduled at,
- the route id, and
- specific incidents, i.e. one of
 - no important incidents at all,
 - special weather conditions, i.e. whether an umbrella was needed or snow covered the city,

- cancellations by participants including the reason they gave (if any),
- disturbances due to the software used for data acquisition,
- a different meeting point had to be chosen, because participants did acknowledge they were unable to find the starting point of the route.

As I give this data for the sake of completeness, only, no English translation is given.

Date	Time	Participant	Route	Comment
12.11.2014	14:45	mka-vp99	28	N/A
13.11.2014	09:30	mka-vp60	3	N/A
	14:45	mka-vp55	13	N/A
14.11.2014	09:30	mka-vp0	22	Verschoben auf Donnerstag, 20.11.2014, auf Grund technischer Probleme
	11:00	mka-vp43	5	N/A
17.11.2014	08:00	mka-vp49	4	onResume-Problem
	11:00	mka-vp75	27	2-maliger Absturz während Bewertung des ersten Objekts (Home-Button-Problem)
	13:15	mka-vp105	1	Absturz nach SoD (Home-Button-Problem) + onResume-Problem
18.11.2014	08:00	mka-vp74	23	Auf Grund der Witterung (Regen) verschoben auf Donnerstag, 20.11.2014
	09:30	mka-vp98	37	Antwortet nicht auf Absagemail; kommt aber auch nicht zum Treffpunkt; letzter Versuch: auf Mittwoch verlegen
	11:00	mka-vp37	13	VP sagt Termin wegen Regenwetter auf Grund einer ohnehin bereits vorhandenen Erkrankung ab; im Januar erneut kontaktieren
	13:15	mka-vp19	14	Bei Regen (mit Schirm)
19.11.2014	08:00	mka-vp41	1	Regenschirm nötig
	09:30	mka-vp98	37	Antwortet nicht auf Absagemail; kommt aber auch nicht zum Treffpunkt; erneute Mail mit Termin für heute; Antwort, dass geantwortet u. Januar präferiert ; auf Januar verschoben
	11:00	mka-vp53	32	nur 6 Objekte bewertet (fehlt mka_objects.uniqueid \$= 3505) (onResume-Problem) + Absturz nach SoD
	14:45	mka-vp32	23	Problem beim Laden der Frage, nach dem eigenen Standort im Vergleich zum Objekt und dem weiteren Routenverlauf
20.11.2014	11:45	mka-vp74	23	N/A
	13:00	mka-vp0	22	N/A
03.12.2014	09:30	mka-vp50	18	Absturz vor Schreiben von final aber alle Daten vorhanden
04.12.2014	09:30	mka-vp14	11	Problem, dass zweites Objekt auf Route nicht angezeigt wurde, folglich auch nicht bewertet
	11:00	mka-vp38	44	N/A
	13:15	mka-vp72	9	kurzfristig abgesagt
	14:45	mka-vp68	52	Objekt 397 wird vor 221 angezeigt; daher 221 nach 397 nachgeholt; dann ganz nach Reihenfolge mit 98 weiter
05.12.2014	09:30	mka-vp89	31	Versuchsperson erscheint nicht am vereinbarten Treffpunkt
	11:00	mka-vp23	46	Objekt 390 zunächst nicht angezeigt, stattdessen nach 291 gleich die 404; nach der 404 daher Objekt 390 nachgeholt
	13:15	mka-vp27	24	Objekt 136 (eigentlich das vorletzte auf der Route) zunächst nicht angezeigt; nach dem letzten Objekt nachgeholt
	14:45	mka-vp44	40	Nach Objekt 160 wurde gleich 234 angezeigt; sofort bemerkt und neu gestartet; 160 und 234 daher in richtiger Reihenfolge
08.12.2014	09:30	mka-vp94	41	N/A
	11:00	mka-vp73	54	Auf Grund einer Erkrankung auf 12.01. verschoben
	14:45	mka-vp18	38	N/A

Continued on next page

A APPENDIX

Tabelle 48 – *Continued from previous page*

Date	Time	Participant	Route	Comment
09.12.2014	13:15	mka-vp100	25	App zeigt falsche Bilder an; dann erneutes Treffen um 14:30 funktioniert, aber Abstürze
	14:45	mka-vp92	21	sagt am 08.12.2014 ab; in den Januar verschoben
10.12.2014	09:30	mka-vp93	36	zunächst falsch als VP 36 angemeldet; nach createpart u. sod korrigiert
	11:00	mka-vp25	34	Absturz
	13:15	mka-vp58	31	N/A
	14:45	mka-vp13	50	N/A
11.12.2014	09:30	mka-vp91	7	N/A
	11:00	mka-vp28	15	Bereits am Treffpunkt Verbindungsstörungen im Mobilfunk; createpart-Speicherung dauert extrem lange, Datei schließlich nicht vorhanden; Nach Neustart der App im Park funktioniert laden der Experimente nicht mehr; Abbruch nach 3 Objekten; vollständiger Fehlschlag
12.12.2014	09:30	mka-vp12	17	N/A
	11:00	mka-vp108	2	N/A
15.12.2014	09:30	mka-vp101	12	bei Regen (mit Schirm)
	11:00	mka-vp21	43	teilw. mit Schirm
	13:15	mka-vp31	33	N/A
	14:45	mka-vp22	49	teilw. mit Schirm
16.12.2014	11:00	mka-vp78	19	Probandin erscheint nicht am vereinbarten Treffpunkt
	13:15	mka-vp70	51	Probandin sagt kurzfristig krankheitsbedingt ab
	14:45	mka-vp80	45	Während der Bewertung des ersten Objekts „hängt“ die Eingabe manchmal; Neustart des Systems behebt Problem; bei Regen (mit Schirm)
17.12.2014	09:30	mka-vp109	30	sehr schnelle Versuchsperson (45 Minuten bei 8 Objekten)
	11:00	mka-vp81	48	Absturz während der Bewertung des letzten Objekts auf Grund von Home-Button; nochmalige Bewertung funktioniert
	13:15	mka-vp95	26	Person meldet sich auf Erinnerungsmail hin, dass sie bereits an einem URWalking-Experiment teilgenommen hat und entfällt somit als VP
	14:45	mka-vp11	55	mit Regenschirm; Absturz nach SoD-Fragebogen; nach Neustart App keine Probleme; Drittleztes Objekt liegt eigentlich woanders; vorletztes Bild aus falscher Perspektive, deshalb zunächst daran vorbei gegangen
18.12.2014	09:30	mka-vp84	53	Absturz nach SoD, danach reibungslos
	11:00	mka-vp87	16	VP erscheint nicht am vereinbarten Treffpunkt
	13:15	mka-vp45	47	mit Regenschirm
	14:45	mka-vp09	10	mit Regenschirm
19.12.2014	09:30	mka-vp86	8	VP sagt krankheitsbedingt ab; wird sich einen Termin im Januar suchen
	11:00	mka-vp29	39	DB korrupt; auf die Schnelle behebbar
	13:15	mka-vp102	20	N/A
	14:45	mka-vp87	16	N/A
12.01.2015	09:30	mka-vp1	15	N/A
	11:00	mka-vp73	9	letztes Objekt wird als zweites angezeigt; über already-done-Mechanismus behoben.
	13:15	mka-vp103	21	N/A
	14:45	mka-vp2	51	bei den letzten drei Objekten wird die Frage nach dem eigenen Standort im Vergleich zum Objekt und dem weiteren Routenverlauf nicht richtig angezeigt
13.01.2015	09:30	mka-vp15	8	VP erscheint nicht am vereinbarten Treffpunkt
	13:15		32	VP erscheint nicht am vereinbarten Treffpunkt
	14:45	mka-vp17	11	VP sagt über Nacht krankheitsbedingt ab
14.01.2015	09:30	mka-vp20	26	Versuchsperson sagt am Tag vorher wegen eines Krankenhausaufenthalts des Vaters ab.
	11:00	mka-vp24	4	sagt am 12.01. ab, weil er bereits an einem Experiment zur Fußgängernavigation teilgenommen hat
	13:15	mka-vp26	35	N/A
	14:45	mka-vp30	42	bei 3. u. 4. Objekt wird die Frage, nach dem eigenen Standort im Vergleich zum Objekt und dem weiteren Routenverlauf nicht richtig dargestellt
15.01.2015	09:30	mka-vp33	37	Versuchsperson sagt auf Grund von Lernstress ab
	11:00	mka-vp34	19	N/A
	13:15	mka-vp35	29	N/A
	14:45	mka-vp37	54	N/A

Continued on next page

A.3 PARTICIPANTS

Tabelle 48 – *Continued from previous page*

Date	Time	Participant	Route	Comment
16.01.2015	09:30	mka-vp39	6	N/A
	11:00	mka-vp3	32	N/A
	15:00	mka-vp40	11	N/A
17.01.2015	12:15	mka-vp42	8	am Beginn des ersten Objekts reagiert Tablet schlecht auf Eingaben; Neustart behebt Problem
	13:15	mka-vp104	4	N/A
	14:45	mka-vp46	26	N/A
19.01.2015	09:30	mka-vp47	37	N/A
	11:00	mka-vp15	10	N/A
	13:15	mka-vp16	23	N/A
	14:45	mka-vp17	33	Absturz nach SoD; Neustart führt gleich auf „Abmelden“-Schirm; dann erneuter Neustart der App behebt Problem
20.01.2015	11:00	mka-vp36	54	Absturz unmittelbar nach SoD und während der Bewertung des drittletzten Objekts
	13:15	mka-vp48	1	VP sagt am 19.01.2015 spätabends, ohne Gründe zu nennen, ab
	14:45	mka-vp4	27	N/A
21.01.2015	08:40	mka-vp52	6	N/A
	11:00	mka-vp51	20	Versuchsperson sagt am Vorabend ab, weil sie bereits an einem Fußgängernavigationsexperiment teilgenommen hat
	13:15	mka-vp54	16	Absturz nach SoD
	14:45	mka-vp56	13	N/A
22.01.2015	09:30	mka-vp57	22	N/A
	11:00	mka-vp59	29	Zweimal hintereinander stürzt Software ab, als Benennung für erstes Objekt geschrieben wird; dritter Versuch klappt und ab dann reibungsloser Verlauf.
	13:15	mka-vp5	36	N/A
	14:45	mka-vp61	39	N/A
23.01.2015	09:30	mka-vp62	30	N/A
	11:00	mka-vp70	44	Absturz beim 6ten Objekt - Home-Button-Problem
	13:15	mka-vp33	38	Absturz beim 1sten Objekt - Home-Button-Problem
	14:45	mka-vp63	14	Abbruch nach zwei Objekten, da Kleinkind nicht mehr zu beruhigen
24.01.2015	09:00	mka-vp92	45	N/A
26.01.2015	09:30	mka-vp6	50	N/A
	11:00	mka-vp72	11	N/A
	13:15	mka-vp7	37	N/A
	14:45	mka-vp8	43	N/A
27.01.2015	13:15	mka-vp10	42	N/A
	14:45	mka-vp20	28	N/A
28.01.2015	11:00	mka-vp24	1	N/A
	12:30	mka-vp48	4	N/A
	15:00	mka-vp51	3	N/A
29.01.2015	11:00	mka-vp64	25	N/A
	13:15	mka-vp		VP erscheint nicht am vereinbarten Treffpunkt
	14:45	mka-vp66	47	N/A
30.01.2015	09:30	mka-vp67	19	Absturz während SoD; Schneedecke
	11:00	mka-vp69	7	N/A; Schneedecke
	13:15	mka-71	12	N/A; Schneedecke
	14:45	mka-vp76	55	Drittletzes Objekt liegt eigentlich woanders; vorletztes Bild aus falscher Perspektive, deshalb zunächst daran vorbei gegangen; Schneedecke
31.01.2015	11:00	mka-vp96	17	Absturz bei Bewertung des dritten Objekts; Schneedecke
	13:00	mka-vp		VP bittet um Aufschub auf Grund eines Todesfalls in der Familie
02.02.2015	09:30	mka-vp		VP erscheint ohne Entschuldigung nicht am vereinbarten Treffpunkt
	11:00	mka-vp77	32	N/A; Schneedecke
	12:30	mka-vp78	52	N/A; Schneedecke
	14:45	mka-vp79	34	N/A; Absturz beim ersten Objekt; Schneedecke + Schneefall
03.02.2015	11:00	mka-vp		VP erscheint unentschuldig nicht am vereinbarten Treffpunkt
	13:15	mka-vp83	15	N/A; Schneedecke

Continued on next page

A APPENDIX

Tabelle 48 – *Continued from previous page*

Date	Time	Participant	Route	Comment
	14:45	mka-vp85	5	Treffen am E-R-Platz nötig, da VP nicht zum Cafe findet; am E-R-Platz creatpart u. SoD ausgefüllt, dann ohne App zum Cafe gegangen; Absturz nach SoD; Schneedecke
04.02.2015	09:30 11:00 13:00 14:45	mka-vp mka-vp mka-vp89 mka-vp	48	VP sagt am Montagmorgen ab. VP sagt am 03.02.2015 krankheitsbedingt ab N/A; 1.5 h Dauer VP ist um 14:55 nicht mehr am Treffpunkt; unklar, ob nicht erschienen oder nicht mehr da
06.02.2015	09:30 11:00 13:15 14:45	mka-vp110 mka-vp90 mka-vp95 mka-vp82	24 40 49 2	Absturz bei creatpart; dann als falsche VP für creatpart u. SoD angemeldet; danach reibungsloser Verlauf komme um 11:05 an; VP nicht da; erscheint nicht bis 11:25; zunächst keine Reaktion auf Anruf; um 11:50 erfolgt Rückruf und eine erneute Verabredung für 13:15 Uhr; für salientobject 1952 wird falsches Bild angezeigt (Zahlendreher in der DB) die Bewertungen für dieses Objekt aus der final getilgt; die anderen aber brauchbar VP sagt auf Grund einer Erkrankung tags zuvor ab N/A
09.02.2015	11:00 14:45	mka-vp106 mka-vp98	20 18	N/A N/A
10.02.2015	11:00 13:15 14:45	mka-vp97 mka-vp115 mka-vp112	35 8 31	Absturz beim ersten Objekt VP sagt krankheitsbedingt ab. N/A
11.02.2015	11:00 13:15	mka-vp86 mka-vp116	51 41	durch Erkrankung von MKA Absage nötig; VP findet keinen weiteren geeigneten Termin durch Erkrankung MKA verschoben auf 17.2.
12.02.2015	13:15	mka-vp114	14	N/A
14.02.2015	11:00 13:00 14:30	mka-vp65 mka-vp95 mka-vp107	21 49 46	N/A Absturz nach SoD N/A
17.02.2015	09:30 13:15	mka-vp111 mka-vp116	53 41	N/A; VP muss vor letztem Objekt zur Toilette; dauert etwa 13 Minuten N/A
18.02.2015	09:30 11:00 13:15 14:45	mka-vp88 mka-vp86 mka-vp113 mka-vp118	9 51 26 16	Das letzte Objekt wird wieder als zweites angezeigt, obwohl die DB korrekt ist; sofort bemerkt und händisch gelöst; dann das letzte Objekt nochmals nach Neustart händisch gelöst. VP bewertet bei zweitem Objekt zunächst den danebenliegenden Zaun N/A VP bewertet bei erstem Objekt zunächst das Danebenliegende; wird bei Benennungsfrage klar; Neustart Probandin bittet auf Grund beruflicher Verpflichtungen um Verschiebung
19.02.2015	10:30 12:30 14:45	mka-vp117 mka-vp119 mka-vp115	40 15 8	N/A N/A N/A

Tabelle 48.: The research diary, which was kept for each day experiments were scheduled for. If participants did cancel their participation the originally assigned ID was reused.

B

LIST OF ABBREVIATIONS

2D	two-dimensional
3D	three-dimensional
ADV	Visibility in Advance
approx.	approximately
AVE	Average Variance Extracted
BCa	bias-corrected and accelerated
CB-SEM	Covariance-based Structural Equation Modeling
cf.	confer
CI	confidence interval
CLM	Collaborative Landmark Mining
CM	Cognitive Map
COS	Cognitive Saliency
CRel	composite reliability
DB	database
DP	decision point
DV	dependent variable
e.g.	exemplum gerens
etc.	et cetera
FRS	Fragebogen Räumliche Strategien
GIS	Geographic Information Systems
GIScience	Geographical Information Science
HOC	higher order component
HTMT	Heterotrait-Monotrait Ratio of Correlations
ID3	Iterative Dichotomiser 3
i.e.	id est
IR	Information Retrieval
IS	Information Science
IV	independent variable
LM	landmark

B LIST OF ABBREVIATIONS

LOC	lower order component
LV	latent variable
MANOVA	Multivariate Analysis of Variance
MIMIC	Multiple Indicators, Multiple Causes
MV	measured variable
NB:	nota bene
OLS	Ordinary Least Squares
OSM	OpenStreetMap
OVSAL	Overall Saliency
p.	page
PLS-PM	Partial Least Squares Path Modeling
PLSc	Consistent Partial Least Squares Path Modeling
POI	point of interest
PNS	pedestrian navigation system
PRO	Protoypicality
SE	standard error
SEM	Structural Equation Model
SoD	sense of direction
SO	salient object
STS	Structural Saliency
TFIDF	Term Frequency Inverse Document Frequency
VGI	Volunteered Geographic Information
VIF	Variance Inflation Factor
VIS	Visual Saliency
VR	Virtual Reality

LIST OF FIGURES

Figure 1	An example of a pedestrian route in <i>Regensburg</i>	2
Figure 2	Locations of LMs with respect to DP adapted from Klippel and Winter (2005)	25
Figure 3	Bayesian network for <i>salience</i> estimations according to Caduff (2007)	28
Figure 4	SEM terminology explained	47
Figure 5	Variable names in PLS-PM explained	50
Figure 6	Reflective measurement explained	52
Figure 7	Formative measurement explained	53
Figure 8	The reading behavior of pre-study participants .	70
Figure 9	The way pre-study participants rated <i>salience</i> . .	70
Figure 10	The highly salient <i>Regensburg</i> Cathedral	77
Figure 11	The least salient building of all objects rated . .	77
Figure 12	A non-building example (circuit braker panel) .	78
Figure 13	The study area in <i>Regensburg</i> downtown	79
Figure 14	Route 22 as an example of a route course	82
Figure 15	An example screen of the survey application . .	85
Figure 16	A part of the SoD-rating screen	86
Figure 17	A part of the <i>salience</i> rating screen	88
Figure 18	A part of the <i>visual salience</i> rating screen	88
Figure 19	A specific tree used in the sample	99
Figure 20	A second specific tree used in the sample	100
Figure 21	The course of route 33 (least complex route) . .	102
Figure 22	The course route 17 (most complex)	103
Figure 23	Subsamples for the analysis of snowy weather impact	104
Figure 24	Forest plot of snowy weather influence	105
Figure 25	Model I: Independence of subdimensions	106
Figure 26	Model I: The SEM before estimation	108
Figure 27	Convergent validity assessment of the formative model part (before estimation)	118
Figure 28	The estimated convergent validity model	119

List of Figures

Figure 29	The second order model for <i>visual salience</i>	134
Figure 30	An example of simple mediation	138
Figure 31	An example of multiple mediation	140
Figure 32	An example of multiple mediation including a three-path mediation	141
Figure 33	Model II: Complex structural relationships . . .	142
Figure 34	Model II: Estimation results	148

LIST OF TABLES

Table 1	Verbs and prepositions used to describe the extents of SOs	17
Table 2	Difference between Lohmöller's and Wold's LV scores	56
Table 3	The questions used to measure OVSAL	60
Table 4	The questions used to measure PRO	61
Table 5	The questions used to measure ADV	62
Table 6	The questions used to measure COS	63
Table 7	The questions used to measure STS	64
Table 8	Visual dimensions revealed in existing studies .	65
Table 9	The questions used to measure VIS	67
Table 10	The reflective VIS questions	68
Table 11	The questions used in the pre-study	69
Table 12	Starting points of routes	81
Table 13	Survey application: Allocation of questions to screens	86
Table 14	Frequency of object categories	98
Table 15	Amount of time needed to finish a SO survey . .	101
Table 16	Composite reliability thresholds	111
Table 17	Reflective measurement model results	114
Table 18	The assessment of the Fornell-Larcker-Criterion	115
Table 19	The assessment of cross loadings	116
Table 20	AVE and CRel for elimination candidate items . .	116
Table 21	Assessment of reflective items for VIS	120
Table 22	VIFs for all causal-formative indicators	122
Table 23	Model I: Outer weights (OW) and loadings (OL) of causal-formative VIS indicators	126
Table 24	Model I: Predictive capabilities	132
Table 25	Model I: Results of HOC-LOC analysis of VIS . .	135
Table 26	Model I: Results of HOC-LOC-analysis	136
Table 27	A guide to interpreting results of mediation . .	138
Table 28	Model II: R^2 and Q^2 values	147

List of Tables

Table 29	Model II: Predictive capabilities for OVSAL . . .	149
Table 30	Model II: Predictive capabilities for STS	149
Table 31	Model II: Simple path coefficients	150
Table 32	Model II: Assessment of specific indirect effects	151
Table 33	Model II: HOC-LOC analysis of VIS	151
Table 34	ρ_A results based on PLS _c	169
Table 35	AVE: contrasting PLS-PM with PLS _c results	170
Table 36	Bootstrapping results for HTMT	171
Table 37	Different weights of dimensions according to Winter et al. (2005)	173
Table 38	Model I: PLS _c vs. PLS-PM results	174
Table 39	Model II: PLS _c vs. PLS-PM	175
Table 40	Model II: Comparing simple path coefficients of PLS _c and PLS-PM	176
Table 41	Model II: PLS _c vs. PLS-PM results for VIS	176
Table 42	Model II: PLS _c vs. PLS-PM specific indirect effects	177
Table 43	Model II: Predictive capabilities (PLS _c)	178
Table 44	Model II: R^2 of endogenous constructs for PLS-PM and PLS _c results	179
Table 45	Model II: Predictive capabilities for OVSAL based on PLS _c estimations	180
Table 46	The number of trials per day	203
Table 47	Route statistics based on starting points	205
Tabelle 48	The research diary	212

REFERENCES

- Aguirre-Urreta, M. I., & Marakas, G. M. (2014a). A Rejoinder to Rigdon et al. (2014). *Information Systems Research*, 25(4), 785–788.
- Aguirre-Urreta, M. I., & Marakas, G. M. (2014b). Research Note – Partial Least Squares and Models with Formatively Specified Endogenous Constructs: A Cautionary Note. *Information Systems Research*, 25(4), 761–778.
- Alwin, D. F., & Hauser, R. M. (1975). The Decomposition of Effects in Path Analysis. *American Sociological Review*, 40, 37–47.
- Appleyard, D. (1969). Why Buildings Are Known: A Predictive Tool for Architects and Planners. *Environment and Behavior*, 1(1), 131–156.
- Bainter, S. A., & Bollen, K. A. (2014). Interpretational Confounding or Confounded Interpretations of Causal Indicators? *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 125–140.
- Bainter, S. A., & Bollen, K. A. (2015). Moving Forward in the Debate on Causal Indicators: Rejoinder to Comments. *Measurement: Interdisciplinary Research and Perspectives*, 13(1), 63–74.
- Balaban, C. Z., Röser, F., & Hamburger, K. (2014). The effect of emotions and emotionally laden landmarks on wayfinding. In P. Bello, M. Guarani, M. McShane, & B. Scasselati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society, Austin, Texas* (pp. 3315–3320). Cognitive Science Society.
- Barclay, D. W., Higgins, C. A., & Thompson, R. (1995). The partial least squares approach to causal modeling: personal computer adoption and use as illustration. *Technology Studies*, 2(2), 285–309.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.

References

- Becker, J.-M. (2015, September). *Re: consistent PLS Bootstrapping*. online. Retrieved from <http://forum.smartpls.com/viewtopic.php?f=5&t=3531> (last accessed on April 18th, 2016)
- Bidwell, N. J., Lueg, C., & Axup, J. (2005). The territory is the map: designing navigational aids. In *Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: making CHI natural* (pp. 91–100). New York, NY, USA: ACM.
- Bienk, S., Kattenbeck, M., Ludwig, B., Müller, M., & Ohm, C. (2013). I Want to View it My Way – Interfaces to mobile maps should adapt to the user's orientation skills. In M. Kranz, K. Synnes, S. Boring, & K. Van Laerhoven (Eds.), *MUM '13: Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (p. article no. 34). New York, NY, USA: ACM.
- Boguraev, B., & Kennedy, C. (1997). Salience-based Content Characterisation of Text Documents. In *Advances in Automatic Text Summarization* (pp. 2–9). The MIT Press.
- Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology*, 17, 37–69.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. (2002). Latent Variables in Psychology and the Social Sciences. *Annual Review of Psychology*, 53, 605–634.
- Bollen, K. A., & Davis, W. R. (2009a). Causal Indicator Models: Identification, Estimation, and Testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 498–522.
- Bollen, K. A., & Davis, W. R. (2009b). Two Rules of Identification for Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 523–536.
- Bollen, K. A., & Diamantopoulos, A. (2015). In Defense of Causal-Formative Indicators: A Minority Report. *Psychological Methods*, no pagination specified. (Advance online publication)

- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation Perspective. *Psychological Bulletin*, 110(2), 305–314.
- Box, G. E. P., & Draper, N. R. (1987). *Empirical Model Building and Response Surfaces*. New York, NY: John Wiley & Sons.
- Brunyé, T. T., Gardony, A., Mahoney, C. R., & Taylor, H. A. (2012). Going to town: Visualized perspectives and navigation through virtual environments. *Computers in Human Behavior*, 28(1), 257–266.
- Burnett, G. E. (1998). "Turn right at the King's Head". *Drivers' requirements for route guidance information* (Unpublished doctoral dissertation). Loughborough University.
- Burnett, G. E., Smith, D., & May, A. (2001). Supporting the Navigation Task: Characteristics of 'good' Landmarks. In *Contemporary Ergonomics* (pp. 441–446). Taylor & Francis.
- Burt, R. S. (1976). Interpretational Confounding of Unobserved Variables in Structural Equation Models. *Sociological Methods & Research*, 5, 3–52.
- Caduff, D. (2007). *Assessing Landmark Salience for Human Navigation* (Unpublished doctoral dissertation). Mathematisch-naturwissenschaftliche Fakultät der Universität Zürich, Zürich.
- Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing*, 9(4), 249–267.
- Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A Practical guide for medical statisticians. *Statistics in Medicine*, 19, 1141–1164.
- Case, D. O., & Spink, A. (Eds.). (2012). *Looking for Information A Survey of Research on Information Seeking, Needs and Behavior* (3rd ed.). Emerald Group Publishing Limited.
- Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, 26(4), 435–446.

References

- Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of Formative Measurement in Information Systems Research. *Management Information Systems Quarterly*, 33(4), 689–707.
- Chernick, M. R. (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers*. Hoboken, NJ: John Wiley & Sons.
- Chin, W. W. (1998). The Partial Least Squares Approach to Structural Equation Modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–358). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Chin, W. W., & Newsted, P. R. (1999). Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares. In R. H. Hoyle (Ed.), *Statistical Methods for Small Sample Research* (pp. 307–342). SAGE Publications.
- Claramunt, C. (2007). Structural salience of elements of the city. *Environment and Behavior*, 34(6), 1030–1050.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). New York and London: Routledge.
- Colluccia, E., & Louse, G. (2004). Gender differences in spatial orientation. A review. *Journal of Environmental Psychology*, 24, 329–340.
- Conklin, E. J., & McDonald, D. D. (1982). Saliency: The Key to the Selection Problem in Natural Language Generation. In *Proceedings of the 20th Annual Meeting on Association for Computational Linguistics* (pp. 129–135). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Couclelis, H., Golledge, R., Gale, N., & Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7(2), 99–122.
- Crandall, D., Backstrom, L., Huttenlocher, D. P., & Kleinberg, J. (2009). Mapping the world's photos. In Y. Maarek & W. Nejdl (Eds.),

- International world wide web conference* (pp. 761–770). Madrid, Spain: ACM Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological tests. *Psychological Bulletin*, *52*(4), 281–302.
- Cumming, G. (2012). *Understanding the New Statistics. Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.
- Daniel, M.-P., & Denis, M. (1998). Spatial Descriptions as Navigational Aids: A Cognitive Analysis of Route Directions. *Kognitionswissenschaft*, *7*, 45–52.
- Daniel, M.-P., Przytula, E., & Denis, M. (2009). Spoken versus written route directions. *Cognitive Processing*, *10*, 201–203.
- Davies, C., & Peebles, D. (2010). Spaces or Scenes: Map-based Orientation in Urban Environments. *Spatial Cognition & Computation*, *10*(2-3), 135–156.
- Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive - Current Psychology of Cognition*, *16*(4), 409–458.
- Denis, M., Pazzaglia, F., Cornoldi, C., & Bertolo, L. (1999). Spatial discourse and navigation: an analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, *13*(2), 145–174.
- DeVellis, R. F. (2012). *Scale Development. Theory and Applications* (3rd ed.). Thousand Oaks, CA et al.: SAGE Publications.
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*, 434–449.
- Diamantopoulos, A., & Siguaw, J. A. (2006). Formative Versus Reflective Indicators in Organizational Measure Development. *British Journal of Management*, *17*(4), 263–282.

References

- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, 38, 269–277.
- Dijkstra, T. K. (1983). Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics*, 22(1-2), 67–90.
- Dijkstra, T. K., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, 81, 10–23.
- Dijkstra, T. K., & Henseler, J. (2015b). Consistent Partial Least Squares Path Modeling. *Management Information Systems Quarterly*, 39(2), 297–316.
- Doodle AG. (2016). *Doodle: Termine einfach schneller finden*. Retrieved from <http://doodle.com/de/> (last access April 29th, 2016)
- Drolet, A. L., & Morrison, D. G. (2001). Do We Really Need Multiple-Item Measures in Service Research? *Journal of Service Research*, 3, 196–204.
- Duckham, M., Winter, S., & Robinson, M. (2010). Including landmarks in routing instructions. *Journal of Location Based Services*, 4(1), 28–52.
- Dunietz, J., & Gillick, D. (2014). A New Entity Salience Task with Millions of Training Examples. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers* (pp. 205–209). Gothenburg, Sweden: Association for Computational Linguistics.
- Eddelbuettel, D. (2015). random: True Random Numbers using RANDOM.ORG [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=random> (R package version 0.2.5)
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.

- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1–26.
- Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, 82(397), 171–185.
- Elias, B. (2003a). Determination of landmarks and reliability criteria for landmarks. In *Proceedings of the Fifth ICA Commission on Map Generalization Workshop on Progress in Automated Map Generalization*.
- Elias, B. (2003b). Extracting landmarks with data mining methods. In W. Kuhn, M. Worboys, & S. Timpf (Eds.), *Spatial Information Theory, Proceedings: Foundations of Geographic Information Science* (Vol. 2825, pp. 375–389).
- Elias, B. (2006). *Extraktion von Landmarken für die Navigation* (Unpublished doctoral dissertation). Fakultät für Bauingenieurwesen und Geodäsie der Universität Hannover, Hannover.
- Fang, Z., Li, Q., Zhang, X., & Shaw, S.-L. (2012). A GIS data model for landmark-based pedestrian navigation. *International Journal of Geographical Information Science*, 26(5), 1–22.
- Farr, A. C., Kleinschmidt, T., Yarlagadda, P., & Mengersen, K. (2012). Wayfinding: A simple concept, a complex process. *Transport Reviews*, 32(6), 715–743.
- Fassott, G., & Eggert, A. (2005). Zur Verwendung formativer und reflektiver Indikatoren in Strukturgleichungsmodellen: Bestandsaufnahme und Anwendungsempfehlungen. In F. Bliemel, A. Eggert, G. Fassott, & J. Henseler (Eds.), *Handbuch PLS-Pfadmodellierung. Methoden, Anwendung, Praxisbeispiele* (pp. 31–47). Schäffer-Poeschel.
- Feuersänger, C., & Tantau, T. (2015). PGF and TikZ (v 3.0.1) – Graphic systems for TeX [Computer software manual]. Retrieved from <https://sourceforge.net/projects/pgf/> (last access on April 30th, 2016)
- Fontaine, S., Edwards, G., Tversky, B., & Denis, M. (2005). Expert and Non-expert Knowledge of Loosely Structured Environments. In

References

- A. Cohn & D. Mark (Eds.), *Spatial Information Theory* (Vol. 3693, pp. 363–378). Springer Berlin / Heidelberg.
- Fornell, C. G., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 19, 440–452.
- Fornell, C. G., & Cha, J. (1994). Partial least squares. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 52–78). Cambridge, England: Blackwell Publishing Ltd.
- Fornell, C. G., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 3–50.
- Furlan, A., Baldwin, T., & Klippel, A. (2007). Landmark Classification for Route Directions. In *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions* (pp. 9–16). Prague.
- Gärling, T., Böök, A., & Lindberg, E. (1986). Spatial orientation and wayfinding in the designed environment: a conceptual analysis and some suggestions for postoccupancy evaluations. *Journal of Architectural and Planning Research*, 3, 55–64.
- Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, 4, Article 7.
- Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika*, 61(1), 101–107.
- Glenn, B. T., & Chignell, M. H. (1992). Hypermedia: Design for Browsing. In H. R. Hartson & D. Hix (Eds.), (Vol. 3). Norwood, NJ: Ablex Publishing Corporation.
- Godbold, N. (2006). Beyond information seeking: towards a general model of information behaviour. *iR – Information Research*, 11(4), paper 269.
- Goferman, S., Zelnik-Manor, L., & Tal, A. (2012). Context-Aware Saliency Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10), 1915–1926.

- Golledge, R. G. (Ed.). (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore, MA: The Johns Hopkins University Press.
- Goodchild, M. F. (1992). Geographical Information science. *International Journal of Geographical Information Systems*, 6(1), 31–45.
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal. Spatially Integrated Social Sciences and Humanities*, 69(4), 211–221.
- Goodchild, M. F. (2010). Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science*, 1, 3–20.
- Goodman, J. M., Brewster, S. A., & Gray, P. (2005). How can we best use landmarks to support older people in navigation? *Behavior & Information Technology*, 24(1), 3–20.
- Google Inc. (2016). *Google Maps*. Retrieved from <https://www.google.com/maps/dir/Reichstagsmuseum,+Rathauspl.+1,+93047+Regensburg,+Deutschland/49.0184516,12.099593/@49.0193424,12.0963348,18z/am=t/data=!3m1!4b1!4m9!4m8!1m5!1m1!1s0x479fc1af5106d7cf:0x87bf6f75c3c6277e!2m2!1d12.0945518!2d49.0203349!1m0!3e2> (last access April 22nd, 2016)
- Gould, W., & Pitblado, J. (2015). *How large should the bootstrapped samples be relative to the total number of cases in the dataset?* Retrieved from <http://www.stata.com/support/faqs/statistics/bootstrapped-samples-guidelines/> (last access April 4th, 2015)
- Haenlein, M., & Kaplan, A. M. (2004). A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics*, 3(4), 283–297.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis. A Global Perspective* (7th ed.). Upper Saddle River, NJ: Person Education.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19(2), 139–151.

References

- Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Los Angeles et al.: SAGE Publications.
- Haklay, M. M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4), 12–18.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Hamburger, K., & Röser, F. (2011). The Meaning of Gestalt for Human Wayfinding - How Much Does it Cost to Switch Modalities? *Gestalt Theory*, 33(3), 363–382.
- Hamburger, K., & Röser, F. (2014). The Role of Landmark Modality and Familiarity in Human Wayfinding. *Swiss Journal of Psychology*, 73(4), 205–213.
- Hansen, S., Richter, K.-F., & Klippel, A. (2006). Landmarks in OpenLS—A Data Structure for Cognitive Ergonomic Route Directions. In M. Raubal, H. Miller, A. Frank, & M. Goodchild (Eds.), *Geographic Information Science* (Vol. 4197, pp. 128–144). Berlin / Heidelberg: Springer.
- Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, 12(1), 1–17.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimator. *Journal of Educational Statistics*, 6(2), 107–128.
- Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30(5), 425–447.
- Heisenberg, W. (1958). *Physics and philosophy*. Retrieved from https://en.wikiquote.org/wiki/Werner_Heisenberg (last access May 2nd, 2016)
- Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics*, 25(1), 107–120.

- Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., . . . Calantone, R. J. (2014). Common Beliefs and Reality About PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods, 17*(2), 182–209.
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial Management & Data Systems, 116*(1), 2–20.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135.
- Hidayetoglu, M. L., Yildirim, K., & Akalin, A. (2012). The effects of color and light on indoor wayfinding and the evaluation of the perceived environment. *Journal of Environmental Psychology, 32*(1), 50–58.
- Hirtle, S. C., & Raubal, M. (2013). Many to Many Mobile Maps. In M. Raubal, D. M. Mark, & A. U. Frank (Eds.), *Cognitive and Linguistic Aspects of Geographic Space. New Perspectives on Geographic Information Research* (pp. 141–157). Heidelberg et al.: Springer.
- Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal, 20*, 195–204.
- Iachini, T., & Logie, R. H. (2003). The Role of Perspective in Locating Position in a Real-World, Unfamiliar Environment. *Applied Cognitive Psychology, 17*(6), 715–732.
- Iaria, G., Palermo, L., Committeri, G., & Barton, J. (2009). Age differences in the formation and use of cognitive maps. *Behavioural Brain Research, 196*, 187–191.
- Ingwersen, P., & Järvelin, K. (2005). *The Turn. Integration of Information Seeking and Retrieval in Context* (Vol. 18). Berlin/Heidelberg: Springer.

References

- Ishikawa, T., & Nakamura, U. (2012). Landmark Selection in the Environment: Relationships with Object Characteristics and Sense of Direction. *Spatial Cognition & Computation*, 12(1), 1-22.
- Itti, L. (2007). Visual salience [rev 72776]. *Scholarpedia*, 2(9), 3327.
- Itti, L., & Koch, C. (2000). A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention. *Vision Research*, 40, 1489-1506.
- Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254-1259.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills: Sage Publications.
- Jansen-Osmann, P., Schmid, J., & Heil, M. (2007). Spatial knowledge of adults and children in a virtual environment: The role of environmental structure. *European Journal of Developmental Psychology*, 4(3), 251-272.
- Janzen, G., & C., J. (2010). A neural wayfinding mechanism adjusts for ambiguous landmark information. *NeuroImage*, 52(1), 364-370.
- Janzen, G., & Turennout, M. v. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, 7, 673-677.
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, 30, 199-218.
- Jones, K. S. (1999). Automatic summarising: factors and directions. In I. Mani & M. Maybury (Eds.), *Advances in automatic text summarisation* (pp. 1-12). Cambridge, MA: MIT Press.
- Jöreskog, K. (1971). SIMULTANEOUS FACTOR ANALYSIS IN SEVERAL POPULATIONS. *Psychometrika*, 36, 409-426.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602-619.

- Karimi, H. A. (2011). *Universal Navigation on Smartphones*. Springer New York Dordrecht Heidelberg London.
- Kattenbeck, M. (2012). Handlungsorientierte Modellierung von Landmarken im Innen- und Außenbereich. Eine informationswissenschaftliche Fragestellung. *Information – Wissenschaft & Praxis*, 63(5), 338–342.
- Kattenbeck, M. (2015a). Empirically Measuring Saliency of Objects for Use in Pedestrian Navigation. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 3:1–3:10). New York, NY, USA: ACM.
- Kattenbeck, M. (2015b). How Do Participants Deal with Questions if Asked to Rate Visual Dimensions of an Object? In F. Pehar, C. Schlögl, & C. Wolff (Eds.), *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science, ISI 2015, Zadar, Croatia, May 19-21, 2015* (pp. 313–319). Verlag Werner Hülsbusch.
- Kattenbeck, M., Brockelmann, M., Hammwöhner, R., Jackermeier, R., & Ludwig, B. (2013). Vermessungsdaten - OpenStreetMap - In-Situ-Experimente. Die Datengrundlage von URWalking. In H.-C. Hobohm (Ed.), *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Proceedings des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013) Potsdam, 19. bis 22. März 2013* (pp. 72–383). Glückstadt: Verlag Werner Hülsbusch (Reihe Nr. 63), p. 372-383.
- Kattenbeck, M., Müller, M., Ohm, C., & Ludwig, B. (2015). Der Weg ist das Ziel? Fußgängernavigation ist Forschung zu Information Behavior. *Information – Wissenschaft & Praxis*, 66(1), 45–55.
- Kelley, K. (2005). The Effects of Nonnormal Distributions on Confidence Intervals Around the Standardized Mean Difference: Bootstrap and Parametric Confidence Intervals. *Educational and Psychological Measurement*, 65(1), 51–69.
- Kennedy, L. S., & Naaman, M. (2008). Generating Diverse and Representative Image Search Results for Landmarks. In *Proceedings of*

References

- the 17th International Conference on the World Wide Web, WWW* (pp. 297–306). ACM, New York.
- Kenny, D. A. (1979). *Correlation and Causality*. John Wiley & Sons.
- Kettunen, P., Irvankoski, K., Krause, C. M., & Sarjakoski, L. T. (2013). Landmarks in nature to support wayfinding: the effects of seasons and experiment methods. *Cognitive Processing*, 14(3), 245–253.
- Klarner, P., Sarstedt, M., Hoeck, M., & Ringle, C. M. (2013). Disentangling the Effects of Team Competences, Team Adaptability, and Client Communication on the Performance of Management Consulting Teams. *Long Range Planning*, 43(3), 258–286.
- Kline, R. B. (2013). *Beyond Significance Testing. Statistics Reform in the Behavioral Sciences* (2nd ed.). American Psychological Association.
- Klippel, A., Hansen, S., Richter, K.-F., & Winter, S. (2009). Urban granularities—a data structure for cognitively ergonomic route directions. *Geoinformatica*, 13(2), 223–247.
- Klippel, A., & Winter, S. (2005). Structural salience of landmarks for route directions. In *Proceedings of the 2005 International Conference on Spatial Information Theory* (pp. 347–362). Springer-Verlag Berlin / Heidelberg.
- Lazem, S. Y., & Sheta, W. M. (2005). Automatic Landmark Identification in Large Virtual Environment: A Spatial Data Mining Approach. In *Ninth International Conference on Information Visualisation (IV'05)* (pp. 752–757). Los Alamitos, CA, USA: IEEE Computer Society.
- Lee, Y., Kwong, A., Pun, L., & Mack, A. (2001). Multi-Media Map for Visual Navigation. *Journal of Geospatial Engineering*, 3(2), 87–96.
- Lindenberger, U., & Ghisletta, P. (2009). Cognitive and Sensory Declines in Old Age: Gauging the Evidence for a Common Cause. *Psychology and Aging*, 24(1), 1–16.
- Lioma, C., Blanco, R., & Moens, M.-F. (2009). A Logical Inference Approach to Query Expansion with Social Tags. In L. Azzopardi et al. (Eds.), *Advances in Information Retrieval Theory* (pp. 358–361). Springer Berlin/Heidelberg.

- Lloret, E., & Palomar, M. (2012). Text Summarisation in Progress: A Literature Review. *Artificial Intelligence Review*, 37(1), 1–41.
- Lohmöller, J. B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica.
- Lovelace, K. L., Hegarty, M., & Montello, D. R. (1999). Elements of good route directions in familiar and unfamiliar environments. In Freksa, C. and Mark, D. M. (Ed.), *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science* (pp. 65–82).
- Ludwig, B., Bienk, S., Kattenbeck, M., Müller, M., Ohm, C., Einmal, M., . . . Schubart, L. (2013). Do You Recognize that Building's Facade? *Künstliche Intelligenz*, 27(3), 241–246.
- Lynch, K. (1960). *The Image of the City*. Cambridge, MA and London, England: The M.I.T. Press.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence Limits for the Indirect effect: Distribution of the Product and Resampling Methods. *Multivariate Behavioral Research*, 39, 99–128.
- Mani, I., & Maybury, M. (1999). *Advances in automatic text summarization*. Cambridge, MA: The M.I.T. Press.
- Mark, D. M. (2003). Geographic information science: Defining the field. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of geographic information science* (pp. 1–18). New York: Taylor & Francis.
- Mateos-Aparicio Morales, G. (2011). Partial Least Squares (PLS) Methods: Origins, Evolution, and Application to Social Sciences. *Communications in Statistics - Theory and Methods*, 40(13), 2305–2317.
- Matsumoto, T., Satake, S., Kanda, T., Imai, M., & Hagita, N. (2012). Do you remember that shop?—Computational model of spatial memory for shopping companion robots. In *Proceedings of the seventh annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 447–454). New York, NY, USA: ACM.

References

- May, A. J., Ross, T., Bayer, S. H., & Tarkiainen, M. J. (2003). Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing*, 7(6), 331–338.
- McDonald, R. P. (1996). Path Analysis with Composite Variables. *Multivariate Behavioral Research*, 31, 239–270.
- Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, 32, 89–115.
- Michon, P.-E., & Denis, M. (2001). When and Why Are Visual Landmarks Used in Giving Directions? In D. Montello (Ed.), *Spatial Information Theory* (pp. 292–305). Berlin / Heidelberg: Springer.
- Miller, J., & Carlson, L. (2011). Selecting landmarks in novel environments. *Psychonomic Bulletin & Review*, 18, 184–191.
- Montello, D. R., & Xiao, D. (2011). Linguistic and Cultural Universality of the Concept of Sense-of-Direction. In M. Egenhofer, N. Giudice, R. Moratz, & M. Worboys (Eds.), *Spatial information theory* (pp. 264–282). Berlin/Heidelberg: Springer.
- Mummidi, L., & Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal. Spatially Integrated Social Sciences and Humanities*, 72(3), 215–227.
- Münzer, S., & Hölscher, C. (2011). Entwicklung und Validierung eines Fragebogens zu räumlichen Strategien. *Diagnostica*, 57(3), 111–125.
- Niu, Y., Todd, R. M., Kyan, M., & Anderson, A. K. (2012). Visual and Emotional Salience Influence Eye Movements. *ACM Transactions on Applied Perception*, 9(3), 13:1–13:18.
- Nothegger, C., Winter, S., & Raubal, M. (2004). Selection of Salient Features for Route Directions. *Spatial Cognition & Computation*, 4(2), 113–136.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw & Hill.

- O'Day, V. L., & Jeffries, R. (1993). Orienteering in an Information Landscape: How Information Seekers Get from Here to There. In B. Arnold, G. van der Veer, & T. White (Eds.), *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (pp. 438–445). New York, NY, USA: ACM.
- Ohm, C., Bienk, S., Kattenbeck, M., Ludwig, B., & Müller, M. (2016). Towards interfaces of mobile pedestrian navigation systems adapted to the user's orientation skills. *Pervasive and Mobile Computing*, 26, 121–134.
- Onsrud, H., & Kuhn, W. (Eds.). (2016). *Advancing Geographic Information Science: The Past and Next Twenty Years*. Global Spatial Data Infrastructure Association.
- Park, Y., & Li, Y. (2006). Extracting Salient Keywords from Instructional Videos Using Joint Text, Audio and Visual Cues. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL, New York, June 2006* (pp. 109–112). Association for Computational Linguistics.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107–123.
- Partala, T., Nurminen, A., Vainio, T., Laaksonen, J., Laine, M., & Väänänen, J. (2010). Salience of visual cues in 3D city maps. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 428–432). British Computer Society Swinton, UK.
- Peters, D., Wu, Y., & Winter, S. (2010). Testing Landmark Identification Theories in Virtual Environments. In C. Hölscher, T. Shipley, M. Olivetti Belardinelli, J. Bateman, & N. Newcombe (Eds.), *Spatial Cognition VII* (pp. 54–69). Springer Berlin / Heidelberg.
- Piccardi, L., Risetti, M., & Nori, R. (2011). Familiarity and Environmental Representations of a City: A Self-report Study. *Psychological Reports*, 109(1), 309–326.

References

- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers*, 36(3), 717–731.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891.
- Presson, C. C., & Montello, D. R. (1988). Points of reference in spatial cognition: Stalking the elusive landmark. *British Journal of Developmental Psychology*, 6(4), 378–381.
- QGIS Development Team. (2014). QGIS Geographic Information System, v2.6.0 [Computer software manual].
- Quesnot, T., & Roche, S. (2015). Quantifying the Significance of Semantic Landmarks in Familiar and Unfamiliar Environments. In S. Fabrikant, M. Raubal, M. Bertolotto, C. Davies, S. Freundschuh, & S. Bell (Eds.), *Spatial Information Theory* (Vol. 9368, pp. 468–489). Springer International Publishing.
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1, 81–106.
- R Core Team. (2013). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/> (last access May 1st, 2016)
- Raubal, M., Mark, D. M., & Frank, A. U. (Eds.). (2013). *Cognitive and Linguistic Aspects of Geographic Space. New Perspectives on Geographic Information Research*. Heidelberg et al.: Springer.
- Raubal, M., & Winter, S. (2002). Enriching Wayfinding Instructions with Local Landmarks. In M. Egenhofer & D. Mark (Eds.), *Geographic Information Science* (pp. 243–259). Berlin / Heidelberg: Springer.
- Rehrl, K., Häusler, E., & Leitinger, S. (2010). Comparing the effectiveness of GPS-enhanced voice guidance for pedestrians with metric- and landmark-based instruction sets. In *Proceedings of the 6th International Conference on Geographic Information Science* (pp. 189–203). Berlin / Heidelberg: Springer-Verlag.

- Reinartz, W. J., Haenlein, M., & Henseler, J. (2009). An Empirical Comparison of the Efficacy of Covariance-based and Variance-based SEM. *International Journal of Research in Marketing*, 26, 332–344.
- Richter, K.-F. (2007). A Uniform Handling of Different Landmark Types in Route Directions. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *Spatial Information Theory* (pp. 373–389). Berlin / Heidelberg: Springer.
- Richter, K.-F. (2013). Prospects and Challenges of Landmarks in Navigation Services. In M. Raubal, D. M. Mark, & A. U. Frank (Eds.), *Cognitive and Linguistic Aspects of Geographic Space. New Perspectives on Geographic Information Research* (pp. 83–97). Heidelberg et al.: Springer.
- Richter, K.-F., & Duckham, M. (2008). Simplest Instructions: Finding Easy-to-Describe Routes for Navigation. In *Proceedings of the 5th International Conference on Geographic Information Science* (pp. 274–289). Berlin / Heidelberg: Springer-Verlag.
- Richter, K.-F., & Winter, S. (2011a). Citizens as Database: Conscious Ubiquity in Data Collection. In *12th International Symposium on Spatial and Temporal Databases* (pp. 445–448).
- Richter, K.-F., & Winter, S. (2011b). Harvesting User-Generated Content for Semantic Spatial Information: The Case of Landmarks in OpenStreetMap. In *Proceedings of the SURVEYING & SPATIAL SCIENCES BIENNIAL CONFERENCE 2011, 21-25 November 2011, Wellington, New Zealand*.
- Richter, K.-F., & Winter, S. (2014). *Landmarks. GIScience for Intelligent Services*. Springer International Publishing.
- Rigdon, E. E. (2012). Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods. *Long Range Planning*, 45, 341–358.
- Rigdon, E. E., Becker, J.-M., Rai, A., Ringle, C. M., Diamantopoulos, A., Karahanna, E., . . . Dijkstra, T. K. (2014). Conflating Antecedents and Formative Indicators: A Comment on Aguirre-Urreta and Marakas. *Information Systems Research*, 25(4), 780–784.

References

- Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. SmartPLS GmbH. Retrieved from <http://www.smartpls.com> (last access May 2nd, 2016)
- Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222–245.
- Rönkkö, M. (2014). The Effects of Chance Correlation on Partial Least Squares Path Modeling. *Organizational Research Methods*, 17(2), 161–184.
- Rönkkö, M., & Evermann, J. (2013). A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. *Organizational Research Methods*, 16(3), 425–448.
- Rönkkö, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences*, 87, 76 – 84.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, 8, 382–439.
- Röser, F. (2015). *The cognitive observer-based landmark-preference model—What is the ideal landmark position at an intersection?* Retrieved from http://geb.uni-giessen.de/geb/volltexte/2015/11640/pdf/RoeserFlorian_2015_07_15.pdf (last access May 2nd, 2016)
- Röser, F., Hamburger, K., & Knauff, M. (2011). The Giessen virtual environment laboratory: human wayfinding and landmark salience. *Cognitive Processing*, 12, 209–214.
- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19, 305–335.
- Sadeghian, P., & Kantardzic, M. (2008). The New Generation of Automatic Landmark Detection Systems: Challenges and Guidelines. *Spatial Cognition & Computation*, 8(3), 252–287.

- Sándor, A., & de Waard, A. (2012). Identifying Claimed Knowledge Updates in Biomedical Research Articles. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse* (pp. 10–17). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Sarjakoski, T., Kettunen, P., Halkosaari, H.-M., Laakso, M., Rönneberg, M., Stigmar, H., & Sarjakoski, T. (2013). Landmarks and a Hiking Ontology to Support Wayfinding in a National Park During Different Seasons. In M. Raubal, D. M. Mark, & A. U. Frank (Eds.), *Cognitive and Linguistic Aspects of Geographic Space. New Perspectives on Geographic Information Research*. Heidelberg et al.: Springer.
- Sarstedt, M., Ringle, C. M., & Hair, J. F. (2014). PLS-SEM: Looking Back and Moving Forward. *Long Range Planning*, 47(3), 132–137.
- Schaeffer, N. C., & Presser, S. (2003). The Science of Asking Questions. *Annual Review of Sociology*, 29, 65–88.
- Schroder, C. J., Mackaness, W. A., & Gittings, B. M. (2011). Giving the 'Right' Route Directions: The Requirements for Pedestrian Navigation Systems. *Transactions in GIS*, 15(3), 419–438.
- Sefelin, R., Bechinie, M., Müller, R., Seibert-Giller, V., Messner, P., & Tscheligi, M. (2005). Landmarks: Yes; but Which? Five Methods to Select Optimal Landmarks for a Landmark- and Speech-based Guiding System. In M. Tscheligi, R. Bernhaupt, & K. Mihalic (Eds.), *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services* (pp. 287–290). New York, NY, USA: ACM.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3), 379–423.
- Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In H. Reese (Ed.), *Advances in child development and behaviour* (pp. 9–55). New York: Academic Press.
- Sin, S.-C. J. (2011). Towards Agency-Structure Integration: A Person-in-Environment (PIE) Framework for Modeling Individual-Level

References

- Information Behaviours and Outcomes. In A. Spink & J. Heinström (Eds.), *New Directions in Information Behaviour* (pp. 181–209). Howard House, UK: Emerald Group Publishing Limited.
- Snowdon, C., & Kray, C. (2009). Exploring the Use of Landmarks for Mobile Navigation Support in Natural Environments. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 13:1–13:10). New York, NY, USA: ACM.
- Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In *Sociological methodology* (pp. 290–312). Jossey-Bass.
- Soper, D. (2015). *A-priori Sample Size Calculator for Structural Equation Models [Software]*. Retrieved from <http://www.danielsoper.com/statcalc/default.aspx> (last access September 29th, 2015)
- Sorrows, M., & Hirtle, S. (1999). The Nature of Landmarks for Real and Electronic Spaces. In C. Freksa & D. Mark (Eds.), *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science* (pp. 37–50). Berlin / Heidelberg: Springer.
- Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221–257). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111–147.
- Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the Three-Path Mediated Effect. *Organizational Research Methods*, 11(2), 241–269.
- Tezuka, T., & Tanaka, K. (2005). Landmark Extraction: A Web Mining Approach. In A. Cohn & D. Mark (Eds.), *Spatial Information Theory* (pp. 379–396). Springer Berlin / Heidelberg.

- The GIMP Team. (1997-2013). *GNU Image Manipulation Program, GIMP 2.8.6*. Retrieved from <http://www.gimp.org> (last access May 2nd, 2016)
- The Inkscape Team. (2016). *Inkscape 0.91*. Retrieved from <https://inkscape.org/de/> (last access April 26th, 2016)
- Tom, A., & Denis, M. (2003). Referring to Landmark or Street Information in Route Directions: What Difference Does It Make? In W. Kuhn, M. Worboys, & S. Timpf (Eds.), *Spatial Information Theory. Foundations of Geographic Information Science* (p. 362-374). Berlin / Heidelberg: Springer.
- UNESCO World Heritage Convention. (2006). *Old Town of Regensburg with Stadtamhof – inscribed property*. online resource. Retrieved from <http://whc.unesco.org/document/117721> (last access April 20th, 2016)
- University Consortium for Geographic Information Science. (2016). *By Laws of the University Consortium for Geographic Information Science: ARTICLE II PURPOSE*. online. Retrieved from <http://ucgis.org/basic-page/laws> (last access on April 22nd, 2016)
- Valuch, C., Becker, S. I., & Ansorge, U. (2013). Priming of fixations during recognition of natural scenes. *Journal of Vision*, 13(3), 1–22.
- Vinson, N. G. (1999). Design Guidelines for Landmarks to Support Navigation in Virtual Environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit* (pp. 278–285). New York, NY, USA: ACM.
- Vinzi, V. E., Trinchera, L., & Amato, S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares* (pp. 47–82). Berlin/Heidelberg: Springer-Verlag.
- Weiber, R., & Mühlhaus, D. (2010). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Heidelberg et al.: Springer.

References

- Werts, C., Linn, R., & Jöreskog, K. (1974). Intraclass Reliability Estimates: Testing Structural Assumptions. *Educational and Psychological Measurement*, 34(1), 25–33.
- West, R., & Leskovec, J. (2012). Human Wayfinding in Information Networks. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 619–628). New York, NY, USA: ACM.
- Wikipedia. (2015, Oct). *Partial least squares regression*. last access May 2nd, 2016. Retrieved from https://en.wikipedia.org/wiki/Partial_least_squares_regression
- Williams, J., & MacKinnon, D. P. (2008). Resampling and Distribution of the Product Methods for Testing Indirect Effects in Complex Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(1), 23–51.
- Wilson, T. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270.
- Winter, S. (2003). Route Adaptive Selection of Salient Features. In W. Kuhn, M. Worboys, & S. Timpf (Eds.), *Spatial Information Theory. Foundations of Geographic Information Science* (pp. 349–361). Berlin / Heidelberg: Springer.
- Winter, S., Bennett, R., Truelove, M., Rajabifard, A., Duckham, M., Kealy, A., & Leach, J. (2010). Spatially Enabling 'Place' Information. In R. Abbas, C. Joep, M. Kalantari, & B. Kok (Eds.), *Spatially Enabling Society, Research, Emerging Trends and Critical Assessment* (chap. 4). Belgium: euvn University Press.
- Winter, S., Raubal, M., & Nothegger, C. (2005). Focalizing Measures of Saliency for Wayfinding. In L. Meng, Z. A., & T. Reichenbacher (Eds.), *Map-based Mobile Services: Theories, Methods, and Design Implementations* (pp. 125–139). Springer Geosciences.
- Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, 14(3), 138–146.
- Wold, H. O. A. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin,

- R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307–357). New York: Academic Press.
- Wold, H. O. A. (1982). Soft Modelling: The Basic Design and Some Extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation. Causality, structure, prediction, part II* (pp. 1–54). Amsterdam: North-Holland.
- Wold, H. O. A. (1985). Partial Least Squares Regression. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (pp. 581–591). New York: John Wiley.
- Wold, S. (1993). Discussion: PLS in Chemical Practice. *Technometrics*, 35(2), 136–139.
- Wolfe, J. M., & Horowitz, J. M. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 1–7.
- Wolfensberger, M., & Richter, K.-F. (2015). A Mobile Application for a User-Generated Collection of Landmarks. In J. Gensel & M. Tomko (Eds.), *Web and Wireless Geographical Information Systems: 14th International Symposium, W2GIS 2015, Grenoble, France, May 21-22, 2015, Proceedings* (pp. 3–19). Cham: Springer International Publishing.
- Wu, W. W. (2010). Linking Bayesian networks and PLS path modeling for causal analysis. *Expert Systems with Applications: An International Journal*, 37(1), 134–139.
- Xia, J. C., Richter, K.-F., Winter, S., & Arnold, L. (2011). A Survey to Understand the Role of Landmarks for GPS Navigation. In *The 7th PATREC Research Forum [Planning and Transport Research Centre], PATREC Research Forum* (pp. 1–15).
- yWorks GmbH. (2016). *yEd 3.15.0.2*. Retrieved from <http://www.yWorks.com> (last access April 26th, 2016)
- Zha, H. (2002). Generic Summarization and Key Phrase Extraction Using Mutual Reinforcement Principle and Sentence Clustering.

References

- In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland* (pp. 113–120).
- Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian Framework for Saliency Using Natural Statistics. *Journal of Vision*, 8(7), 32.1-20.
- Zhao, Q., & Koch, C. (2013). Learning saliency-based visual attention: A review. *Signal Processing*, 93(6), 1401–1407.
- Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *The Journal of Consumer Research*, 37(2), 197–206.
- Zheng, Z., & Pavlou, P. A. (2010). Toward a Causal Interpretation from Observational Data: A New Bayesian Networks Method for Structural Models with Latent Variables. *Information Systems Research*, 21(2), 365–392.