



Luís Miguel Alves Fernandes

**Cultural Gestures and Virtual and Augmented
Reality to enable Non-Experts to Explore
Systemic Concepts and Complexity**

**Doctoral Program in Computer Science
of the Universities of Minho, Aveiro, and Porto**



July 2021



Luís Miguel Alves Fernandes

**Cultural Gestures and Virtual and Augmented
Reality to enable Non-Experts to Explore
Systemic Concepts and Complexity**

*Thesis submitted to the Faculty of Engineering of the University of Porto
for the Doctor Degree in Computer Science within the Joint Doctoral Program in
Computer Science of the Universities of Minho, Aveiro and Porto*



Computer Science Department
Faculty of Engineering of the University of Porto
July 2021

Supervisors

Professor Leonel Morgado

Associate Professor with Habilitation
Universidade Aberta

Professor António Coelho

Associate Professor with Habilitation
Faculty of Engineering, University of Porto

Professor Hugo Paredes

Associate Professor with Habilitation
University of Trás-os-Montes e Alto Douro

This work is co-financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project: UID/EEA/50014/2019, and by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 and the Lisboa2020 under the PORTUGAL 2020 Partnership Agreement, and through the Portuguese National Innovation Agency (ANI) as a part of project “CHIC: POCI-01-0247-FEDER-024498”. It is done in collaboration with USA Fulbright and IIE (Grantee ID E0607422) through the University of Montana, Salish Kootenai College, and Tech4Good, Montana, USA. Luís Miguel Alves Fernandes is supported by a FCT/MAPi PhD research grant (PD/BD/128141/2016).



REPÚBLICA
PORTUGUESA



I will treasure the knowledge like a squirrel treasures acorns.

Unknown

To my whole family, and girlfriend Mel

Author's declaration

I hereby declare that the work in this thesis was completed in accordance with the University's Regulations and Code of Practice for Research Degree Programmes, and that it has not been submitted for any other academic award. The work is the candidate's own work, unless otherwise indicated by a specific reference in the text. Work done in collaboration or with the assistance of others is labeled as such. Any opinions expressed in the dissertation are the author's.

SIGNED: 

DATE: 20-04-2022

Abstract

Complexity is a core part of our lives. People, whether aware of it or not, need to understand and communicate complex ideas and perspectives. Interactive simulations can help people understand and communicate complexity. Doing so in the physical world, on the other hand, is frequently impractical due to information overload and ambiguity, prohibitive costs, and safety concerns. Innovative immersive technology may hold the key to transforming how we approach understanding and communicating complexity. We propose empowering user agency and perception to participate in and create complex learning experiences by combining two factors: enhanced visual and spatial context provided by location-awareness, immersive environments, and somatic, embodied agency; and enhanced cultural and social context provided by leveraging the rich semantics of cultural-social gestures and rituals. To test the viability of this argument, we developed and now present a culture-aware prototype based on the culture of Montana's Confederated Salish and Kootenai Tribes (CSKT).

Keywords. Complexity · Complex Systems · Complex learning · Immersive environments · Emergence · Culture · Context · HCI · Gestures

Resumo

A complexidade é uma parte fulcral das nossas vidas. As pessoas, conscientes ou não dela, precisam de compreender e comunicar ideias e perspectivas complexas. As simulações interactivas podem ajudar as pessoas a compreender e a comunicar a complexidade. Fazê-lo no mundo físico, por outro lado, é frequentemente impraticável devido à sobrecarga de informação e ambiguidade, custos proibitivos, e preocupações de segurança. A tecnologia imersiva inovadora pode ter a chave para transformar a forma como abordamos a compreensão e a comunicação da complexidade. Propomos o potenciar da agência e percepção do utilizador para participar e criar experiências de aprendizagem complexas através da combinação de dois factores: contexto visual e espacial melhorado proporcionado pela percepção da localização, ambientes imersivos, e agência somática, encarnada; e contexto cultural e social melhorado proporcionado pelo aproveitamento da riqueza semântica de gestos e rituais culturais-sociais. Para testar a viabilidade deste argumento, desenvolvemos e apresentamos agora um protótipo com sensibilidade cultural baseado na cultura das Tribos Confederadas Salish e Kootenai (CSKT) de Montana.

Palavras-chave. Complexidade · Sistemas Complexos · Aprendizagem complexa · Ambientes imersivos · Emergência · Cultura · Contexto · HCI · Gestos

Acknowledgments

Fabricando fit faber - Practice makes perfect

I would like to start by expressing my gratitude to Professors Leonel Morgado, António Coelho, and Hugo Paredes for their guidance throughout my PhD. I am grateful for their wise counsel, their efforts to keep me motivated, and their guidelines to keep me on the right track. I am grateful to the MAP-i doctoral program, INESC TEC, FCT, and Fulbright for their encouragement and funding of this research. A special thanks to the University of Montana (UM), Salish Kootenai College (SKC), the Confederated Salish, Kootenai, and Pend d'Oreille Tribes (CSKT), Friedrich-Alexander University (FAU), and Bahcesehir University (BAU) for hosting me at various points throughout my research. If these institutions and their collaborators did not believe in the work, this thesis and its contributions would be impossible to achieve. I'd like to thank my family, and my girlfriend Mel, for always being there for me, sharing my joys and encouraging me when things got tough. Thank you to everyone who contributed in some way to this PhD.

To all of you, a big thanks!

Luís Miguel Alves Fernandes

FCUP, Porto

30th of July, 2021

Agradecimentos

Fabricando fit faber – É fazendo que se aprende

Gostaria de começar por expressar a minha gratidão aos Professores Leonel Morgado, António Coelho, e Hugo Paredes pela orientação ao longo do meu doutoramento. Estou grato pelos sábios conselhos, pelo esforço de me manterem motivado, e por me manterem no caminho certo. Estou grato ao programa de doutoramento MAP-i, INESC TEC, FCT, e Fulbright, pelo seu encorajamento e financiamento desta investigação. Um agradecimento especial à Universidade de Montana (UM), Colégio Salish Kootenai (SKC), às Tribos Confederadas Salish, Kootenai, e Pend d'Oreille (CSKT), Universidade Friedrich-Alexander (FAU), Universidade Bahcesehir (BAU), e à própria UTAD por me acolherem em diferentes pontos da minha investigação. Se estas instituições e os seus colaboradores não acreditassem neste trabalho, esta tese e as suas contribuições seriam impossíveis de realizar. Gostaria de agradecer à minha família, e à minha namorada Mel, por estarem sempre presentes, e partilharem as minhas alegrias e encorajando-me quando as coisas ficavam difíceis. Obrigado a todos os que contribuíram, de alguma forma, para este doutoramento.

A todos, um muito obrigado!

Luís Miguel Alves Fernandes

FCUP, Porto

30 de Julho, 2021

General Index

Abstract	xi
Resumo	xiii
Acknowledgments	xv
Agradecimentos	xvii
Table Index	xxv
Figure Index	xxix
1 Introduction	1
1.1 Context and Motivation	1
1.2 Main Contributions and Goals	2
1.3 Document Organisation	4
1.4 Bibliographic Note	6
1.5 Disclaimer	7
2 Systematic Literature Review	9
2.1 Introduction	9
2.2 Background	11
2.2.1 Complexity	11
2.2.2 Learning and Expressing Complexity	14

2.2.3	Learning in Context	19
2.2.4	Immersive Environments	24
2.2.5	Multimodal Human-Computer Interaction	28
2.3	Method	33
2.3.1	Research Questions	34
2.3.2	Search Process	35
2.3.3	Inclusion and exclusion criteria	38
2.3.4	Quality assessment	41
2.3.5	Data collection	42
2.3.6	Overview of search process	42
2.4	Thematic Analysis	46
2.4.1	[F6] Complex Topic being learn/taught	46
2.4.2	[F7] Technology for Visual and Spatial Context	48
2.4.3	[F8] Techniques to provide more visual and spatial context	51
2.4.4	[F9] Tech for cultural-aware interaction	57
2.4.5	[F10] Type of interaction	59
2.4.6	[F11] Techniques for additional Cultural Context	61
2.4.7	[F12] Complex learning	66
2.4.8	[F13] Target	73
2.5	Results	76
2.5.1	Quality evaluation	76
2.5.2	Thematic Analysis Results	78
2.6	Discussion	90
2.6.1	F6 - What is the complex topic being learnt/taught?	90
2.6.2	F7 - What technologies are used to provide visual and spatial context?	92
2.6.3	F8 - What techniques are used to provide more visual and spatial context?	94
2.6.4	F9 & F10 - What technologies are used to provide culture- aware interaction? And what types of interaction?	99
2.6.5	F11 - What methods/techniques are used to provide cultural context (and meaning)?	101
2.6.6	F12 - How is the understanding/expression of a complex topic approached versus traditional methods?	105
2.6.7	F13 - What are the main targets of these studies?	112
2.7	Conclusions	113
3	Software Architectures	115
3.1	Introduction	115
3.2	Design Principles	118

3.3	Architectural Styles	118
3.3.1	Benefits & Basic Properties	119
3.3.2	Common Styles	121
3.3.3	Enterprise	126
3.3.4	Mobile Oriented	127
4	Methodology	131
4.1	Introduction	131
4.2	Objectives	132
4.3	Design Science	135
4.4	Practical Application	138
5	CSKT Case Study	143
5.1	Introduction	143
5.2	Proposal	144
5.3	Luli's Journey Multimodal Analysis	145
5.3.1	Research Questions	146
5.3.2	Propositions	146
5.3.3	Units of Analysis	147
5.3.4	Evidence Collection	148
5.3.5	Propositions' Support	156
5.3.6	Conclusions	165
6	Multimodal HCI Architecture Specification	167
6.1	Introduction	167
6.2	Travelogue	168
6.3	Requirements Analysis	172
6.4	Architecture Design	176
7	InMERSE Architecture Implementation	181
7.1	Introduction	181
7.2	Inception	182
7.2.1	JavaScript Prototype	183
7.3	Design	188
7.3.1	Adapter Module	189
7.3.2	Decoder Module	191
7.3.3	Framework Module	192
7.4	Culmination	194
7.4.1	Asynchronous Programming	195
7.4.2	Communication	196

7.4.3	Design Patterns	196
8	CSKT Prototype	201
8.1	Introduction	201
8.2	Narrative	202
8.2.1	Grounding	203
8.3	VR Animated Concept	217
8.3.1	Environment and 3D Modelling	218
8.3.2	Dialogue	219
8.3.3	Interactions	223
8.4	VR Prototype	226
8.4.1	Foundation	226
8.4.2	Drawing Detection	229
8.4.3	Gesture Detection	236
9	User Testing	243
9.1	Introduction	243
9.2	User Sample Distribution	244
9.3	Pretest	245
9.4	Task Execution and Observation	252
9.5	Data Analysis	257
9.6	Main Findings	272
10	Final Considerations	277
10.1	Contributions and Limitations	277
10.2	Future Research Directions	279
A	SLR's Studies and Quality evaluation	281
A.1	Included Studies	281
A.2	Quality evaluation	289
B	CSKT Collected Evidence	291
B.1	Luli's Journey collected evidence	291
B.2	SKC IRB Protocol	307
B.3	Socio Demographic Pre-Questionnaire	315
B.4	Interview's Form of Consent	317
B.5	Interview's Questionnaire	319
B.5.1	Interview with Salish Student	321
B.5.2	Chat with Professor Frank Finley	331

C CSKT Narrative	333
C.1 Narrative Details	333
C.2 Narrative Script	334
D Machine Learning Detailed Models	345
E User Trials Data Compilation	347
E.1 Observational Compiled Data	347
E.1.1 Final Compilation	433
E.2 Data Analysis	435
Acronyms	445
Abbreviations	451
References	453

Table Index

2.1	Keywords, not including syntactic variants, for the Complexity Category– part of a more extensive search string (<i>qq.v.</i> Table 2.5).	14
2.2	Keywords, not including syntactic variants, for the Cultural Category– part of a more extensive search string (<i>qq.v.</i> Table 2.5).	24
2.3	Keywords, not including syntactic variants, for the Immersive En- vironments’ Category–part of a more extensive search string (<i>qq.v.</i> Table 2.5).	28
2.4	Keywords, not including syntactic variants, for the Interaction Category– part of a more extensive search string (<i>qq.v.</i> Table 2.5).	33
2.5	Categories and respective keywords. Categories with two asterisks appended (**) were not included in the automatic searches.	36
2.6	Databases and respective search scopes used in automatic searches. . .	37
2.7	Number of papers per publication type	43
2.8	Systematic Review’s Data Collection Form	45
2.9	Average quality score for studies by publication date.	77
2.10	Studies included in each F6 theme.	79
2.11	Studies included in each F7 theme.	79
2.12	Studies included in each F8 theme.	82
2.13	Studies included in each F9 theme.	84

2.14	Studies included in each F10 theme.	85
2.15	Studies included in each F11 theme.	88
2.16	Studies included in each F12 theme.	88
2.17	Studies included in each F13 theme.	90
4.1	Comparison between Positivism and Interpretativism in light of four key philosophical questions. Adapted from Figueiredo and Afonso (2006).	133
5.1	Propositions specification	147
5.2	Example of collected evidence for each unit of analysis, for a single dialogue and respective full-sized picture of the book Luli's Journey. .	150
5.3	Collected evidence for each unit of analysis, for the appendix on the storytellers of the past of the book Blue Jay brings the Chinook Wind.	151
5.4	Collected evidence for interview to Salish Student in CSKT – Ap- pendix B.5.1.	152
5.5	Collected evidence for interview to Salish Student in CSKT Part 2 – Appendix B.5.1.	153
5.6	Collected evidence for interview to Salish Student in CSKT Part 3– Appendix B.5.1.	154
5.7	Collected evidence for interview to Salish Student in CSKT Part 4– Appendix B.5.1.	155
8.1	Full narrative's script dissection.	203
9.1	General sample's description.	245
9.2	Compiled pre-form with corresponding number of occurrences.	248
9.3	Tasks and actions, and corresponding default gestures.	252
A.1	Studies included in final review.	281
A.2	Quality evaluation of included studies.	289
B.1	Collected evidence for each unit of analysis, for pages 4 and 5 of the book Luli's Journey.	292

B.2	Collected evidence for each unit of analysis, for pages 6 and 7 of the book Luli's Journey.	293
B.3	Collected evidence for each unit of analysis, for pages 8 and 9 of the book Luli's Journey.	294
B.4	Collected evidence for each unit of analysis, for pages 10 and 11 of the book Luli's Journey.	295
B.5	Collected evidence for each unit of analysis, for pages 12 and 13 of the book Luli's Journey.	296
B.6	Collected evidence for each unit of analysis, for pages 14 and 15 of the book Luli's Journey.	297
B.7	Collected evidence for each unit of analysis, for pages 16 and 17 of the book Luli's Journey.	298
B.8	Collected evidence for each unit of analysis, for pages 18 and 19 of the book Luli's Journey.	299
B.9	Collected evidence for each unit of analysis, for pages 20 and 21 of the book Luli's Journey.	300
B.10	Collected evidence for each unit of analysis, for pages 22 and 23 of the book Luli's Journey.	301
B.11	Collected evidence for each unit of analysis, for pages 24 and 25 of the book Luli's Journey.	302
B.12	Collected evidence for each unit of analysis, for pages 26 and 27 of the book Luli's Journey.	303
B.13	Collected evidence for each unit of analysis, for pages 28 and 29 of the book Luli's Journey.	304
B.14	Collected evidence for each unit of analysis, for pages 30 and 31 of the book Luli's Journey.	305
E.1	Observation compilation of User's one experience.	347
E.2	Observation compilation of User two's experience.	356
E.3	Observation compilation of User three's experience.	366
E.4	Observation compilation of User four's experience.	374
E.5	Observation compilation of User five's experience.	382
E.6	Observation compilation of User six's experience.	389

E.7 Observation compilation of User seven’s experience. 398

E.8 Observation compilation of User eight’s experience. 406

E.9 Observation compilation of User nine’s experience. 416

E.10 Observation compilation of User ten’s experience. 425

E.11 User Trials Final Compilation before its analysis. 433

Figure Index

2.1	Overview of search process	44
2.2	Quality Assessment Scores of primary filtered studies based on Mahdavi et.al.	77
2.3	F6 (Complex Topic) Distinct Themes' Totals	78
2.4	F7 (Technologies for Visual and Spatial Context) Themes' Distinct Studies' Totals)	80
2.5	F8 (Techniques for more Visual and Spatial Context) Themes' Distinct Studies' Totals	81
2.6	F9 (Techniques for Cultural-aware interaction) Themes' Distinct Studies' Totals	83
2.7	F10 (Interaction Types) Themes' Distinct Studies' Totals	84
2.8	F11 (Techniques for more Cultural Context) Themes' Distinct Studies' Totals	86
2.9	F12 (Complex Learning) Themes' Distinct Studies' Totals	87
2.10	F13 (Target Public) Themes' Distinct Studies' Totals	89
2.11	F6 (Complex Topics) General Clusters' Totals	91
2.12	F7 (Technologies for Visual and Spatial Context) Context Dimensions' Totals	93
2.13	F7's Context Dimensions Side by Side Percentages	94

2.14 F8 (Techniques for Visual and Spatial Context) Context Dimensions'	
Totals	95
2.15 F8's Context Dimensions side by side percentage	96
2.16 F8 (Techniques for Visual and Spatial Context) Type of Techniques'	
Totals	98
2.17 F10 (Interaction Types) Side by Side Percentages	100
2.18 F9 (Technology for Cultural-aware Interaction) Types of Technolo-	
gies' Totals	102
2.19 F11 (Techniques for Cultural Context) Types of Techniques' Totals .	103
2.20 F12 (Complex Learning) Learning Methods' Totals	106
2.21 F12 (Complex Learning) Practices and Strategies' Totals	107
2.22 F12's Complex Learning Aspects' Totals	109
2.23 F13 (Target) Target Groups' Totals	112
3.1 Layer architectural style example	122
3.2 Client-Server architectural style example	123
3.3 Blackboard architectural style example	124
3.4 Publish-Subscribe architectural style example	125
3.5 Event-based architectural style example	126
3.6 Peer-to-Peer architectural style example	127
4.1 Picture taken from Hevner (2007) . Design Science Research Cycles. .	136
4.2 Picture taken from Peffers et al. (2007) . Design science research pro-	
cess (DSRP) model.	137
4.3 HCI Mental map.	139
5.1 P1 HASS' proposition support	157
5.2 P1 Science, Medicine, and everyday life's proposition support	158
5.3 P2 HASS' proposition support	159
5.4 P2 Science, Medicine, and everyday life's proposition support	160
5.5 P3 HASS' proposition support	160
5.6 P3 Science, Medicine, and everyday life's proposition support	161
5.7 P4 HASS' part 1/4 of proposition support	161
5.8 P4 HASS' part 2/4 of proposition support	162

5.9	P4 HASS' part 3/4 of proposition support	163
5.10	P4 HASS' part 4/4 of proposition support	164
5.11	P4 Science, Medicine, and everyday life's proposition support	165
6.1	InMERSE logo and involved institutions	168
6.2	Digital Signage screenshot showing the reproduction of a swipe right gesture.	169
6.3	Original concept for the First Army of India with Player 1, the helmsman, on the left, and Player 2, Adamastor, on the right.	170
6.4	First Army experience. Photo by Quinto Canal at the Manga Comic Event (MCE) 2015, which took place on 22nd and 23rd of August of 2015, in Faro, Portugal.	171
6.5	High quality render of First Army of India's environment.	171
6.6	Overview of the InMERSE Framework.	176
6.7	Improved InMERSE Architecture.	177
7.1	InMERSE library setup example.	186
7.2	Leap Motion Earthquake Visualisation showcasing a two-handed interaction with a globe within a web application.	187
7.3	Leap Motion 360 degrees Interaction showcasing a full two-hand rotational interaction system with an object in VR.	188
7.4	Consumption Mode Class Diagram	194
7.5	Bootstrap Example Class Diagram	197
8.1	CSKT Prototype low poly environment.	219
8.2	2D animated concept without avatar.	220
8.3	CSKT Animated Concept with Narrator's Avatar.	221
8.4	Tvori's Puppetry showcasing strings attached to the puppet's head and elbows.	222
8.5	Bear rigging and animation.	224
8.6	Chick rigging and animation.	225
8.7	Coyote rigging and animation.	225
8.8	Goose rigging and animation.	225
8.9	Bird rigging and animation.	225

8.10	Example of preprocessed image.	231
8.11	Architecture of Drawing CNN Model.	232
8.12	Camera capturing drawing's layer only.	234
8.13	Drawing system's predictions for a laptop drawing.	234
8.14	Drawing system's predictions for a shorts drawing.	235
8.15	Drawing system in action within the VR Prototype.	235
8.16	Drawing system's predictions for a tree drawing.	236
8.17	Picture taken from Portalble (Qian et al., 2019). Each fingertip is represented by three features (Fx, Fy, Fz). Fz is a projection formed by a normal vector to the palm and the fingertip to the palm center ($NPalm$). The remaining two features are computed in the same manner, using a projection from the palm vector ($NPalm$) and each of the two local vectors PUp and $PRight$	237
8.18	Closed gesture.	238
8.19	One gesture.	238
8.20	V gesture.	239
8.21	Horn gesture.	239
8.22	Open gesture.	240
8.23	Gesture detection system's 10-fold cross-validation's results.	242
9.1	User Trials dominant hand	245
9.2	User Trials - vision issues	246
9.3	User Trials - vision specific issues	246
9.4	User Trials - Auditory issues	246
9.5	User Trials - Education level	247
9.6	User Trials - Prior experience with immersive technology	247
9.7	User Trials - used Immersive technology before	247
9.8	User Trials - last time since the use of immersive HMDs	248
9.9	Action A1 - Look around	258
9.10	Action A2 - Wave for Hello	259
9.11	Task T1 - Grab Microphone	260
9.12	Task T1.1 - Talk	260
9.13	Action 3 - Give Microphone	261

9.14	Task T2 - Look at V-shaped peak	261
9.15	Task T3 - Look back at narrator	262
9.16	Task T4.1 - Interact with pictures	263
9.17	Task T5.1 - Interact with pictures	263
9.18	Task T6.1 - Resize Luli's book	264
9.19	Task T7 - Hold brush	265
9.20	Task T7.1 - Draw bird	265
9.21	Task T8 - Make Luli grow	266
9.22	Task T9 - Adjust Map	267
9.23	Task T10 - Make Luli fly	267
9.24	Task T10.1 - Look at the Geese in the sky	268
9.25	Task T11 - Look at light leading to the sky	268
9.26	Task T11.1 - Look at Geese flying back	269
9.27	Tasks T12 - Yes/No Answer for more information on tribes	270
9.28	Tasks T13 and T14 - Yes/No Answer	270
9.29	Action 19 - Exploration phase with 'change'	271
9.30	Action 20 - Goodbye Wave	272
9.31	Users' computer experience	273
9.32	Users trials observation codes' total count	274
9.33	User trials observation code's percentages	275
B.1	Picture from Luli's Journey book that lead to chatting with Professor Frank Finley.	332
D.1	Detailed Drawing (left), and Gesture (right), Machine Learning Model Views using Netron	346
E.1	Task T4 - Look at reservation pictures	435
E.2	Action A4 - Look at car and bear demonstration	436
E.3	Task T5 - Look at art pictures	436
E.4	Action A5 - Look at coyote	437
E.5	Look at Luli's book cover	437
E.6	Action A6 - Look at Luli's book cover	438
E.7	Action A7 - Look at Luli at the beach	438

E.8	Action A8 - Look at Geese and bear in lake	439
E.9	Action A9 - Look at the sun and planets models	439
E.10	Action A10 - Look at the earth globe	440
E.11	Action A11 - Look at the animals disappearing (hibernating)	440
E.12	Action A12 - Look at the mountains snow coming down like a blanket	441
E.13	Action A13 - Look at Geese flying at different speeds	441
E.14	Action A14 - Look at goose going out of formation and recovering . .	442
E.15	Action A15 - Look at the Geese formation's leader rotating back when it gets tired	442
E.16	Action A16 - Look at the Geese flying and motivating the ones in front to continue	443
E.17	Action A17 - Look at a goose falling and two others staying with him until he recovers	443
E.18	Action A18 - Look at the Geese joining back the same flock	444



Introduction

1.1 Context and Motivation

In general, one can say that there are recurring accessibility issues, particularly in the areas of **Human-Computer Interaction (HCI)** and inclusion. People, as users, must constantly learn how to interact with newer devices, which, due to the field's rapid evolution, begins to be accepted as the norm. This may be related to Moore's Law, which states that "the number of transistors on a microchip doubles about every two years, even though the cost of computers is halved". Despite the fact that possible improvements are already having an impact on society, a number of obstacles remain, and there is still a long way to go. One of the major positive aspects of technological evolution that is of particular interest for this thesis is the ability to simulate and/or forecast certain events or situations that would otherwise be impossible (*e.g.*, disasters), which is becoming increasingly transversal to various areas (*e.g.*, health, industry). People often use oversimplified models and abstractions to learn emergent properties when addressing these types of topics, which can lead to errors (*e.g.*, false cause-effect relationships ([Chi et al., 2012](#))). The study of complex systems, or the study thereof, may hold the key to our understanding in this regard. The reductionist paradigm is being replaced by a more holistic approach in

complexity science. As a result, the process of learning about complexity, known in the literature as Complex Learning (Ifenthaler and Eseryel, 2013; van Merriënboer et al., 2002), presents unique challenges. The ability to communicate an idea and its understanding is considered a challenge from the moment people gain insights into a given complex system. The use of visualisations (*e.g.*, Three-dimensional (3D) models in Two-dimensional (2D) screens), which augmented, mixed, and virtual reality (*i.e.*, immersive environments) in particular can leverage, is one possible approach to solving the already known understanding problem. The visualisations are especially effective when made interactive by utilising computational modelling tools such as simulators. Authoring with modelling software, where both its analysis and design involve the creation and definition of rules and behaviours—which, despite going by many names (*e.g.*, scripting, configuring, setting up), ultimately implies programming by the end user, who is often technically limited to do so because they have no background in the field. To make a computer program (or script or specification) work, a person must convey concepts and intentions in accurate and precise ways, implying that these individuals must become aware of flaws in the exactness and clarity of their conceptions—that is, the lack thereof. This entails giving concepts context, connecting abstractions or diffuse ideas to the practical implications of reality, or, as theoretical educator Seymour Papert once put it, submitting ideas for testing in reality (Papert, 1999). An approach that appears to be promising (Morgado et al., 2015b; Alves Fernandes et al., 2019), with promising early results (Carvalho, 2019), is based on the idea of leveraging the multidimensionality of context through cultural constructions as a means of providing context for concepts, connecting abstractions with the concrete (Holbert and Wilensky, 2014), and, in particular, using their own cultural expressions such as gestures and body movements, voice, and object manipulation, in conjunction with immersive environments.

1.2 Main Contributions and Goals

This thesis is part of an ongoing research effort (Morgado et al., 2015b; Alves Fernandes et al., 2019) in HCI and immersive environments with the overarching goal

of empowering non-experts to understand and communicate systemic concepts and ideas. Philosophical arguments contend that, in comparison to more traditional interfaces, people will be able to use a broader range and depth of richer interactions with cultural gestures. An earlier paper (Pinto et al., 2015) proposed a method for determining this by analysing how people behave in a videogame or other immersive systems. Subsequently, Carvalho (Carvalho, 2019) found that users employing cultural gestures for interaction have a higher recall rate and fewer errors than users employing non-cultural gestures, after a 15-day time gap. This in spite of both cohorts of users expressing equal satisfaction, *i.e.*, regardless of having employed cultural or non-cultural gestures.

Based on these three points, this thesis explores the conjecture that combining context and cultural gestures for interaction with immersive environments, will leverage richer interactions and support better learning of complex concepts and systems. This exploration of the conjecture pursues an innovative combination of two factors that can make a significant contribution: (1) enhanced visual and spatial context provided by location-awareness, immersive environments, and somatic, embodied agency; and (2) enhanced cultural and social context by leveraging the rich semantics of cultural-social gestures and rituals as input methods. There was no pre-existing method for approaching or achieving this, and this thesis provides clarity in this regard.

A systematic review of the literature on how to teach and learn about complex concepts is presented, emphasising the importance of context’s multidimensionality first. This significance demonstrates that context and gestures are not being used effectively for this problem, at least not yet, so it was impossible to tell if it was even a relevant problem.

We present a set of complex concepts based on the children’s book Luli’s Journey after conducting ethnographic research on the Flathead Indian Reservation (Confederated Salish and Kootenai College (CSKT)). We iteratively design a Virtual Reality (VR) concept and prototype based on this analysis, leveraging interaction through an architecture (InMERSE) that we also improve in an attempt to free end-user

applications from specific input devices and modalities. We first use this final combination to see if the prototype complies with the complex learning literature and its guidelines. We observe, record, and register all of the interactions that users have, and then analyse them to learn what they explore and what they don't. What commands/gestures they use, why they use them, if they repeat them throughout the experience, and so on. Then we have a viable, tried-and-true method of reiterating the artefact (the prototype) with a new, improved, default set of gestures that should make more sense in Portuguese culture.

To summarise, this thesis sets forth from a cultural exemplar (a child's tale) and provides (1) a process for identification and extraction of complex concepts, (2) an improved version of the InMERSE software architecture to enable device-independent gesture interaction, (3) a method for designing immersive interactive environments for exploring complex concepts associated with a cultural exemplar, and (4) a method for creating cultural gestures for interaction with the environment. This manuscript represents the culmination of the entire process, not only of this doctoral program, but also of ideas that our research group has been developing for years. It isn't better or worse than other approaches; it's just the first step toward something new. After all, what is science if not a selfless, naive attempt to better society?

Finally, the goal is to present a model and a sufficiently robust proof of concept, without becoming a dogma, to contribute to support for the resolution of the aforementioned challenges. The research is beginning to be disseminated to the scientific community via articles submitted to conferences and events, as well as journals.

1.3 Document Organisation

Complex systems are, as the name implies, complex. Given the various fields involved in this particular work, ranging from grounding in the literature to the conceptualisation and investigation of how the multidimensionality of context can aid in the learning and understanding of complex systems and concepts, this thesis is

structured to present all of this process in the clearest and logical manner possible, as further discriminated:

- **Chapter 2 - Systematic Literature Review:**

Presentation of a broad systematic literature review on complexity, learning and expressing complexity, learning in context, immersive environments, and human-computer interaction, including a thematic analysis of the included studies.

- **Chapter 3 - Software Architectures:**

Presentation of the overall architectural concept and its adaptation to the area of computing, as well as the considered principles and requirements in the conceptualisation and design of a software architecture.

- **Chapter 4 - Methodology:**

Presentation of the adopted philosophical paradigms, as well as the metamethod used to guide and develop this work, and how it is effectively used throughout.

- **Chapter 5 - CSKT Case Study:**

Presentation of the formal case study, titled CSKT, including the multimodal analysis of collected data, such as from books, and interviews.

- **Chapter 6 - Multimodal HCI Architecture Specification:**

Presentation of the work already done by the research group, as building blocks of an improved architecture, its requirements, and subsequent design.

- **Chapter 7 - InMERSE Architecture Implementation:**

Presentation of implementation of a proof of concept and improved version of the original InMERSE Framework, based on the conceptualised architecture and technologies available for that purpose.

- **Chapter 8 - CSKT Prototype:**

Presentation of the developed CSKT prototypes, with their foundation on an established and grounded narrative, including a (VR) animated concept, and a VR prototype application.

- **Chapter 9 - User Testing:**

Presentation of the conducted user tests, as well as their analysis, and main findings.

- **Chapter 10 - Final Considerations:**

Presentation of the key points of this thesis, as well as encouragement of future work.

1.4 Bibliographic Note

This thesis includes some work that has already been published and/or submitted. The following list provides the references to those publications:

- [Immersive Learning Experiences for Understanding Complex Systems](#), June 2019;
- Immersive environments that leverage context and cultural interaction to increase the understanding and expression of complexity - A systematic literature review – submitted for peer review to ACM Computing Surveys (CSUR) on March 2021.

This work, however, would not be possible without the foundation of previous works, which were also developed by our research group, namely:

- [Separating Gesture Detection and Application Control Concerns with a Multimodal Architecture](#), October 2015;
- [Bringing User Experience empirical data to gesture-control and somatic interaction in virtual reality videogames: an Exploratory Study with a multimodal interaction prototype](#), November 2015;
- [Primeira Armada da Índia: novo conceito de jogo misturando realidades aumentada e virtual, gestos finos e amplos](#), November 2015;

- [Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype](#), October 2016.

1.5 Disclaimer

This work does not represent the CSKT tribes in any way. Research was conducted in the reservation, but further approval and validation by the respective Cultural Committees and User trials are still needed.

2

Systematic Literature Review

2.1 Introduction

Stephen Hawking once said he thought the 21st century would be the century of complexity (Chui, 2000). Undeniably, the world is an extremely complex place, comprised of complex and interconnected systems, creating novel challenges that call for a new way of thinking. Ecosystems, societies, weather, and the human body are examples of such complex systems. They are complex in the sense that they act in interdependent and unpredictable ways. The concepts of complexity and system are traditionally difficult to understand by non-topic professionals (non-specialists). Specifically, aspects such as a system consisting of several layered or network sub-systems; the existence of communication processes between the system and the environment that surrounds it; the control process; and especially the notion that a system has emergent properties, observable when looking at the system as a whole, but absent from their individual components (Checkland, 2012), causing uncertainty to be inherent. Examples of that include the behaviours of water (*e.g.*, surface tension), or a bird flock (*e.g.*, flight formation). Generally, people resort to oversimplified models and abstractions for the learning of emergent properties, which often can induce people into error, *e.g.*, false cause-effect relationships (Chi

et al., 2012). Complex systems, or the study thereof, may hold the key to our understanding in this regard. Complexity science replaces the reductionist paradigm with a more holistic approach. Thus, the process of learning about complexity, which the literature traditionally calls Complex Learning (Ifenthaler and Eseryel, 2013; van Merriënboer et al., 2002), poses specific challenges. From the moment people reach insights of a given complex system, the ability to communicate this idea and its understanding is also considered a challenge. A possible approach to solving the understanding problem that is already known is the use of visualisations (*e.g.*, 3D models in 2D screens), something that augmented, mixed, and virtual reality (*i.e.*, immersive environments) in particular, can leverage. The visualisations are particularly effective when made interactive, taking advantage of computational modelling tools, such as simulators. A possible approach to solving the expression problem is authoring with modelling software, where both its analysis and its design involve the creation and definition of rules and behaviours—which despite going by many names (*e.g.*, scripting, configuring, setting up), ultimately implies the programming by the end user who is often limited technically to do so, as they have no background in the field. For a computer program (or script or specification) to work, a person has to convey concepts and intentions in accurate and precise ways, thus implying that these need to become aware of shortcomings in the exactness and clarity of their conceptions—that is, the lack thereof. This involves providing context for concepts, linking abstractions or diffuse ideas with the practical implications of reality or, as the theoretical educator Seymour Papert once put it, subjecting ideas to the test of reality (Papert, 1999).

An approach that might be promising (Alves Fernandes et al., 2019; Morgado et al., 2015b), with favourable early results (Carvalho, 2019), is grounded on the idea of leveraging the multi-dimensionality of context through cultural constructions as the means to provide context for the concepts, for connecting abstractions with the concrete (Holbert and Wilensky, 2014) and, in particular, use their own cultural expressions such as gestures and body movements, voice and manipulation of objects, in conjunction with immersive environments. The expectation is that both exerting agency and perception within an enhanced visual-spatial context, exploring

the interconnected concepts and data, together with enhanced socio-cultural context of input methods, support more powerful and diverse semantics for that agency, thus contributing to a deeper perception and therefore ease the understanding and the communication of complex concepts and ideas. This should particularly ease the process for non-topic professionals (non-specialists) and technology novices, through the creation of representations of those ideas that are dynamic experiences in their own right, which can then be explored by third-parties. By performing a systematic literature review, we can appraise how much work has leveraged this approach, in a rigorous and reliable way. Given the subjectivity and qualitative nature of the collected data for all the included relevant fields, we performed a thematic analysis to help us understand it further, followed by its scrutinisation. The resulting panorama revealed important background knowledge gaps across several fields of research, thus contributing to the establishment of future research recommendations.

2.2 Background

2.2.1 Complexity

Complexity has several definitions proposed by authors with different backgrounds, objectives, and research perspectives, as one would expect from its interdisciplinarity. It is therefore difficult to find a strict definition that applies to all domains—inherently subjective, due to the difficulty of its quantification (*e.g.*, number of variables, connections). It may even be self-defeating to pin down its definition, Cilliers argues (Cilliers, 2013), as there is something inherently reductionist about it, whereas complexity calls for a more holistic view instead. However, in order to establish the scope and criteria for a systematic literature review that tackles it, one will approach its main properties and dynamics for grounding towards definitions of scope and criteria.

Complexity comes in various shapes and forms, and has a myriad of possible applications across the sciences, medicine, and everyday life (*e.g.*, traffic jams, market crashes) (Johnson, 2009, p. 10). It is often tied to science alone, particularly to the

fields of physics and mathematics, but science is not enough to speak and explain the richness of life, agency, value, meaning, doing, and understanding (Kauffman, 2008, pp. 15, 21, 340)—all of which emerged through the evolution of the biosphere. Not everything can be reduced to physics alone (*e.g.*, biology). There is more to it; Humanities, Arts, and Social Sciences (HASS) are such examples. As Kauffman pointed out, “we live in a universe, biosphere, and human culture that are not only emergent but radically creative” (Kauffman, 2008, p. 25), and “story, including poetry, music, and other arts, is how we know ourselves” (Kauffman, 2008, p. 338). Science, as a human activity, has “its similarities to the arts, law, business, medicine, crafts, pastoring, all domains of diverse human creativity and actions in the face of uncertainty” (Kauffman, 2008, p. 341).

The interdisciplinarity of complexity, as suggested above, calls for a more holistic view (base of systems thinking, systems theory, and complex thought (Morin, 2011)) and equally interdisciplinary studies. One notorious example of this is in medicine, which has been increasingly applying the breadth of complexity theory to health care, health care reform, and the future of the field by bringing holism (*e.g.*, taking into account the mental and social factors, rather than just the symptoms) and by leveraging the context (*e.g.*, the role of the family and of their interactions with the patient) and the cultural background of the patients to better understand their diseases and therapeutic responses (Sturmberg et al., 2014). Apparently dissimilar systems can share important commonalities both in structure and behaviour, so it should be possible to find universal or more fundamental laws by which their properties are governed (Kwapień and Drożdż, 2012). Emergence is often described, originally by Aristotle, as “the whole is something besides the parts” (Aristotle, y BC): the collective phenomena at a higher level are not the simple sum of the individual behaviours or characteristics of the constituent elements—a composite whole (Doolittle, 2014). This is also called synergy (though it is a more general term), such as the division of labour of the ants or sounds’ destructive interference. Complexity is associated with the intertwining and inter-connectivity of the constituent elements of a system and the surrounding environment (Fischer et al., 2012; Kwapień and Drożdż, 2012). By extension, the definitions of complex systems (Doolittle, 2014;

Fischer et al., 2012; Ifenthaler and Eseryel, 2013; Kwapień and Drożdż, 2012; Steenbeek and van geert, 2013; Sturmberg et al., 2014), typically based on ideas from authors such as Holland, Funke, and Dorne, agree that a large number of elements, interacting at various levels, result in the emergence of a new level of organisation: self-organisation. Thus, there is a hierarchical and asymmetrical structure, though not neatly nested as pointed out by Cilliers (2013), with all these levels affecting each other.

They are interconnected and interdependent on each other, and it is not possible to completely isolate a whole system or reduce the whole thing to one level without losing the capabilities that make it a complex system to begin with (Cilliers, 1998, p. 80). The very emergence of order out of apparently chaotic or noisy behaviour, an idea first proposed by von Foerster (2003), is a manifestation of complexity. The coexistence of order and noise and their interweaving appearance is the complexity. A more general notion of complex systems can then be seen as a specific type/class of system with properties that lead to emergence, namely: (a) multiple constituent elements, relations, and levels of hierarchy; (b) exhibiting non-linearity and phase transitions, that is: joining two things does not necessarily mean that the result is simply the sum of the properties of the individual elements, and occasions where a minimal change of a value, through feedback loops, can trigger large systemic effects—Self-organised Criticality (SOC) (Kwapień and Drożdż, 2012); (c) connectivity, as a driver of complexity, meaning the system is topologically a network (*cf.*, Connectionism—(Cilliers, 1998, p. 25)); and (d) autonomy, variability, and adaptation, as enablers of self-organisation and evolution.

This notion and understanding of complex systems is important to acquire, not only for the process of complex learning, as argued throughout the next section (*q.v.* Section 2.2.2 - Learning and Expressing Complexity), but also to identify and label complex topics in this systematic review through the use of certain keywords that may point towards a study being about complexity (even if not referred as such), *q.v.* Table 2.1, below.

Table 2.1 – Keywords, not including syntactic variants, for the Complexity Category—part of a more extensive [search string](#) (*qq.v.* Table 2.5).

Complexity Keywords						
complex	complexity	non-linear	emergence	systemic	self-organisation	systems theory

2.2.2 Learning and Expressing Complexity

Literature suggests that today’s education remains, in general, linear and mechanised (Doolittle, 2014; Figueiredo, 2005; Figueiredo and Afonso, 2006), and there is room to adopt its true nature, that is, non-linear, adaptive and constructive (Doolittle, 2014)—indeed, learning is itself a complex process. Therefore, learning about complexity is the application of a complex process (learning) to a complex concept (complexity). The resulting hypercomplex (Morin, 2011, p. 37) system is thus a challenging and distinctive social cognitive process (Ifenthaler and Eseryel, 2013). This nature has traditionally been expressed in the literature as “complex learning” (Ifenthaler and Eseryel, 2013; van Merriënboer et al., 2002), despite its ring of pleonasm. Visualisation has been one of the most utilised methods to deal with complexity in the past, often accomplished by showing 3D models in 2D displays (Drew and Hendley, 1995; Hendley and Drew, 1995)—which was referred to, back then, as virtual reality, whereas nowadays, however, the use of the meaning of virtual reality has shifted to the use of Virtual Reality headsets—something a few researchers argue against (Han, 2020). Immersive environments (*q.v.* Section 2.2.4 - [Immersive Environments](#)) and technology have the ability to reach further, with a whole different level of virtual reality, by making these visualisations come alive and fully immerse the user in the data, offering new perspectives, and modes of interaction, which may result in positive learning outcomes (Bock et al., 2018; Laha et al., 2012; Sicat et al., 2019).

Little wonder then that teaching complex concepts is challenging for teachers, communicating complex concepts and ideas is challenging across intercultural contexts, and even societal clashes may arise from lack of common understanding. Several efforts have been carried out attempting to understand and explain the phenomena.

One particular aspect is that often teachers cannot put themselves in the cognitive status of the novice student, no longer remembering the difficulties themselves faced when learning the same topic. This is called the curse of knowledge bias (Wiemann, 2007) or hindsight bias (Fischhoff, 1975). In this regard, Ifenthaler and Eseryel (2013) argues that effective learning environments for complex knowledge domains (*e.g.*, Science, Technology, Engineering, Mathematics (STEM)–and HASS) should target three complex learning attributes: (a) understanding complex systems (*q.v.* Section 2.2.1 - Complexity); (b) developing adaptive expertise; and (c) acquiring soft skills such as collaboration (*e.g.*, peer-learning to tackle the aforementioned curse–(Crouch and Mazur, 2001)), communication, and task coordination (*e.g.*, active learning and its sub methods such as discovery, Problem-based Learning (PBL), project-based learning–(Prince and Felder, 2006)). It is then not surprising that there are works specifically focusing on this field, from learning theories that explain the phenomenon like Complex Constructivism (Doolittle, 2014) and Connectivism (Siemens, 2005), tools and strategies for its practical application such as the Four-Component Instructional Design (4C/ID) model (Van Merriënboer and Kester, 2014) or the Complex Problem Solving (CPS) (Fischer et al., 2012), to papers that analyse particular aspects such as the aforementioned hindsight bias or the importance of feedback and repetition (Argyris and Schön, 1974; Stermann, 1994).

It may be argued that both Complex Constructivism and Connectivism are based on Constructivism, applying the principles of complexity and self-organisation theories, and thus adopting the non-linear and adaptive nature of learning. Connectivism was conceived for the present times, for the “digital age”, the Internet and the chaotic flow of information, therefore also integrating the principles of chaos theory. Although both learning theories defend the impact and importance of context (*e.g.*, social and cultural) in learning, as well as of the established connections and contacts (nodes or sources of information in Connectivism’s terminology), in Connectivism, learning is accomplished through a digital context and the flow of information (*i.e.*, the person-network-organisation cycle *cf.*, (Siemens, 2005)), suggesting learners’ involvement in networks and Internet use. In Complex Constructivism, however, learning is a process between the interaction of the student and his/her

mental models, combining the social and radical views of constructivism, and emphasising the active, self-organising construction of knowledge (Prince and Felder, 2006), and its adaptive nature (Doolittle, 2014).

Also with a constructivist basis, the CPS perspective applies problem-solving techniques to tackle complex learning. Similarly, but addressing the teaching of complex learning, the 4C/ID model is also—in some way—problem-based, using a mix of (social) Constructivist and Instructivist approaches. Despite being different, both methods share some ideas, particularly the importance of schema construction for complex learning. The CPS process consists of two phases, namely knowledge acquisition (that by itself goes through five stages—(Dreyfus et al., 2000, p. 20)) and goal-oriented knowledge application (Fischer et al., 2012), and can be subdivided further. The 4C/ID model (van Merriënboer et al., 2002), however, argues that four interrelated components are essential to complex learning, while providing instructional methods for each of them. The parts of the CPS process that might, to some extent, have a correspondence in the 4C/ID model will be indicated in bold. The objective of the first two components of the model (1st phase in CPS: **knowledge acquisition**), learning tasks and supportive information, is to promote the construction of schemata—which in turn contributes to developing adaptive expertise (one of the challenges/requirements of complex learning (Ifenthaler and Eseryel, 2013)). In spite of not being explicit, the model begins by acknowledging the need to understand the problem (calling for **information retrieval**), another attribute of complex learning (Ifenthaler and Eseryel, 2013).

Only then it is possible to provide the learners, through scaffolding (*e.g.*, guidance and support) and **Zones of Proximal Development (ZPD)**—essential to attain learning otherwise unachievable as argued originally by Vygotsky (Obukhova and Korepanova, 2009; Vygotsky, 1980, p. 86)—with a series of (adaptive) learning tasks (**goal elaboration** and **goal balancing**). Ideally, it should confront the students with non-recurrent skills that constitute the whole complex skill (calling for task coordination and integration—also an important attribute to complex learning (Ifenthaler and Eseryel, 2013)). The second component of the model, supportive information, is about connecting newly presented information to existing mental models

(**information integration**) in order to reach the proposed objectives or tasks (2nd phase in **CPS: goal-oriented knowledge application**). These internal representations, or mental models, as well as impressions, intuitions, and many decisions, are silently created and integrated with the context in our mind (Sterman, 1994; Kahneman, 2011, p. 4) at the level of sensory perceptions (Heald et al., 2018; Körding and Wolpert, 2006), constantly fed by sensory cues or stimuli, including visual, auditory, haptic, olfactory, and environmental cues (Goldstein, 2010). The aim is to interpret almost instantaneously what surrounds us, an ability improved throughout the evolution of the species through survival instinct. Despite happening mostly unconsciously, this perceptual process, comprised of four processes (stimulus, electricity, experience and action, and knowledge, *cf.*, Goldstein (2010)) can also be attended through what is known as attentional shift, or shift of attention (Posner, 1980). This makes it possible for (complex) mental models to also leverage reflective expertise, that is, monitoring and assessing the progress of solving a task (**action planning** and **decision making**) and own performance (**self management**), as well as the implications of the received feedback which is essential for learning (*cf.*, Argyris and Schön (1974)).

These proposals hinge on planning and providing circumstances and processes within the context of learning environments in support of complex learning, where the learner interacts (exerts agency) and perceives the outcome. The underlying perspective is that our personal mental models, states, constructs, or entities (depending on the theoretical lens on knowledge) change, not only by self-reflection and inner interactions over time, but by interaction with the physical world. Our agency and perception in this interaction are heavily influenced by prior knowledge and world-views (dependent on social interactions and culture). The outcome then impacts subsequent agency and perception. However, carrying out some learning experiences in the physical world is often infeasible, not only due to monetary issues or the risk they may entail (*e.g.*, simulation of catastrophes/accidents), but also because of their dynamic complexity, the presence of confounding and ambiguous variables, and the fact that feedback is often misinterpreted or surfaces with a time delay. If this occurs, the number of iterations of these cycles decreases and consequently slows the entire learning process, including the ability to accumulate experience,

test hypotheses and improve (Serman, 1994). Back in the 80s, Papert and Seymour (1980) and Schön (1983) suggested a way they deem feasible to carry out such experiments, through simulation spaces or virtual worlds (also later used by Serman (1994, 2006)). Other terms might include serious games (gamification) and edutainment, which when particularly coupled with semantics, might take its realism to a whole new level and impact the learning aspect of the game experience positively (Repetto and Catalano, 2015). According to these perspectives, the main benefits of using immersive learning experiences (*i.e.*, virtual, augmented, mixed and cross reality, in present terminology) for learning purposes, and more specifically in fields of complex knowledge, are: (a) the removal of everything that at first sight should be of interest to the experience, isolating it the most; and (b) having a greater control over the variables, so that the decision maker can obtain and process complete, accurate and immediate feedback. An example of this isolation and control is that of turbulence systems (*e.g.*, chaotic fluids), which have long been observed to follow a universal rule known as Batchelor’s law but only recently has been proved mathematically (Bedrossian et al., 2020). Another such example was presented by Ifenthaler and Eseryel (2013), and combines the benefits of mobile learning, Virtual Learning environments (VLEs) and Augmented Reality (AR). As he further argues (Ifenthaler and Eseryel, 2013), AR (and Mixed Reality (MR)) open up new opportunities for both supporting and facilitating complex learning.

Unlike complexity and other categories relevant to the research questions (*q.v.* Section 2.3 - Method’s subsections), learning by itself tends to be more general and subjective to search for through keywords and can be assessed more effectively through manual screening instead (*i.e.*, inclusion and exclusion criteria—*q.v.* subsection 2.3.3). This background section on learning and expressing complexity is still particularly important, however, for the processes of data collection (*q.v.* subsection 2.3.5) and data processing, namely to define codes and themes (*q.v.* Section 2.4 - Thematic Analysis) that are grounded on previous literature.

2.2.3 Learning in Context

Learning can be seen as the change of our view of the world (*q.v.* learning loops in Section 2.2.2 - [Learning and Expressing Complexity](#)), highly influenced by the surrounding environment, particularly by social interactions and culture ([Sterman, 1994, 2006](#)), or, in short, shaped by context. This influence is often overlooked or frequently understood as trivial or shared common ground ([Roque et al., 2004](#)). It is also considered one of the challenges of complex learning (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)) due to its situativity ([Greeno and Moore, 1993](#)) in real life contexts ([Ifenthaler and Eseryel, 2013](#)). VLEs should therefore simulate its practices by leveraging the context of the surroundings (*e.g.*, physical, social, cultural)–becoming situated learning ([Lave and Wenger, 1991](#)) environments, that is, unintentional and embedded within activity, context and culture. A dimension of context often forgotten, however, is the mind. The contextual-binding theory ([Yonelinas et al., 2019](#)) argues that the mind state, that also comprises the thoughts, emotions (also proven to be correlated with learning, *cf.*, [D’Mello and Graesser \(2011\)](#); [Xu and Woodruff \(2017\)](#)), and other mental activity experienced at a given moment, is associated with memories along with the other context dimensions (*e.g.*, physical location, part of spatial context) ([Norman, 2013](#), p. 47). As Godden and Baddeley argue ([Godden and Baddeley, 1975](#)), “recall is better if the environment of original learning is reinstated”, and situated learning idealises just that.

Situated learning environments are, however, more than just location-based services, despite often being portrayed as such. Oftentimes, location-based services are also labelled as context-aware, which comes to show an incomplete vision of context; one that ignores its multi-dimensionality altogether. But the root of this issue goes deeper than just agreeing on a definition, or its main properties and entities; the problem seems to lie on how pedagogical practice is approached nowadays.

Prior to the industrial revolution, and contrary to what happens now, learning was mainly done in context. Knowledge was acquired, often by necessity (just-in-time), through real-world experiences ([Figueiredo and Afonso, 2006](#)). Education, through

its massification, has become mechanised (just-in-case), where students of the same age learn the same content through books and teachers, and often without visible application (Figueiredo and Afonso, 2006). It has even (mostly) replaced storytelling, a long-established tradition started 100.000 years ago of maintaining cultural heritage and transferring knowledge from generation to generation, with PowerPoint. When the Human species started developing language, this technique was used through a more or less established narrative (*i.e.*, actual content of the story, including time, place, objects and characters). Other significant features of storytelling include perspective, interaction, and medium (Lugmayr et al., 2017). One might recall the illustrious movie “Dead Poets Society”, where Professor Keating, performed by Robin Williams, stands on his desk and asks his students why. His answer is “just when you think you know something, you have to look at it in another way”. Perspective reminds us of what is important and what is not. Chances are that the perspective of the one teaching and the one learning are different, and it is critical to human learning to match them (Berlin et al., 2006). Gaining perspective and context is something that grows with experience, and immersive environments have the potential of accelerating that process. Despite the obvious importance of context in learning (Figueiredo and Afonso, 2006; Roque et al., 2004; Schön, 1983; Sterman, 2006), it has not been explicitly orienting pedagogical practice, however.

Finding a single, formal definition of context is by itself an arduous task, particularly because of its specificity and multi-dimensionality that varies from domain to domain (Bauer and Dey, 2016; Malinowski, 1967). It also depends on the adopted paradigm (Figueiredo and Afonso, 2006). For example, given the positivist current, the context is external, delimited, totally independent, stable, and predictable. In contrast (Figueiredo and Afonso, 2006), from a constructivist point of view, more recommended for socially complex contexts, the context cannot be delimited nor localised (fuzzy), and is perceived through its interactions with the student—“context is the interactions”. One of the most accepted and used definitions today is that of Dey et al. (2001), although there are several others, created over time as the need to include one or more variables arose. Combining it with that of Henricksen (2003), thus enriching it with the information of the action that the user (agent) intends to

perform, which allows the system to be anticipatory, one obtains the following definition: “*any type of information and relevant circumstances, which can be used to represent the action that a particular entity intends to carry out and the restrictions it undergoes at a given moment, including the relevant previous actions*” (Sousa, 2017).

Considering this definition, one common type of information used to represent an agent/entity, even if unknowingly, is the visuals: everything that relates to the seeing and sight (visual perception, *q.v.* Section 2.2.2 - Learning and Expressing Complexity), which constitutes visual context. In the same way, it should now be clear that location is yet another possibility of representing an agent/entity, not necessarily through visual means (Alves Fernandes, 2016), under the umbrella of spatial context—which comprises anything that might be used to describe the space (*i.e.*, sensory cues, *q.v.* Section 2.2.2 - Learning and Expressing Complexity), that is, the surrounding environment, *e.g.*, sound, smell, temperature, haptics, and indeed the location, including disposition of agents in relation to objects. Spatial-awareness can be approached in a user-centred perspective in which “conceptualisations imbue spatial structures with meaning” through the interactions between three entities: a (cognitive) agent, the environment, and a cartographic map (Freksa et al., 2007). This trilateral relationship is important to understand (1) why maps are useful for our (spatial) orientation in an environment in which we are immersed, and to which we have direct visual access; and (2) the role of cognition (and the mind) and agency, while solving orientation and wayfinding tasks. Specifically, the interaction between agents and their environments will determine the affordances: some emerge, and others are already imposed due to physical limitations (*i.e.*, anatomy and physiology). The latter will, to a large extent, determine where the agent will move, but the cognition of the agent can further help by carrying out operations mentally that would otherwise require physical actions in the surrounding (spatial) environment itself. One such example is through the use of cognitive maps, that is, internal representations of the surrounding environment, that help the cognitive agents plan and control their movements in space. These mental operations are much less dangerous than the corresponding physical actions, just like a shortcut or running a simulation in a virtual world (*q.v.* Section 2.2.2 - Learning and Expressing Complexity).

The visual and spatial dimensions of context are intertwined, and are included under an even broader dimension, known as physical context. While this ontological view of physical context might seem too dissected, it is particularly relevant given that one of the aspects being investigated in this systematic review has to do with enhanced visuals (comprising the entire spectrum of immersive environments), and enhanced spatial-awareness (provided by location, and somatic, embodied agency), to leverage the understanding and expression of something deemed as complex.

A good example of the importance and application of the socio-culture dimensions of context, in particular, is in the field of general practice/family medicine (complex adaptive system), as a way to improve the understanding and responses to illnesses (*q.v.* [Sturmberg et al. \(2014\)](#) in Section 2.2.1 - Complexity). Another example is the influence of context when learning a language is the variability of the meaning assigned to a word ([Bronislaw, 1923](#)). Its interpretation is a concern of the field of semantics, which caters to prior usage experiences (*i.e.*, social and cultural background and prior knowledge—([Norman, 2013](#))) and to the situation or conditions/constraints in which the word is employed. Namely, it can be determined by the language itself (*e.g.*, its rules and available words—([Winawer et al., 2007](#))), time (and its perception ([Boroditsky et al., 2011](#))) and place of the conversation (*e.g.*, urban or rural center, place of work or home), speaker and receiver of the message (*e.g.*, proximity, social status, professional level and culture), tone of voice (*e.g.*, deep, quiet), the words that precede and succeed, among others ([Ogden and Richards, 1923](#), p. 224).

In this perspective of integrating the context in learning, each student should, ideally, have an individualised and personalised plan, leveraged by real-life learning contexts, including the accompaniment of someone who enhances their experience according to both individual and common needs, expectations and objectives. These would always depend on the skill level of the learner, which can be categorised into five different stages according to Dreyfus' model of skill acquisition ([Dreyfus et al., 2000](#), p. 20). However, given the society's labour needs, and the difficulty—or cost—of allocating a specialist (*i.e.*, competent, proficient, or expert in [Dreyfus et al. \(2000\)](#)'s model) to each student, this is also not possible or feasible. As such, the

use of technology for this purpose, considering its potential to leverage activity, interaction, and culturally rich social immersive environments, is discussed by several authors (Figueiredo and Afonso, 2006; Ifenthaler and Eseryel, 2013; Morgado et al., 2010, 2015b; Roque et al., 2004; Roque, 2006) and is adopted as the basis of the hypothesis investigated in this systematic review, *cf.* Alves Fernandes et al. (2019)).

Tackling this with technology, however, creates new challenges concerning the use, design, and development of these environments, taking into account that the current technology still does not provide us with the authentic context of most daily basis complex situations. This occurs for numerous reasons; from immersive environments (*e.g.*, Virtual Field Trips (VFTs) (Han, 2020)) not being (for the most part) tangible (contrary to what happens in conventional learning), to the use of technology as a mediator and the importance of the social dimension (Figueiredo and Afonso, 2006), and interaction (discussed further, *q.v.* Section 2.2.5 - Multimodal Human-Computer Interaction). In order to tackle these challenges, and in accordance to different theories (*e.g.*, Activity Theory (AT), Cultural-Historical Activity Theory (CHAT) (Roth and Lee, 2007), Actor-Network Theory (ANT)), several authors (Bauer and Dey, 2016; Dey et al., 2001; Figueiredo and Afonso, 2006; Ifenthaler and Eseryel, 2013; Roque, 2006) propose methods, models, and recommendations for context analysis and design. Some of these proposals put emphasis on the learning paradigm used to support context design in immersive learning environments. If we consider constructivism, and depending on the associated complexity, predicting something will only be possible to a certain extent, thereby recommending a balance between the designed and the emergent (Figueiredo and Afonso, 2006), that is, between what can actually be predicted (*e.g.*, directions, boundaries), and what may emerge during the learning process (*e.g.*, socio-material quasi-objects within a constructivist theory of mapping relations between material and concept such as ANT—(Hylving, 2017; Serres, 1982, p. 224). It is therefore given space to explore and learn because ultimately, it is the student who creates the necessary connections for effectively learning a concept (*e.g.*, discovery learning—“all forms of obtaining knowledge for oneself by the use of one’s own mind” (Bruner, 1961, p. 22)).

Similarly to the learning category (*q.v.* Section 2.2.2 - Learning and Expressing

Complexity), context also tends to be more general and subjective to search for through keywords and can be assessed more effectively through manual screening instead (*i.e.*, inclusion and exclusion criteria—*q.v.* subsection 2.3.3 - **Inclusion and exclusion criteria**). This background section on learning in context is still particularly important to realise that context appears in several forms and dimensions, some of which invisible, and its benefits (and challenges it poses) to learning, particularly when it comes to complexity as “the context in which a persistent emergent pattern is embedded determines its function” (Holland, 2000, p. 226). Context is the main pillar of the conjecture being investigated, *cf.* Alves Fernandes et al. (2019)), and subsequent research question of this systematic review (*q.v.* subsection 2.3.1 - **Research Questions**), thus being essential for the processes of data collection (*q.v.* subsection 2.3.5 - **Data collection**) and data processing, namely to define codes and themes (*q.v.* Section 2.4 - **Thematic Analysis**) that are grounded on previous literature. The cultural dimension of context, of particular interest to this review, can still, however, be used to help narrow down results more effectively through keywords (*q.v.* Table 2.2).

Table 2.2 – Keywords, not including syntactic variants, for the Cultural Category—part of a more extensive **search string** (*qq.v.* Table 2.5).

Culture Keywords			
culture	heritage	narrative	storytelling

2.2.4 Immersive Environments

The definition of immersion (and, consequently, immersive environments) ultimately depends on the research field and perspective lens, leading to a lack of consensus in the literature, which often uses other terms such as *presence* and *realism* interchangeably. While some consider immersion to be purely a mental state (Witmer and Singer, 1998), others tend to tie its meaning with the employed technology (Routhier, 2018; Slater, 2009). Recent studies (Agrawal et al., 2019; Beck et al., 2020; Nilsson et al., 2016), however, reach a more holistic view on the matter, which are crucial for this **Systematic Literature Review (SLR)** to avoid such subjectivity

by adopting a definition of immersion and immersive environments before moving any further. That said, this study is, in addition to other particular aspects (*q.v.* Section 2.3.2 - Search Process), looking for immersive environments as per Beck et al. (2020) definition *apud* Agrawal et al. (2019) and Nilsson et al. (2016), but only those that leverage computing technologies (*apud* Slater (2009)) and cultural (human-computer) interaction (*apud* Morgado et al. (2015b))—as discussed further.

To clarify, immersion goes beyond system or technology-centric perspectives, and brings into account the other dimensions that constitute Nilsson’s immersion taxonomy (*i.e.*, narrative immersion and challenge-based immersion). Specifically, immersive environments as “*the surroundings in which a person may experience immersion. It is described by the a) physical and technical system and its properties, b) narrative content, and c) encompassing context (physical setting and its contextual conditions), in the stead of Nilsson et al.’s perspective.*” (Beck et al., 2020).

We further refine the scope of this review by looking for experiences that leverage computing technologies. By that we mean, using Slater (2009) terminology, higher order systems that are used to provide Place Illusion (PI)—sense of being there—and some degree of Plausibility Illusion (Psi)—what appears to be happening is really happening. And, as Slater (2009) put it, “*if we consider PI as based on the extent to which normal Sensorimotor Contingencies (SCs) apply to perception*” then PI cannot occur in traditional 2D screen mediums (*e.g.*, desktop systems). This means that the scope of this review is excluding experiences that rely exclusively on the use of a 2D traditional screen as the display method (not to be confused with 3D stereoscopic display, projection based experiences, or touch-based displays). One consequence of this option is this review having a narrower scope than Beck et al. (2020)’s, and thus excluding their “*plain interactive text adventure on an old-school computer*”. To clarify, despite this scope refinement option restricting the selection of studies considering only one of immersion’s dimensions, the data extraction from said studies still considers the remaining ones.

The quality of an immersive experience is ultimately tied to perception, in turn affected by SCs (using the body to perceive) and therefore tackled sensory modalities (*e.g.*, visual, audio, haptic, *q.v.* Section 2.2.5 - Multimodal Human-Computer

Interaction). If we take Immersive Virtual Reality (IVR) as an example, in principle, experiences with a high frame rate, wide field of view, low latency, stereo, and a head-tracking system are enough to trigger appropriate and realistic human responses even in a wire-frame rendered scenario (Slater et al., 2009). Evidence isn't yet clear on the influence of visual quality on presence (Slater et al., 2009), other than in terms of the expectations of people based on their past experience with the quality of video games. *Presence*, as *being in a place* does not rely on perceptual realism (how well the surrounding environment looks and sounds when compared to the physical world—(McMahan, 2003)), but rather on the SCs (and interaction, *q.v.* Section 2.2.5 - Multimodal Human-Computer Interaction) it supports. While some authors divide *Realism* into geometrical and illumination factors (Slater et al., 2009; Yu et al., 2012), others also add a social dimension to it (Lombard and Ditton, 1997; McMahan, 2003)—which for the former studies is part of Psi within the concept of *Presence* instead. Animated cartoons or low poly, for example, could have low perceptual realism but a high degree of social realism (and Psi), which can be achieved by designing the virtual world to match the physical one, with streets, landmarks, as well as organising rituals and ceremonies that enable users to identify their social place within the virtual world (McMahan, 2003). One could conjecture, then, that based on the realistic responses observed in several studies with lower graphics quality (Slater et al., 2009; Yu et al., 2012), the rendering quality itself doesn't seem to matter much. As everything, however, it most likely depends on the experience itself and what is being depicted. For instance, in a simulation that is too risky to undertake in the physical world, higher Psi is a must, otherwise the responses from the user (agent) might not correspond to what would happen in a real scenario. At the same time, it also might be important to have a higher degree of perceptual realism. For such scenarios, technology and techniques to perform 3D scanning and 3D mapping (*e.g.*, Photogrammetry, Geographical Information System (GIS) map generation, laser scanning—particularly with the recent outspread with a Light Detection and Ranging (LiDAR) sensor being incorporated in iPhone 12), and superimposition of 3D models could be employed. Realistic avatars might also be important in some situations, but responses appear to be shaped both by the agents' behaviours and by people's expectations of the technology (*i.e.*, the effect

might be diminished with more experienced users) (Garau et al., 2005). More recently (Seinfeld and Müller, 2020) it has been shown that (virtual) hands detached from a virtual body decrease embodiment when compared to connected hands, even with precise hand tracking. In other words, an experience that depends on the response of the user’s hands for some event, wouldn’t fully simulate the physical world if the “virtual” representation of his/her hands were not connected (*i.e.*, arms are not tracked or depicted in the virtual experience).

This review is also being aligned with the current trend of immersive technology and HCI and only considering studies from 2013 onwards (similarly to two other systematic reviews (Jensen and Konradsen, 2018; Snoswell and Snoswell, 2019)). We conjecture that 2013 was the point of change for both VR and AR with the release of Oculus Rift DK1, Google Glass (a year later), and perhaps most importantly, the outburst of smartphones consumer wise. While that is true, it is important to note that equivalent classes of systems that can be used to approximately simulate a 3 Degrees of Freedom (3DoF) Head-mounted Display (HMD) (Slater, 2009) are still considered, such as Cave Automatic Virtual Environment (CAVE), power wall, or similar projection-based experiences, despite being older technologies. They are included on the basis that we are also looking for Cultural human computer interaction (Morgado et al., 2015b).

Immersion—and immersive environments, by extension—is yet another category of utmost importance for the conjecture being investigated (Alves Fernandes et al., 2019) and subsequent research question of this systematic review (*q.v.* subsection 2.3.1 - Research Questions). Given its vast application and research, this background section on immersion and immersive environments, in conjunction with previous ones (*q.v.* Section 2.2.2 - Learning and Expressing Complexity) is important to not only make clear the adopted definition, but to also understand the scope of this review and consequently help setting appropriate keywords (*q.v.* Table 2.3, below) and inclusion and exclusion criteria that satisfies it (*q.v.* subsection 2.3.3). Similarly to the previous background sections, this subsection is still particularly important to support the processes of data collection (*q.v.* subsection 2.3.5) and data processing, namely to define codes and themes (*q.v.* Section 2.4 - Thematic Analysis) that are

grounded on previous literature.

Table 2.3 – Keywords, not including syntactic variants, for the Immersive Environments’ Category—part of a more extensive [search string](#) (*qq.v.* Table 2.5).

Immersive environments Keywords						
Immersive	Immersion	Virtual reality	Augmented reality	Mixed reality	Cross reality	X-reality
Serious game	Microworld	Virtual world	Edutainment	Learning game	Educational game	

2.2.5 Multimodal Human-Computer Interaction

[HCI](#) also challenges the development of context-aware environments (*q.v.* Section 2.2.3 - [Learning in Context](#)). There is a steadily increasing effort in the research community to improve communication techniques through speech, gestures, facial expressions and other interactive system communication modalities ([Turk, 2014](#)) since human interaction and communication with the world is inherently multimodal ([Bunt, 1998](#); [Quek et al., 2002](#)). The research objectives in this field are the development of technologies, methods and interfaces that employ and combine the human senses in order to try to provide a more natural interaction on the part of the users. The “natural” allegation of the so-called [Natural User Interfaces \(NUIs\)](#) has been criticised ([Malizia and Bellucci, 2012](#)), with empirical support ([Bowman et al., 2017](#); [Liebold et al., 2020](#)), due to the high level of artificiality it entails and even for the fact that most such “natural” interfaces do not follow basic principles of interaction design ([Norman, 2010](#)). They imply the learning of a set of predefined gestural commands by users—rather than the other way around (*i.e.*, the interfaces accepting whatever commands users would already deem “natural”). For instance, one can consider the interaction with current head-mounted devices such as Microsoft HoloLens, Oculus Rift and Quest, HTC Vive, and Magic Leap. Despite gesturing being natural and innate, switching between these devices’ interfaces requires learning different arrays of gestures/commands using one’s hands or joysticks—just as what happened with classic gestures and interfaces. Gestures, in particular, despite being natural and innate, lack studies from a semiotic perspective ([Roth and Lawless, 2002](#)), ignoring the cultural and social dimensions of context. While some argue that gestures have a crucial role in the development of language, and address

these obstacles by selecting gestures or gestural emblems that have a more or less widespread meaning across cultures, attempting to minimise critical failures (*i.e.*, multimodal HCI), others still claim they are merely an auxiliary form of expression. To support the former, a study from 2003 provides data for the “overarching hypothesis that gestures have a transitional function between ergotic/epistemic movements of hands and symbolic expressions” (Roth, 2003). It is time to change the current strategy and put people’s needs in first place and use technology to bridge the gaps in our performance (*i.e.*, Human-centred Design (HCD)) (Norman, 2014, 2013, p. 8).

As knowledge about interaction design for gestural interfaces becomes more consolidated, current gestural command methods for applications will become obsolete. This obsolescence coupled with the emergence of a number of low-cost somatic and gestural interaction devices (*e.g.*, Kinect, Leap Motion, Myo Gesture, and Maestro Gesture Glove) should not prevent current applications from being available under new modes of gestural command. Multimodal HCI addresses these obstacles by selecting gestural gestures or emblems that are mostly widespread across cultures, trying to minimise critical flaws. It is noteworthy, however, that very few gestures are universal (if any), so a better option might be giving the user the option to select the adequate ones. There are a few simple, yet meaningful examples that can show the cultural diversity often ignored while travelling (Axtell, 1997), let alone while using tracking technology. A good example of cultural significance when it comes to interaction is that of pointing. While most western cultures use the index finger to point (some even call it the “pointer” finger), eastern countries, such as China, use an open hand to do so. In Korea, in addition to the open hand, one can also use the middle finger (which in most Western countries is seen as an offence). In other Asian countries it can even be considered rude to use a finger (or foot) to point. Another common example is the thumbs up. While many can consider it being a sign for “okay” or hitch-hiking, in Iran and Nigeria it is considered an insult. In Germany and Japan it means “one” instead. Finger counting, as simple as it may seem, is also not universal. In Germany, if one wants to show the number three (3), they will use the thumb, index, and middle fingers, while most of the other European countries use the index, middle, and ring fingers instead. In Japan, counting to ten (10) can be performed with one hand alone, by starting with all fingers extended

and folding one by one into the palm (starting with the thumb) until five (5), and then the opposite process (*i.e.*, extending the folded fingers) from six (6) to ten (10). Other countries will use both hands sequentially. All these are subtle, yet meaningful differences that can make the **User Experience (UX)** much richer and immersive if taken into account. This was sought to be applied in the architecture (Morgado et al., 2015a) developed under the InMERSE¹ project as the foundation of the prototype “First Army of India” (Alves Fernandes et al., 2016) so that users could choose the devices and preferred modes of interaction. An example that, in this particular prototype, could solve or lessen one of its obstacles—the lack of coordination of the player Adamastor when using the Google Glass and Kinect devices simultaneously—would be to replace the current grab and throw gesture detection, accomplished with Kinect, with Myo Gesture bracelets instead (Alves Fernandes et al., 2015).

A relation between **HCI** and anthropology ought to be established to create a better, more human relationship with technology. An initial proposal (Alves Fernandes et al., 2019) leverages the cultural and social dimensions of context with immersive environments so that systems become more context-aware, therefore attempting to ease the process of complex learning. Another interesting idea to bridge this gap, put forward Keitchi Matsuda (KamiOS), is grounded on the use of our ancient systems of belief, ritual, and community, by bringing objects to life (known as Tsukumogami in the Japanese culture) and extending them with better attribution and perception of mental states such as intents, desires, and knowledge (*i.e.*, Theory of Mind principles—(Premack and Woodruff, 1978)—and related to anticipatory systems *q.v.* Section 2.2.3 - **Learning in Context**). Previous studies (Alves Fernandes et al., 2016; Morgado, 2014; Morgado et al., 2015a) present preliminary contributions on the potential and feasibility of creating a cultural layer abstracting the interaction processes and elements (*e.g.*, gestures and other somatic aspects) from its system control effects (commands)—the shamanic interface. Critically, recent results point towards the use of cultural-aware gestural emblems for interaction resulting in better remembrance and lesser command input errors than using gestural emblems that

¹ <https://bitbucket.org/Apidcloud/inmerse-framework>

are not cultural-aware (Carvalho, 2019). While other studies seem to either disregard the importance of HCI, often bringing it into the equation because it is novel (*e.g.*, hand tracking in Oculus Quest), or indirectly through the experience (*e.g.*, walking from one Point of Interest (POI) to another), the other extreme goes a bit far and calls it “natural” interaction instead (*e.g.*, hand tracking), which, as argued before, is not the case if the cultural and social dimensions of context are still ignored (*i.e.*, the experience is not tailored to the user). Not leveraging those dimensions does not only affect the experience negatively, but also limits its (learning) potential. Studies and experiences that do not target a specific group (*i.e.*, that are available for the general public), or that do not change accordingly, cannot leverage the participants prior knowledge nearly to its potential—not only intellectually but also interaction wise. Given the problematics of distinguishing between attentive and inattentive interaction interfaces with the current trending literature (and big companies’ advertisement) expressions (*e.g.*, intuitive, fluent, and natural), that take into account the background of the users and surrounding context, we oppose the use of those terms, and suggest the use of more explicit expressions such as context-aware or cultural-aware interaction instead, knowing, however, the problematic of applications labelling themselves as context-aware when, in fact, they are only location-based (*q.v.* Section 2.2.3 - Learning in Context).

Feedback, or the lack thereof, is another challenge when it comes to the usability, accessibility, and development of experiences using gestural and somatic interfaces in particular. Walking (often referred to as locomotion) in VR is an example. Though some headsets support room tracking and 6 Degrees of Freedom (6DoF), walking in a room while in VR is still pretty limited and constrained. Other solutions include the use of multiple teleportation modes (different gestures for different headsets), or the use of a VR treadmill. Even if VR teleportation was set in stone, which is not, the feedback given to the user is still concerning. Feedback is invisible through omnipresence (Stermann, 1994), and haptics, a form of feedback, is no different. The lack of haptics affects both somatic and gestural interfaces. While some smartphones and tables already use some haptic mechanisms to give the impression of a physical press, touchless gestural and somatical interfaces face a bigger problem. Traditionally, the former would use some sort of glove or finger haptics, but they end up falling

short for being invasive and/or requiring more hardware. A more recent and promising solution is that of UltraLeap (a merge of the previously known Leap Motion and Ultra Haptics), using ultrasound to create an impression of virtual touch, that they call mid-air haptics. Their collaboration with Varjo, known for manufacturing high-end VR headsets, might be the step needed to start giving these touchless modalities more attention and standardisation—particularly in a post Coronavirus disease 2019 (COVID-19) world, where touch interfaces will be much more restricted. The creation and exploration of new multimodal techniques and applications towards more context and cultural-aware interaction are thus an opportunity—its combination in particular, since most systems only integrate two modalities, such as speech alongside touch or visual gestures (Turk, 2014). These concepts have also been applied to education and e-learning, in support of a wider variety of student preferences and interests (Sankey et al., 2011), specifically through the development of interactive multimodal learning environments, such as 3D virtual worlds and social media, using verbal and nonverbal modes to represent content knowledge (Moreno and Mayer, 2007). The use of these environments has been shown to enhance learning, rendered more flexible, self-oriented, and enjoyable (Birch, 2008; Picciano, 2009).

Feedback, which can also be seen as consequences, is one part of the equation. On the other end (x) is what causes such feedback (y). In principle, more user (agent) interaction (and SCs, *q.v.* Section 2.2.4 - Immersive Environments) implies more Agency, but it is important to still separate the latter from the actual technical means of interaction (its mechanics). Agency is the capability of an entity (a person or other) to act in a virtual environment or made-up context, and it is a key contributor to enabling a state of presence in the experiencer (along with PI and Psi, *q.v.* Section 2.2.4 - Immersive Environments). Though it does that by playing a huge role in the surrounding context's dimensions in various ways (*q.v.* Section 2.2.3 - Learning in Context), it can further positively affect immersion if also tackling the cultural and social dimensions. As Bock et al. (2018) put it, “interaction is also a path to immersion”. That can be accomplished by leveraging as input methods the rich semantics (the meaning) of cultural-social gestures, rituals, and symbols in contextually relevant situations, and through what is known as embodiment (or embodied agency). Examples of that include the entity giving form to some idea or

personifying a specific character and its role (*e.g.*, scavengers hunting in the woods with certain goals, and abilities and choices to attain them). Symbols, despite being a more generic term (and thus not being used here as search keyword), are a subclass of signs in the field of semiotics such as colours (Caivano, 1998) and drawings, that have a (cultural) meaning tied to them (*i.e.*, a semantic dimension—(Norman, 2013; W. Morris, 1938)). Signs, originally defined by Peirce (1935), and followed by Caivano (1998), are “something that stands for something else and that is understood or has some meaning for somebody” and “is used to transmit a concept” or “substitute something”.

Interaction is another category closely related to context, particularly to its physical, spatial, and cultural (and social) dimensions. Given its importance to the conjecture being investigated (Alves Fernandes et al., 2019) and subsequent research question of this systematic review (*q.v.* subsection 2.3.1), and the fact that its keywords are not so subjective and general (*q.v.* Table 2.4, below), it can be used effectively to narrow down search results before assessing them further through manual screening (*i.e.*, inclusion and exclusion criteria—*q.v.* subsection 2.3.3). This background section on interaction is still particularly important throughout the processes of data collection (*q.v.* subsection 2.3.5) and data processing, namely to define codes and themes (*q.v.* Section 2.4 - Thematic Analysis) that are grounded on previous literature.

Table 2.4 – Keywords, not including syntactic variants, for the Interaction Category—part of a more extensive search string (*qq.v.* Table 2.5).

Interaction Keywords				
interaction	gesture	somatic	embodiment	interface

2.3 Method

Systematic Literature Reviews (SLRs) represent a critical part of research in Evidence-based Software Engineering (EBSE). It is a concise summary of the best available academic evidence that uses explicit and rigorous methods and strategies to identify, critically assess, and synthesize relevant studies on a set of particular topics (Cruzes

and Dybå, 2011). This SLR study is based on the guidelines proposed by Kitchenham and Charters (2007) and Okoli (2015) for software engineering SLRs. One of the main aspects of these guidelines is the creation of a review protocol that defines how the review is conducted by specifying all the steps to be performed (Mahdavi-Hezavehi et al., 2017). Specifically, the protocol contains the research questions, the search process to identify and collect relevant primary studies, inclusion and exclusion criteria for filtering out remaining unrelated papers, quality assessment to filter or assess them further, and methods for extracting data and synthesising them to answer the research questions. In this particular study, the goal is to assess primary studies, thus being categorised as a secondary literature review, where the data is synthesised through what is known as **Thematic Analysis (TA)** (*q.v.* Section 2.4). All the steps taken in this SLR are documented throughout this section and are summarised at the end (*q.v.* figure 2.1).

2.3.1 Research Questions

The main goal of this SLR is to answer a single, yet complicated, research question. As one might notice further, it combines multiple fields and categories (*q.v.* subsection 2.3.2 - **Search Process**). We thus opted to divide it into sub-questions, extracting their answers when collecting data (*q.v.* subsection 2.3.5) from the resulting selected papers as an attempt to, in the end, gain a more holistic and digested view and finally answer the main research question.

1. How can immersive environments leverage context and cultural gestures to increase the understanding and communication of complex concepts by non-technical users?

- 1.1 What is the complex topic being learnt/taught? *e.g.*, complex flight or weather dynamics, cultural heritage and values, learning of complexity.

- 1.2 What technologies are used to provide visual and spatial context?

- 1.3 What techniques (if any) are used to provide more visual and spatial context?

- 1.4 What technologies are used to provide culture-aware interaction?

1.4.1 Is it gestural, somatic, or other? Specify if touch based (*e.g.*, smartphone, tablet, tabletop), *etc.*

1.5 What methods/techniques are used to provide cultural context (and meaning)?

1.6 How is the understanding/expression of a complex topic approached versus traditional methods? How/Which are/of these factors (are) contributing?

1.7 What are the main targets of these studies?

2.3.2 Search Process

The search process consisted of multiple automatic searches across search engines/-databases and exporting the results to a reference manager software (Zotero², in our case). By automatic search we mean performing a search by executing search strings with boolean operators. The research question at hand is multidisciplinary; that is, not tied to a specific field like Computer Science or Medicine, and thus opting out of a manual search across some specific venues (which we still attempted but, given this multidisciplinaryity, did not yield good results). That said, we started by dividing **RQ1** into the following categories:

- **C1.** Complexity
- **C2.** Immersive environments (VR, AR, MR, Extended Reality (XR), *etc.*)
- **C3.** Learning
- **C4.** (Spatial and visual) Context
- **C5.** Interaction
- **C6.** Culture (context)
- **C7.** Public/Target

² <https://www.zotero.org/>

Preliminary searches were then conducted, not only to determine the appropriate search keywords per each of the categories above (*q.v.* table 2.5, below) but also to assess potential relevant studies (“quasi-gold” standard that should be present in the final search) and dates of interest. One of the main challenges of this particular search was the fact that many of the keywords are also used in totally different contexts which are irrelevant for what we are trying to determine. For instance, **C3**, **C4**, and **C7** are particularly **generic**, yielding too many irrelevant results, with keywords like **context**, **learning**, and **public**. We thus **opted** for a search without those categories, and use the inclusion and exclusion criteria instead (*q.v.* subsection 2.3.3) to further refine the search results and make sure each paper matches what is envisioned. It is also important to mention that given what we are aiming at, the search was limited to 2013 onwards, as explained further in the background (*q.v.* Section 2.2.4 - Immersive Environments). Snowballing new research studies through the inspection of the resulting bibliography allowed the scope of examination to be enlarged and consider aspects that were still not addressed in the initial search, and served as a way to partially overtake potential limitations concerning the use of a specific set of search keywords.

Table 2.5 – Categories and respective keywords. Categories with two asterisks appended (**) were not included in the automatic searches.

Categories' Keywords							
C1	Complexity	Complex	nonlinear non-linear	Systems theory	Systemic	Emergent Emergence	Self-organization Self-organisation
C2	Immersive	Virtual reality	Augmented reality	Mixed reality	X-reality Cross reality	microworld microworlds	Serious game
	Immersion	Virtual world	Virtual worlds	Educational game	Edutainment	Learning game	Serious games
C3**	Learning	Understanding	Grasping	Perceiving		Training	Education
C4**	Context	Situated		Simulated		Experience	Environment
C5	Culture	Cultural		Heritage		Narrative	Storytelling
C6	Interaction	Interface		Gesture		Somatic	Embodied
	Interactive	Interfaces		Gestures			Embodiment
				Gestural			
C7**	Non-expert		Non-technical		End-user		Public

All the used categories and respective keywords are depicted in Table 2.5, including the plurals and hyphenated ways of writing a word (*e.g.*, nonlinear and non-linear). Despite some search engines/databases expanding its search to alternative terms automatically, others do not or have changed throughout the years. As a way to

make sure these are always considered (at least the ones we are aware of), we included them manually which resulted in the following search string:

(complex OR complexity OR nonlinear OR non-linear OR systemic OR emergence OR emergent OR self-organisation OR self-organization OR "systems theory") AND (immersive OR immersion OR "virtual reality" OR "augmented reality" OR "mixed reality" OR "cross reality" OR "x-reality" OR "serious game" OR "serious games" OR microworld OR microworlds OR "virtual world" OR "virtual worlds" OR "edutainment" OR "learning game" OR "educational game") AND (interaction OR interactive OR gesture OR gestures OR gestural OR somatic OR embodied OR embodiment OR interface OR interfaces) AND (culture OR cultural OR heritage OR narrative OR storytelling)

The search string was applied to search databases for matching paper **titles**, **key-words**, and **abstracts**. In databases that had other textual fields, we also applied that search string to those fields—but not to the papers’ full text. Table 2.6, below, lists the searched databases, with the respective used search scope.

Table 2.6 – Databases and respective search scopes used in automatic searches.

Engine or Database	Search Scope
Scopus	Title, Abstract, Keywords
ISI Web of Science	All databases - Topic, Title, Abstract, Keywords
PubMed	All fields
IEEE	Metadata - Abstract, Title, indexing terms
AIS	AISeL. Peer Reviewed only. All fields.
arXiv	All fields
ERIC proquest	All fields except full text

As a related note, one might ask why ACM or ERIC are not part of the list. Firstly, both of them did not support multi-queries at the time of the search. That alone should not be reason enough, however, as it is possible to divide the query with more or less effort and achieve the same result. While that is true, ACM papers are well covered in other databases/engines such as Scopus, so we did not consider it

further. Finally, regarding ERIC, we found an alternative named ERIC proquest also targeting education research which yielded good results. The database automatic searches yielded a total of **620** results. That being said, after removing obvious unwanted results (*e.g.*, non-english, certain manuscript types) and duplicates, we were left with a set of **489** studies. To that, we still added several articles for potential snowballing (**29**) by analysing the bibliographic references and contacting the authors of the former set of studies to ask for similar articles, totalling **518** primary studies.

2.3.3 Inclusion and exclusion criteria

The articles were screened according to various criteria especially pertinent to the research questions. We started by a tentative set of inclusion and exclusion criteria, which we tentatively applied to the first few papers. Upon scrutiny of their clarity and effectiveness, we refined them, and reapplied them to the same papers, and then to a few more papers. This refinement process was repeated three times, until the final set of inclusion and exclusion criteria was achieved. The final criteria were then applied to all papers.

All the following **inclusion criteria** had to be met:

- The study should be published between Jan 1st 2013 and June 18th 2019 (*q.v.* Section 2.2.4 - Immersive Environments), or in press but available at the latter date.
- The study should be about aspects of learning, or teaching of a complex topic.
- The study should use immersive environments for that purpose.
 - The study is using a VR headset (HMD) with, at least, 3DoF.
 - Or, the study is using CAVE, Power Wall, or similar projection-based virtual environment experiences (*e.g.*, e-REAL lab), or even a Stereoscopic 3D Display instead of a VR headset.

- Otherwise, the study is using **AR** (*e.g.*, through glasses or smartphone/tablet), or a **MR/XR** headset.
 - * For instance, a Museum **AR** guide/tour or other Cultural Heritage experiences (*e.g.*, lost cities, monuments).
- The study should have some form of “meaningful” (and coherent) physical interaction.
 - When using **AR**, the interaction includes at least the physical movement from the user (*e.g.*, location based) or more meaningful gestures (*e.g.*, V for flashlight, palm swipe for scanning the screen, shaking the device to undo).
 - Otherwise (physical immersion), the interaction includes at least gestural and/or somatic interaction (*e.g.*, head movement for gaze tracking). If only using **VR** joysticks, then the used commands/gestures should be explained, and should not rely exclusively on single clicks/triggers. A good interaction example is controlling the zoom/scaling by moving joysticks closer/apart.
- The study should be (more) context-aware by tackling (some) of its different dimensions, namely physical, social, and cultural.
 - Culture dimension is present in some form (*e.g.*, cultural interaction, culturally relevant objects or songs, surrounding environment, cultural heritage topic)
 - Physical dimension (spatial and visual context) is present through location-awareness, immersive environment (*e.g.*, modeled to resemble a specific place, photogrammetry), and/or somatic, embodied agency.
 - For instance, the AR experience requires the (social) collaboration with people outside the experience.
 - For instance, the AR experience requires creating/collecting some physical material to use/interact within.

At the same time, the articles should avoid any intersection with the **exclusion criteria**:

- The study is not written in English.
- The study is not available to read (full-text).
- The study is not published in a peer-reviewed venue.
- The study is an editorial, abstract, summary, keynote, opinion, tutorial, panel discussion, introduction to a special issue, thesis or dissertation.
- The study is a duplicate or older version of the same study, or presents high levels of similarity with a more comprehensive/complete article.
- The study, when applicable (*e.g.*, a systematic literature review), does not present clear research design (research questions defined, plan) or methodology (data collection and analysis).
- The goal of the experience/study is not to increase the understanding and/or expression of complexity.
- When not using **AR**, the study does not provide physical immersion.
- The study only uses the traditional **2D** screen medium (not to be confused with stereoscopic **3D** displays, projection based experiences, or touch-based displays).
- The study does not provide culture interaction.
 - For instance, the study relies on the use of a mouse and/or keyboard to interact with the described experience, rather than methods that map physical motion to virtual motion.
 - For instance, the study relies exclusively on the use of a single tap/click (*e.g.*, to show additional information about a Point of Interest, open a menu, select an option) on a screen (*e.g.*, tablet, smartphone, smart **Television (TV)**), rather than, for instance, a double tap to fast-forward.

- For instance, the study relies exclusively on reading Quick Response code (QR) codes/markers inside a museum to show more information, rather than using the position/location (*e.g.*, Global Positioning System (GPS), beacons, Radio-frequency identification (RFID)) or detecting a specific place or object without the use of markers (*e.g.*, image recognition).
- The study is exclusively targeting technology experienced people and/or topic professionals. In other words, the experience described in the study is exclusively testing with or targeting users that already know how to use the employed or similar equipment, and/or people that are already at a competent, proficient, or expert level in the given experience topic.

2.3.4 Quality assessment

To analyse the reporting quality—not to confused with research quality—of each filtered primary study, we established a checklist with seven questions originally based on Kitchenham and Charters (2007) and Garcés et al. (2017):

- **QAQ1:** Is the study published in a known non-predatory venue (*cf.*, Bealls List³)?
- **QAQ2:** Is there a rationale for why the study was undertaken?
- **QAQ3:** Is an overview about the state of the art of the area in which the study is developed presented?
- **QAQ4:** Is there an adequate description of the context in which the work was carried out?
- **QAQ5:** Is there a justification and description for the research design?
- **QAQ6:** Is there a clear statement of findings and has sufficient data been presented to support them?

³ <https://beallslist.net/>

- **QAQ7:** Are the credibility and limitations of their findings explicitly discussed?

Answers to each of the questions followed a scale from 0 to 1, but were limited to three possible choices: (i) the study does not meet the quality criterion (0 points); (ii) the study meets the criterion to some extent (0.5 points); and (iii) the study fully meets the given quality criterion (1 point). The final quality score for each primary study was obtained by summing up the scores for all the respective **Quality Assessment Questions (QAQs)**. The score of each study could then fall into one of the following set of ranges: 0-1 (very poor); 1.1-2.0 (poor); 2.1-3.0 (fair); 3.1-4.0 (average); 4.1-5.0 (good); 5.1-6.0 (very good); 6.1-7.0 (excellent). Studies with a score equal to or above 3.0 (fair) were considered for the data collection.

2.3.5 Data collection

The papers that passed the previous screenings (title and abstract, inclusion and exclusion criteria, and quality score equal to or above 3.0 points) were selected. They were then read in detail in order to extract the data needed to answer the research (sub) question(s) (*q.v.* Section 2.3.1). To do so, we used a collection form (*q.v.* table 2.8, below) inspired by [Mahdavi-Hezavehi et al. \(2017\)](#) where we also stored documentation data (*i.e.*, **F1-F5** and **F14**). For consistency purposes, only one researcher reviewed the papers and extracted the (meta) data according to the form.

2.3.6 Overview of search process

The automatic search phase yielded an extensive list of potentially relevant papers ($a=489$) gathered after filtering and merging of results from the different sources, and deduplication, *q.v.* Figure 2.1. While applying the inclusion and exclusion criteria, several articles were identified in the papers' text with potential for containing significant data elements. These were added as potential snowballing contributions ($s=29$). After applying the inclusion and inclusion criteria to the **518** papers

($n=a+s$), the final set of unique filtered papers to collect data from was obtained ($f=38$, *q.v.* Table A.1 - Studies included in final review) and later analysed (*q.v.*, Section 2.4 - Thematic Analysis). Table 2.7, below, shows that 23 primary studies out of 38 (f) were found in conferences, and 15 primary studies belonged to journals.

Table 2.7 – Number of papers per publication type

Publication Type	Number of Papers	Paper IDs (<i>q.v.</i> Table A.1)
Conferences	23	S1, S3, S6, S7, S8, S10, S12, S13, S14, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S29, S30, S33, S37
Journals	15	S2, S4, S5, S9, S11, S15, S16, S17, S28, S31, S32, S34, S35, S36, S38

It is worth keeping in mind that whenever a study was unclear, ambiguous, or did not include all the relevant data, its authors were contacted to try to obtain not only the required information, but also to ask for further studies for potential snowballing. Further data was obtained from older studies from which the most recent study originated (when applicable), or from linked videos of the experience whenever mentioned within the respective studies. There are, however, cases where one can reason about certain aspects that are not explicitly mentioned in a study, such as the use of a camera, and some sort of localisation and pose estimation (e.g., GPS, Inertial Measurement Unit (IMU), Compass, RFID tags, Markers, 3D recognition) for a location-based AR experience. In a VR scenario, even if the study does not mention the specific hardware being used throughout the study (*e.g.*, Oculus Quest, HTC Vive), one can still generalise and safely assume that it does use a VR headset of some sort. Even if a study does mention the hardware, we cannot assume that some important characteristics such as the Degrees of Freedom (DoF) are fully used. That is, we cannot assume that every study using technology with 6DoF or Room tracking capabilities actually uses them to leverage the interaction within the experience unless the study or the described experience mentions, reasons, or clearly shows it through pictures or videos. In case a decision could still not be made without additional data and the authors did not reply until the end of the data collection phase, the respective studies were not considered further.

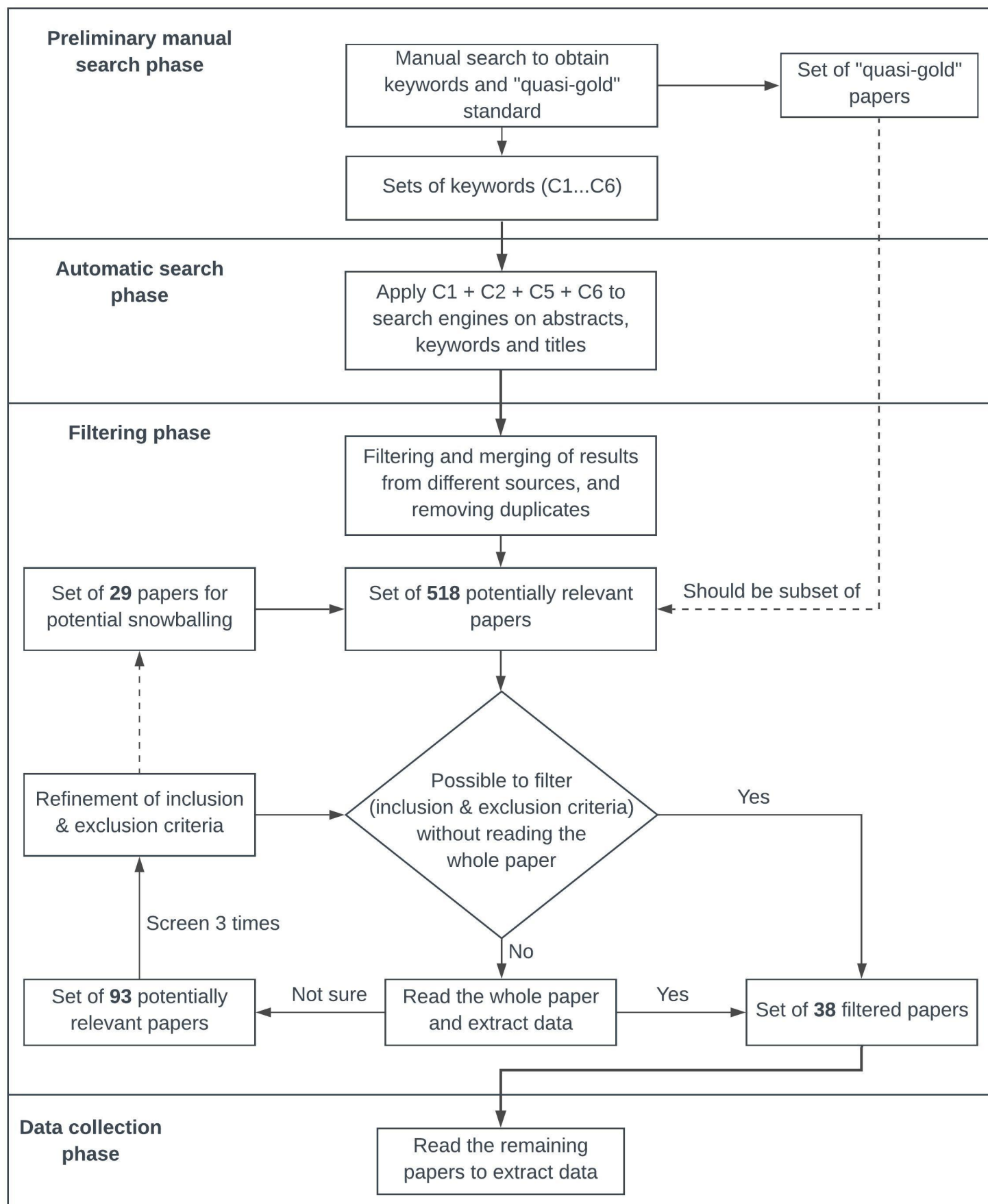


Figure 2.1 – Overview of search process.

Table 2.8 – Systematic Review’s Data Collection Form

#	Field	Concern or RQ	Description
F1	Date of extraction	Documentation	N/A
F2	Reference to the paper, including author, year, and title	Documentation	N/A
F3	Citation Count	Documentation	From Google Scholar
F4	Status	Documentation	Included or Excluded according to the inclusion/exclusion criteria and quality assessment
F5	Quality Assessment Score	Documentation	The result from the quality assessment questions.
F6	Complex Topic	1.1	What is the complex topic/concept being learnt/taught (<i>e.g.</i> , complex flight/fluid/weather dynamics, cultural heritage and values, learning of complexity)
F7	Technologies for visual and spatial context	1.2	HMD, Glasses, and/or through sensors such as GPS, wearables, Leap Motion, Kinect, <i>etc.</i> ? Specify what is used.
F8	Techniques for more visual and spatial context	1.3	Photogrammetry, Spatial Sound, Low poly contents, overhanging text, 3D scanned objects, different avatar genders/hands/3D scanned faces, visual arrangement of objects, <i>etc.</i> Or directly through AR/MR?
F9	Technologies for culture-aware interaction	1.4	Leap motion, Kinect, Myo gesture, <i>etc.</i> Or is it through direct interaction with physical objects?
F10	Type of interaction	1.4.1	Is it Gestural, Somatic, or other? Specify if touch-based (<i>e.g.</i> , smartphone, tablet, tabletop), or other.
F11	Method/Techniques for cultural context (and meaning)	1.5	Is it done through the interaction? Or through storytelling, narratives, language, specific cultural objects, or songs? Or is it through the topic being learnt/taught? If any.
F12	Method(s) for complex learning	1.6	How is the learning being approached? Is scaffolding (with an expert) used? One on one mentorship? Complex problem solving? How does it compare to traditional methods such as visualisation in 2D screens? How/which are/of these factors contributing? If any.
F13	Target	1.7	The general public, specific types of non-experts (<i>e.g.</i> , non-programmers, K12), people that are not experts in the paper subject, even though they are experts in other fields?
F14	Additional Notes	Documentation	N/A

2.4 Thematic Analysis

Thematic analysis (TA) is an umbrella term for some approaches of analysing qualitative data used to identify common themes (Braun et al., 2019), whose definition, in turn, varies. We locate the approach used herein as Reflexive TA, with the aim of providing a coherent and compelling interpretation of the data that is grounded on the data itself. In this approach, themes are defined as “patterns of shared meaning underpinned or united by a core concept” (Braun and Clarke, 2019), evident as the output of multiple iterations of coding (an open and iterative process) through a mix of induction and deduction, and a mix of a conceptual (latent–reading into the sub-text and assumptions underlying the data) and an explicit (semantic–analysing the explicit content of the data) process (Braun et al., 2019). Specifically, our Reflexive TA involved four (4) iterations. Two of which were done independently through the codes (product of coding). That is, without a link to the other fields collected from the same study, as an attempt to create less-biased themes. The third round was more of a review of the themes, this time while cross-matching with the other collected fields of the same study, as an attempt to determine if some information was missing (*e.g.*, the data collected for one field being relevant for another). The last, fourth, TA round was executed over a longer period of time while debating with the other authors of this study for better theme naming, descriptions, possible rollbacks, etc. As a related note, the first collected field (*q.v.* subsection 2.4.1 - [F6] Complex Topic being learn/taught) contains domain summary themes rather than themes, “organised around a shared topic but not shared meaning—they aim to capture the diversity of meaning in relation to a topic or area of focus” (Braun and Clarke, 2019).

2.4.1 [F6] Complex Topic being learn/taught

2.4.1.1 Architectural Design

This theme groups studies on learning about particular architectural aspects (*e.g.*, vaulted systems, styles), including both the process and final product of planning,

designing, and constructing of structures.

2.4.1.2 Archaeological Heritage

This theme groups studies on learning about archaeological aspects. That is, the study of human activity through recovery and analysis of sites (e.g., ancient forums) and artefacts.

2.4.1.3 Monumental Heritage

This theme groups studies on learning about structures that were created to commemorate some event or figure.

2.4.1.4 Arts & Storytelling

This theme groups studies on learning about visual and performing arts, with the exception of studies on architecture (those are placed under [Architectural Design](#), instead). This includes, for instance, paintings, music, and storytelling.

2.4.1.5 Community Heritage

This theme groups studies on learning about several aspects of community culture, such as its historical narrative and values.

2.4.1.6 Societal History

This theme groups studies on learning about historical evolution of societal features, such as demography and criminality.

2.4.1.7 Social Interaction

This theme groups studies on learning about aspects of social interaction, including interpersonal and intercultural communication.

2.4.1.8 STEM Concepts

This theme groups studies on learning about **STEM** concepts. Science, in particular, includes the branches of natural and formal sciences, therefore covering studies focused on physics, biology, logic, *etc.*

2.4.1.9 Skills & Techniques

This theme groups studies on learning about specific skills and techniques, such as Previsualisation, Body Postures and Motions, and Urban Planning.

2.4.2 [F7] Technology for Visual and Spatial Context

2.4.2.1 Speakers

This theme groups studies that employ the use of speakers (and therefore audio) to provide spatial context during the experiences, such as spatial audio (**3D**), directional speakers or even plain stereo. It also includes enhanced spatial context of locations, such as providing location-consistent sound effects, soundscapes, *etc.*

2.4.2.2 360^o Cameras

This theme groups studies that employ the use of 360^o cameras to capture the surrounding environment, to provide visual context.

2.4.2.3 3D Printers

This theme groups studies that employ the use of 3D printers to print 3D replicas (*e.g.*, a building's maquette), which provide visual context.

2.4.2.4 Smart objects

This theme groups studies that employ the use of smart objects to enrich the experience, including contextual lighting, smart toys and other common items (*e.g.*, smart toothbrushes). These can be installed or placed at specific locations, to provide spatial context, and their actual appearance can be used to convey visual context.

2.4.2.5 Positioning and Orientation technology

This theme groups studies that provide or capture spatial context, using technology that estimates the positioning and/or orientation of the user. This includes GPS, IMU, compass, and other sensors, both autonomous or embedded in mobile devices, and also external locators such as RFID tags and bluetooth beacons.

2.4.2.6 Neurophysiologic sensing technology

This theme groups studies that acquire data from neurophysiological sensors: heart rate, skin temperature, and Electrocardiography (ECG). These provide spatial context about the user itself, or some other actor.

2.4.2.7 Weather sensors

This theme groups studies that employ the use of weather sensors to capture spatial context about a location.

2.4.2.8 Markers

This theme groups studies that employ the use of machine-readable markers (including QR codes) to provide spatial context by requiring users to move to a specific space or to tag a specific space for a purpose. All of these cases also leverage the actual aesthetics of the marker for visual context.

2.4.2.9 Mobile devices

This theme groups studies that provide visual context through mobile technologies such as smartphones and tablets. This includes using smartphones to provide visual context when inserted in Cardboard-like HMDs, or visual context from reading/detecting markers or other sensors. Actual spatial-related sensors embedded in these devices are included in a different theme, *Positioning and Orientation technology*: this Mobile devices theme only encompasses the visual context technology of mobile devices.

2.4.2.10 Fixed Cameras for Markers

This theme groups studies that employ fixed cameras at a specific location to recognise and read markers (including QR codes), thus requiring users to move to a specific space for spatial context. If fixed cameras were used for capturing visuals at a specific location, that would be a case of capturing visual context also. However, no such instances were detected in the survey papers, hence this theme is solely about spatial context.

2.4.2.11 Monitors

This theme groups studies that employ various kinds of monitors (*e.g.*, traditional and autostereoscopic) as technologies to provide visual context.

2.4.2.12 Projection technologies

This theme groups studies that employ the use of projection-based technologies (*e.g.*, standard projections, powerwalls, CAVE, 360 screens, tabletops, holograms) to provide visual context either by presenting visual content by itself or by blending the visual content with the physical world seamlessly.

2.4.2.13 HMDs

This theme groups studies that employ head-mounted displays (HMDs) to give the user visual context, by means of visual content, and spatial context, either/or by sense of depth (*i.e.*, stereoscopy) and reacting to head motion (*i.e.*, degrees of freedom). It includes Cardboard-like HMDs meant to be used with smartphones, as well as higher-end HMDs meant to be used with computers, and standalone HMDs.

2.4.2.14 Gestural/Somatic tracking technology

This theme groups studies that employ the use of devices for gestural and/or somatic tracking (*e.g.*, gestures, poses, movements) to provide or capture that spatial context. It therefore includes somatic and gestural tracking cameras (*e.g.*, Kinect, Leap Motion), as well as wearables such as data gloves or smart bracelets (*e.g.*, Myo Gesture).

2.4.3 [F8] Techniques to provide more visual and spatial context

2.4.3.1 3D Avatars

This theme groups studies where users are represented through 3D avatars (*e.g.*, patients) in the virtual experience, to provide visual context.

2.4.3.2 3D Illusion

This theme groups studies that employ the use of a technique so that users perceive a sense of depth/3D effect, sometimes referred to as illusion of depth, to provide spatial context. This includes, for instance, dome theater setups where multiple users experience a 3D effect for as long as the projected image reflects the camera motion. It also includes the use of shadows, or HMDs to give the user a sense of depth (the latter through stereoscopy).

2.4.3.3 Agency

This theme groups studies that empower the user with enriched agency, giving them the ability to, for instance, create and follow their own routes, moving a virtual camera for previsualisation, and therefore leveraging spatial context.

2.4.3.4 Superimposition

This theme groups studies that project/show virtual on top of real objects or buildings, such as a 3D replica, or a real instrument, to provide visual context. It thus also includes what's known as projection mapping (*i.e.*, projecting video on top of objects or buildings).

2.4.3.5 3D Modelling

This theme groups studies that employ the use of virtual or physical content created manually or through volumetric, laser scanned, photogrammetry techniques, to provide visual context. It also includes studies using low poly modelling.

2.4.3.6 Cues

This theme groups studies that employ the use of additional sensory cues (stimuli), that indicate the state of some property of the surrounding environment, to feed

the agent's (perceiver of the environment) sensory perception (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)), and therefore provide more visual and spatial context. These may or may not be attended, or, in other words, these cues may be assimilated subconsciously (*e.g.*, visual cues) or consciously (*e.g.*, directed sound). Examples of these include colour and light (or contrast), spatial sound, soundscapes, haptics, and smell.

2.4.3.7 Feedback

This theme groups studies that employ the use of certain additional feedback elements to provide spatial context, such as the use of (spatial) audio, the use of [Light-emitting Diodes \(LEDs\)](#) to enhance the most dominant colours of a painting, and the use of haptics to, for instance, feel the airflow coming from an opened door.

2.4.3.8 Complementary Display

This theme groups studies that employ the use of additional displays/monitors to provide more information or enable further interaction (*e.g.*, touch-based gestures), therefore leveraging visual context.

2.4.3.9 Immersion-reinforcement elements

This theme groups studies that employ the use of elements that reinforce immersion in the environment by leveraging both visual and spatial context. If one considers the taxonomy introduced by Nilsson, later adopted by Beck, Morgado, O'Shea in their immersion definition/review (*q.v.* Section 2.2.4 - [Immersive Environments](#)), these (reinforcement) elements might induce system immersion, narrative immersion, and challenge immersion. Specifically, this theme includes cases such as providing 1-to-1 scale models (contributing to system immersion), or including elements that provide meaning such as crowd simulations, skyboxes, oceans, smell, soundscapes, and lighting (contributing to narrative immersion), or elements such as task boards, and tutorial indications (contributing to challenge immersion).

2.4.3.10 AR (not *in situ*)

This theme groups studies that employ AR, but not *in situ* (*i.e.*, the experience is not dependent and doesn't benefit from the context provided by the surrounding environment), to provide visual context.

2.4.3.11 AR (location-based)

This theme groups studies that employ the use of AR *in situ*, thus leveraging the visual and spatial context of that surrounding environment (*e.g.*, in a specific museum/building or park), which in turn might provide access to physical objects and/or cultural artefacts.

2.4.3.12 Naked eye MR

This theme groups studies that employ the use of AR or MR techniques, providing visual context, without the direct use of a display or camera (thus designed naked eye), but still with the same principles in mind: joining the virtual and the real. A MR case would be, for instance, a projection onto a tabletop and/or wall, allowing the user to touch a real 3D maquette. An AR case would be, for instance, the presence of holograms that can be seen with naked eye.

2.4.3.13 Image tracking/recognition

This theme groups studies that employ the use of image tracking/recognition (including markers/codes, objects, and buildings) to provide more information and/or allow, for instance, superimposition, therefore leveraging visual context.

2.4.3.14 Localisation

This theme groups studies that employ the use of localisation techniques, which provide spatial context, not only for Points of Interest, or POIs (general case), but

also the smell and sounds available in the site, and time (*e.g.*, dusk) of the experience. These include the use of **RFID**, beacons, **GPS**, *etc.*, the use of shadows so that the user can see his/her position clearly and differentiate himself/herself from the others users (*e.g.*, projection-based experiences), or even the use of physical/virtual maps (for guidance or continuous location).

2.4.3.15 Points of Interest (POIs)

This theme groups studies that employ the use of proximity points/markers (can be both in the virtual or physical worlds) to leverage the visual and/or spatial context of that location (*e.g.*, teleport to another virtual place, retrieving more info using a **QR** code).

2.4.3.16 Media

This theme groups studies that employ the use of (overhanging) text, audio, video, animations, and drawings, to provide visual context. Also includes **High Definition (HD)** content and outside (physical) media such as a physical panel. Media in 360° is not included here, but under **Media (360)** instead.

2.4.3.17 Media (360)

This theme groups studies that employ the use of, specifically, 360° content such as images, and video, to provide visual context. Non-360 content is included under **Media**, instead.

2.4.3.18 Physical Location Virtuality

This theme groups studies where the actions and events in the surrounding physical environment are used to impact the virtual components of the experience, thus leveraging visual and/or spatial context. This includes cases that bring the real and

physical into the virtual and give it some meaning (“Incorporating [...] objects, and the parent and child into the story”), rather than the other way around (*e.g.*, AR and MR).

2.4.3.19 Gestural/Somatic tracking

This theme groups studies that employ the use of gestural and/or somatic tracking techniques (*e.g.*, through a camera and motion recognition) to provide spatial context to the user (*e.g.*, guidance when climbing a wall).

2.4.3.20 Spatial Audio

This theme groups studies that employ the use of spatial audio (including directional) to give a sense of space and direction to the user, therefore providing spatial context. They are often used as cues or for feedback purposes.

2.4.3.21 Visual and Motion arrangement

This theme groups studies where the visual and motion arrangement of objects is used to enrich the experience and narrative, through, for instance, showing feedback to the participants, to provide visual context.

2.4.3.22 Visualisation

This theme groups studies that employ the use of visualisation of graphics, data, and motion to leverage visual context by not only showing them, but by also allowing the participants to switch perspectives and/or control them.

2.4.4 [F9] Tech for cultural-aware interaction

2.4.4.1 Blob/Object tracking cameras

This theme groups studies that employ the use of external cameras (*e.g.*, [Infrared \(IR\)](#)) for blob/object tracking. Even if, technically speaking, the tracking of headsets is mostly done through multiple [IR](#) markers, we are not considering those here, but rather in [Room Head-Tracking](#), instead. This theme is more concerned with more accessory and isolated objects that are brought into the experience.

2.4.4.2 Hand controllers

This theme groups studies that employ the use of hand controllers to control some virtual and/or physical behaviour while moving them (*e.g.*, tracing a route). That includes laser pointers, flashlights, hand held pointers, [VR](#) joysticks (such as HTC Vive's Controllers), and even a steering wheel (*e.g.*, camera movement).

2.4.4.3 Hand Gesture Tracking

This theme groups studies that employ the use of technology that detects hand gestures, including, for instance, gloves and camera-based solutions (*e.g.*, leap motion, kinect).

2.4.4.4 Haptic Actuators

This theme groups studies that employ the use of technology to essentially provide feedback and cues to the user through haptic interfaces. That includes, for instance, wearable vests, walkable and air actuators, and gloves.

2.4.4.5 Mobile Devices

This theme groups studies that employ the use of mobile devices (*e.g.*, smartphones, tablets, wearable bands, touch-pads), including the use of the camera for tracking tags, markers and objects, reading QR codes, typical touch-based interaction, also allowing the user to walk and interact with physical objects, cultural artefacts, and POIs.

2.4.4.6 Naked eye MR

This theme groups studies that employ the use of AR/MR technologies to seamlessly blend virtual content and the physical worlds, allowing the user to walk (*e.g.*, through a museum or park) and interact with physical objects, cultural artefacts, and POIs without the need of a headset or display (thus designed naked eye). That includes, for instance, projection-based experiences, and tabletops. It does not include experiences like CAVE, Powerwall, or similar.

2.4.4.7 Body Motion Capture

This theme groups studies that employ the use of full body motion capture solutions, from full-fledged solutions such as the use of motion capture suits, to the use of domestic cameras and sensors such as Microsoft Kinect and Intel RealSense.

2.4.4.8 Stationary Head-Tracking

This theme groups studies that employ the use of 3DoF or stationary 6DoF head tracking technologies, be it through low-cost VR HMDs such as smartphone-based, or experiences where the user is sitting and therefore doesn't move around the room (even if it supports 6DoF). 6DoF experiences are included under Room Head-Tracking, instead.

2.4.4.9 Room Head-Tracking

This theme groups studies that employ the use of **6DoF** room head tracking technologies, be it through the specialised base stations (*e.g.*, in HTC Vive and Oculus Rift scenarios), or camera marker-based solutions for **CAVE**, *etc.*

2.4.4.10 Multi-touch display table

This theme groups studies that employ the use of multi-touch display tables, including tabletops, to control/interact and/or see content of the experience.

2.4.4.11 Voice Control

This theme groups studies that employ the use of voice recognition (*e.g.*, through Alexa's **Application Programming Interface (API)**) and commands to interact with the experience.

2.4.4.12 Gaze Tracking Technologies

This theme groups studies that employ the use of gaze tracking solutions, be it specifically designed, or inferred through the facing direction of the **HMDs** or cameras (*e.g.*, Kinect, mobile devices).

2.4.5 [F10] Type of interaction

In this field, one could argue that we could easily just join all these themes into gestural, somatic, and voice. That is, we could be less specific about it. While that is indeed a fact, and something we later discuss (*q.v.* Section 2.6.4), dividing them this way from the beginning allows us to dive deeper and argue on some technologies and techniques being used together with others more frequently—which is the essence of this subquestion to begin with.

2.4.5.1 Gestural (dynamics)

This theme groups studies that employ the use of HTC Vive Controller-like cases, wands, torches/flashlights, (laser) pointers, *etc.*, to interact with the experience in some way. It also includes moving the controllers/devices (such as smartphones or tablets) in specific ways (*e.g.*, shaking, or pitching), or even moving a toothbrush in a circular fashion.

2.4.5.2 Gestural (hand/finger tracking)

This theme groups studies that employ the use of hand and finger gestures/movements (*e.g.*, pointing finger) to interact with the experience in some way.

2.4.5.3 Gestural (touch-based–tabletop)

This theme groups studies that employ the use of tabletop (multi) touch-based experiences, supporting tangible 3D objects.

2.4.5.4 Gestural (touch-based)

This theme groups studies that employ the use of traditional touch-based mobile interactions (*e.g.*, smartphone, tablet, wearable bands, touchpad), and gestures on the screen/touchpad (*e.g.*, a “V”). This includes gestures on HTC Vive controllers’ touchpad.

2.4.5.5 Somatic (pointing device)

This theme groups studies that employ the use of movements to point a device (*e.g.*, its camera) to a painting, building, QR code, marker, *etc.* In other words, whenever the said movement isn’t really the concern (*i.e.*, not tracked), but rather the way of capturing or lighting up something else.

2.4.5.6 Somatic (physical object)

This theme groups studies that employ the use of specific movements with physical objects, such as positioning a paper or board with markers in front of a camera, handling tags/objects on top of a table (tabletop), or playing with toys.

2.4.5.7 Somatic (body movement)

This theme groups studies that employ the use of walking, poses, arm movements, and head movements (with no specific action, unlike [Somatic \(gaze and head control\)](#)).

2.4.5.8 Somatic (gaze and head control)

This theme groups studies that employ the use of gaze and/or head movement to trigger events/actions or change the experience and/or its content in some way.

2.4.5.9 Voice commands

This theme groups studies that employ the use of user voice-based interaction, through, for instance, the use of Alexa's [API](#).

2.4.6 [F11] Techniques for additional Cultural Context

2.4.6.1 Agency

This theme groups studies that employ the use of agency to leverage the cultural context through empowering the user with choices and the ability to create (*e.g.*, multiple paths) through some sort of interaction—something that the physical world provides by default. An example of that is, for instance, embodied agency such as scavenging and puppeteering. The actual technical means of interaction, which

are the mechanics of agency, are not part of this theme, but rather of [Interaction methods](#), instead.

2.4.6.2 Thematic itineraries

This theme groups studies where the user (agent) can follow a (cultural) thematic route, including museums, tours, etc. When in a more virtual or mixed reality setting, these are often referred to as [VFTs](#).

2.4.6.3 Attentional Shift

This theme groups studies that employ the use of head control to simulate the human capacity of attentional shift, or shift of attention, by triggering some action/event when the user is focusing some direction for some time (*i.e.*, attended stimuli or cues, *q.v.* Section [2.2.2 - Learning and Expressing Complexity](#)). An example of that is focusing a person or [Non-player Character \(NPC\)](#) in a virtual environment, making its audio source louder and the surrounding environment quieter.

2.4.6.4 Cultural Artefact

This theme groups studies that employ the use of an object/item with cultural and historical interest as part of the experience (physical or virtual).

2.4.6.5 Physical Objects

This theme groups studies that employ the use of physical objects that aren't really artefacts but still leverage cultural context through their cultural and social meaning (*e.g.*, a toothbrush).

2.4.6.6 Choreographies

This theme groups studies that employ the use of Visual and Motion arrangement of pieces according to the users' interaction.

2.4.6.7 Cues

This theme groups studies that employ the use of directional and place cues (*e.g.*, visual, aural, emotional, curiosity)—something that we are used to in real life contexts. Emotions (and vision), in particular, are an integral part of us and our cognition function, and previous research has shown a conclusive correlation between emotions and learning achievement (*q.v.* Section 2.2.3 - [Learning in Context](#)).

2.4.6.8 Feedback

This theme groups studies that employ the use of any form of feedback that tries to mimic the physical world (*e.g.*, through haptics).

2.4.6.9 Cultural data/media

This theme groups studies that include or recur to cultural data, information, images and videos, and show them through visualisation or directly.

2.4.6.10 Cultural heritage site

This theme groups studies that employ the use of the surrounding environment (physical or virtual) for not only its cultural value, but also to influence the visual and spatial context dimensions. Examples of that include situated learning through *in situ*/touring/localised experiences.

2.4.6.11 Interaction methods

This theme groups studies that employ the use of methods to enable user interaction that leverage the user's prior symbolic and physical understanding, sometimes deemed "natural", "intuitive", or "fluent" (*q.v.* Section 2.2.5 - [Multimodal Human-Computer Interaction](#)). This includes gestural and somatic movements/gestures/-poses, voice interaction, *etc.*

2.4.6.12 Native Language

This theme groups studies that employ the use of subtitles or multilingual voice. This is particularly important for endangered languages, such as those from some indigenous tribes.

2.4.6.13 Social Context

This theme groups studies that employ the use of social collaboration, communication, discussion, interaction, inside or outside the (virtual) experience.

2.4.6.14 Multi-user experience

This theme groups studies where multiple users are supported in the same experience at the same time, therefore leveraging cultural and [social context](#).

2.4.6.15 Storytelling

This theme groups studies that employ the use of the storytelling technique to go through a more or less established [narrative](#) (actual content of the story). Storytelling is a long-established tradition used 100.000 years ago, when humans started developing language, to transfer knowledge from generation to generation, and thus being such an important and impactful cultural and social activity. The most significant features of storytelling include perspective, [narrative](#), interaction, and medium

(*q.v.* Section 2.2.3 - [Learning in Context](#)).

2.4.6.16 Narrative

This theme groups studies that employ the use of narratives, the actual content of the stories, to connect the events within an experience by covering features such as mimegis/digenesis, time/space, cause/effect, and sequence/plot (*q.v.* Section 2.2.3 - [Learning in Context](#)). It is created based on story objects, characters, and other narrative components. [Storytelling](#), seen as a performance, is one of many ways of presenting a narrative.

2.4.6.17 Background Knowledge

This theme groups studies that employ the users' knowledge acquired in their previous life experience (both mental and physical, such as training-less interaction).

2.4.6.18 Symbols

This theme groups studies that employ the use of symbols, a subclass of signs in the field of semiotics, such as colours and drawings that have a (cultural) meaning tied to them (*i.e.*, a semantic dimension), *q.v.* Section 2.2.5 - [Multimodal Human-Computer Interaction](#).

2.4.6.19 Topic

This theme groups studies that use the topic being learnt and/or taught during the physical (virtual) experience to leverage the cultural context of the experiences.

2.4.6.20 Tradition

This theme groups studies that employ the use of cultural traditions such as dances, music, folklore, and rituals, to leverage the cultural context of the experiences.

2.4.6.21 Visualisation

This theme groups studies where cultural (and social) relevant data or information are represented/visualised in some way. This includes, for instance, crowd simulations, and weather sensors information.

2.4.7 [F12] Complex learning

One might consider this particular column to be among the most crucial ones for this review—perhaps on par with [F11]. We cannot, however, forget how relevant the others are, and how they all influence complex learning. While we tried to separate the fields, it is important to keep in mind that the border between them can often be fuzzy—and repetitions/overlaps will certainly occur. Gathering this item was an attempt to capture the key aspects of each study that, in some way, influence the process of complex learning.

2.4.7.1 Agency

This theme groups studies that employ the use of agency to empower the user with choices that end up affecting the learning experience, including, for instance, following multiple routes (often referred to as VFTs), choosing from different learning modes, and giving the ability to create (*e.g.*, multiple paths). It also includes examples of embodied agency such as scavenging and puppeteering, where the user (agent) embodies something or someone and takes part in the hands-on simulated experience and scenario (*i.e.*, leveraging context). Though often overlooked (*q.v.* Section 2.2.3 - [Learning in Context](#)), context shapes learning (mostly) through the

surrounding environment, particularly through social interactions and culture. Simulating real life contexts is one out of the three challenges (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)) that effective complex learning environments need to tackle. Generally speaking, simulating a real context implies more agency, which in turn implies more interaction and empowering of the user. The outgoing result is essentially learning by doing, much like it happened before the industrial revolution (*i.e.*, situated theory and learning).

2.4.7.2 Augmented content

This theme groups studies that employ the use of any form of augmentation to improve the learning experience. That includes, for instance, text, 3D models and their superimposition, animations, visual anthropology, and intangible heritage. Augmented content and AR/MR, in general, bridge the gap between the virtual and real by having the (default) possibility of bringing the surrounding environment (and its context) to the virtual experience. AR and MR open up new opportunities for both supporting and facilitating complex learning (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)).

2.4.7.3 Social Context

This theme groups studies that employ the use of a form of [active learning](#) through multiple participants (outside or inside the experience), inciting discussions, engagement, and cooperation, to leverage social context. Humans have an inherent need for a social context in which “knowledge can be understood and assessed”, and acquiring soft skills (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)), such as collaboration (*e.g.*, peer-learning), communication, and task coordination, is one of the three aspects that should be targeted for effective learning environments for complex knowledge domains (*e.g.*, [STEM](#)).

2.4.7.4 Visualisation

This theme groups studies that employ the use of visualisations (one of the most utilised methods to deal with complexity in the past, *q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)) to incite learning through the experiences (*e.g.*, 3D, VR). This includes, for instance, concept maps, astrovisualisations, and documentary-like experiences (*e.g.*, 360, video-guide in [AR](#)).

2.4.7.5 Gamification

This theme groups studies that employ the use of so-called “serious”, “city-wide”, “edutainment”, playful/gamified experiences. When particularly coupled with semantics, it takes realism to a whole new level and impacts positively the learning aspect of the game experience (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)).

2.4.7.6 Guidance

This theme groups studies that aid the user throughout the experience in some way. It includes, for instance, cases that employ the use of projected instructions, [NPCs](#), and multimodal cues. Situated environments (context), is arguably the best learning process one can have. But, depending on the subject and topic being learnt, it might be more complicated to do alone. Guidance can be seen as a form of scaffolding, and is essential, as argued originally by Vygotsky in his [CHAT](#) (*q.v.* Section 2.2.3 - [Learning in Context](#)), to attain learning, otherwise unachievable, through the [ZPD](#), *q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#).

2.4.7.7 Situated learning

This theme groups studies that employ the use of situated learning, with more or less context. It thus includes studies leveraging not only the physical dimension, but also the cultural, social, and mental (*q.v.* Section 2.2.3 - [Learning in Context](#)),

with the ability to present information in a contextualised manner (*e.g.*, through localisation).

2.4.7.8 Tasks

This theme groups studies where the learning (also) happens by defining a set of tasks or goals for the user to accomplish. Learning tasks are the foundation of multiple (complex) learning methods to support the construction of schemata, which in turn contributes to developing adaptive expertise (one of the three challenges/requirements of complex learning, *q.v.* Section 2.2.2 - Learning and Expressing Complexity). Examples of that (*q.v.* Section 2.2.2 - Learning and Expressing Complexity) are CPS, Complex Constructivism, Connectivism, 4C/ID, and problem-based learning, or PBL. Tasks are a part of work done as part of one's duties, while the problem is a difficulty (challenge) that has to be resolved or dealt with, often creatively.

2.4.7.9 First-person simulation

This theme groups studies where the user is part of the simulation itself (*i.e.*, an agent), rather than just an observer. First-person is how we see the world, and using that mode rather than, for instance, a virtualised third-person, might yield better results by bringing more realism to the experience. Doing so in the physical world (first-person by default) is the ideal way, but is often impractical for security and monetary constraints. (Virtual) Simulations of this kind have the advantage of being run over and over again, which in turn may not only increase the acquired learning in a shorter span of time, but also be performed in a safe way (*q.v.* Section 2.2.2 - Learning and Expressing Complexity).

2.4.7.10 Simulation (non-participant)

This theme groups studies where simulations (*e.g.*, temporal, forecast) are run within the experience. Simulations are often infeasible to run in the physical world, due to

the myriad of variables available and ambiguous feedback (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)). Being able to simulate them in virtual environments creates a more isolated and controlled environment, which can be extremely useful to see and understand its dynamics.

2.4.7.11 Storytelling

This theme groups studies where the learning happens (also) by listening to storytelling, that in turn goes through a narrative (actual content of a story). Storytelling is a long-established tradition used 100.000 years ago, when humans started developing language, to transfer knowledge from generation to generation, and thus being such an important and impactful cultural and social activity. It can also be used for entertainment only, but this theme groups a more serious use of storytelling that can be referred to as serious storytelling, with perspective, narrative, interaction, and medium being its most significant features (*q.v.* Section 2.2.3 - [Learning in Context](#)).

2.4.7.12 Perspective

This theme groups studies that employ the use of different perspectives (*e.g.*, eagle-eye, participant & viewer), helping the participants understand which problems are big and which are not. Chances are that the perspectives of the one teaching and the one learning are different, and it is critical to human learning to match them (*q.v.* Section 2.2.3 - [Learning in Context](#)).

2.4.7.13 Active learning

This theme groups studies where the user has an active role (constructivism and reflective practice) in the learning process, rather than merely being presented with information (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)). Active learning is broad and includes more focused methods within. Examples of that

are self-guided learning (*e.g.*, by empowering the user with creation tools), [discovery](#), [problem-based learning \(PBL\)](#), [project-based learning](#), and other forms of collaborative (team-based) learning.

2.4.7.14 Embodied Learning

This theme groups studies where the user has to embody something else through, for instance, puppeteering, and scavenging, to participate in the experience and learn from it in a more realistic context and environment. Using head-mounted displays (HMDs) such as in [VR](#) or [MR](#) enables participants, almost by default, to have more embodied experiences by creating a [first-person perspective](#) within virtual scenes (*q.v.* Section [2.2.5 - Multimodal Human-Computer Interaction](#)).

2.4.7.15 Problem-based Learning (PBL)

This theme groups studies that employ [PBL](#), a form of active learning, where complex, ill-structured, real-world problems provide context for the learning process (*q.v.* Section [2.2.2 - Learning and Expressing Complexity](#)).

2.4.7.16 Project-based Learning

This theme groups studies that employ project-based learning, a form of active learning, where a major project provides context for the learning process (*q.v.* Section [2.2.2 - Learning and Expressing Complexity](#)).

2.4.7.17 Discovery

This theme groups studies where the users visit, for instance, monuments, and [POIs](#), that essentially employ a learning method known as Discovery, a form of active learning (*q.v.* Section [2.2.2 - Learning and Expressing Complexity](#)). It also includes cases where the users go out (*e.g.*, in a scavenger-like fashion), and studies where

the user explores concepts (*e.g.*, through visualisation). From a constructivist point of view, managing learning can be accomplished by giving the participants a set of prescribed directions that create a vision, and clarified goals, while still leaving room for “visionary initiatives and the emergence of creative trajectories (*e.g.*, result from collective learning and shared culture building)” (*q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#)).

2.4.7.18 Feedback

This theme groups studies where feedback is given more importance to leverage complex learning, *q.v.* Section 2.2.2 - [Learning and Expressing Complexity](#) (*e.g.*, showing continuous user location, bidirectional info between real and virtual, haptics).

2.4.7.19 Interaction (HCI)

This theme groups studies where [HCI](#) has more relevance (or reasoning) to leverage complex learning, by, for instance, enabling personalised interaction modes. Often, more interaction mechanics result in more [agency](#), which in turn better represents a real setting and its context (*i.e.*, physical world) and may accelerate the learning process (*q.v.* Section 2.2.5 - [Multimodal Human-Computer Interaction](#)).

2.4.7.20 Multimodal

This theme groups studies that employ the use of multiple forms of input and output (multimodal experiences, sometimes referred to as multimodal [HCI](#)) to leverage the learning process (*q.v.* Section 2.2.5 - [Multimodal Human-Computer Interaction](#)).

2.4.7.21 Individualised learning

This theme groups studies that employ the use of individualised/personalised learning (*e.g.*, through 1 to 1 mentorship, enabling the learner to pick personalised content, change interaction modes). The outcome of this personalised learning experience bridges the gap needed to achieve otherwise unachievable learning results by having ZPD (*q.v.* Section 2.2.2 - Learning and Expressing Complexity).

2.4.7.22 Scaffolding

This theme groups studies where the user is given the necessary support throughout the learning process (*i.e.*, sort of a scaffolding structure to build on). The concept of ZPD, are often seen as scaffolding to attain otherwise unachievable learning results or actions (*q.v.* Section 2.2.2 - Learning and Expressing Complexity).

2.4.7.23 Agent Analysis

This theme groups studies where the performance of the user (agent) is analysed and used to improve not only the experience but also to support the user further with feedback (through self management and reflective expertise). The outcome with this sort of analysis is to essentially guide the user (*i.e.*, guidance), which supports the use of ZPD (*q.v.* Section 2.2.2 - Learning and Expressing Complexity) to attain otherwise unachievable learning results.

2.4.8 [F13] Target

2.4.8.1 General Public

This theme groups studies that target the general public, or common civilians, such as visitors, tourists, and citizens, with no age restrictions (*i.e.*, can range from children to adults). In other words, the experience doesn't really have a specific target

group—it is not tailored—, so the experience within can vary, but is not accounted for—let alone leveraged.

2.4.8.2 Limited mobility

This theme groups studies that target people with limited mobility and/or accessibility.

2.4.8.3 Students

This theme groups studies that specifically target students, with or without specific age groups and intervals.

2.4.8.4 Children

This theme groups studies that specifically target children, with or without specific age groups and intervals.

2.4.8.5 Parents

This theme groups studies targeting parents in a joint experience with their children.

2.4.8.6 Technology Novice

This theme groups studies that target people with little to no experience with the user interaction methods ([HCI](#)) employed in the experience. In other words, following the Dreyfus' skill acquisition model (*q.v.* Sections [2.2.2 - Learning and Expressing Complexity](#), and [2.2.3 - Learning in Context](#)), it translates to users that are considered novice.

2.4.8.7 Technology Experienced

This theme groups studies that target people with a good amount of experience with user interaction (HCI). In other words, it includes users that have already tried the employed, or similar, experience equipment (*e.g.*, touch-based mobile devices, VR-based equipment, controllers, gestural and somatic tracking devices). Following the Dreyfus' skill acquisition model (*q.v.* Sections 2.2.2 - Learning and Expressing Complexity, and 2.2.3 - Learning in Context), it could translate to a skill level ranging from advanced beginner, to expert. Similarly to Topic Professional, one cannot assess the skill level of the users with such depth, and thus grouping all these levels into a single theme.

2.4.8.8 Topic Novice

This theme groups studies that target people with little to no experience in the topic being taught or learnt through the described experiences, including Students, and Children. In other words, following the Dreyfus' skill acquisition model (*q.v.* Sections 2.2.2 - Learning and Expressing Complexity, and 2.2.3 - Learning in Context), it translates to users that are considered novice.

2.4.8.9 Topic Advanced Beginner

This theme groups studies that target people with some experience and situational recollection in the topic being taught or learnt through the described experiences. It includes semi-professionals (*e.g.*, training soldiers) and more advanced students (*e.g.*, practising a musical instrument for a few years but not at a professional level). In other words, following the Dreyfus' skill acquisition model (*q.v.* Sections 2.2.2 - Learning and Expressing Complexity, and 2.2.3 - Learning in Context), it translates directly to users that are considered advanced beginners.

2.4.8.10 Topic Professional

This theme groups studies that target people with a professional level and/or experience in the topic being taught or learnt through the described experiences. It includes professionals and other topic-knowledgeable users (*e.g.*, in healthcare, multimedia, military, science). Following the Dreyfus' skill acquisition model (*q.v.* Sections 2.2.2 - Learning and Expressing Complexity, and 2.2.3 - Learning in Context), professionals can range from competent, to proficient, to finally expert. Without such details in the studies, however, there is no reasonable way to assess with such specificity what level of skill these experiences were targeting, and thus grouping the three of them into Topic Professional.

2.5 Results

This section summarises the results of this SLR. In particular the results yielded from the quality evaluation, and the TA results. It is important to note that the totalling results from each TA theme shown throughout this section are distinctive. That is, during data collection and consequent thematic analysis, it was possible for a study to have multiple (coding) entries refer to the same theme. Since that was not exhaustively pursued, however, and to avoid possible biases, the results throughout this section show the total of unique studies within a given theme.

2.5.1 Quality evaluation

We assessed the studies for quality using some determined and literature-based criteria (*q.v.* Section 2.3.4 - Quality assessment), which yielded a total quality average of 5.11 out of 7 (*q.v.* Table A.2 - Quality evaluation of included studies). We also investigated the possible relationship between the quality score of a primary study and the year when it was published. The score for each study is shown in Figure 2.2, below, followed by the average quality score for the studies within a given year (*q.v.* Table 2.9).

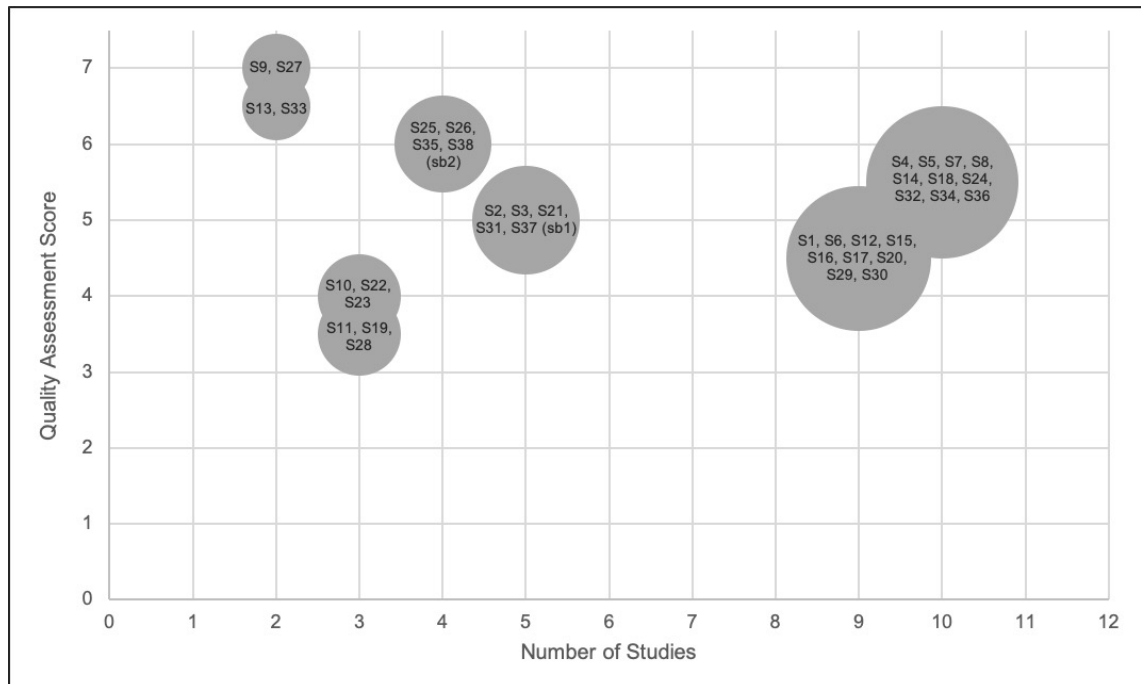


Figure 2.2 – Quality Assessment Scores of primary filtered studies based on Mahdavi-Hezavehi et al. (2017).

Table 2.9 – Average quality score for studies by publication date.

	Year							
	2013	2014	2015	2016	2017	2018	2019	Total
Number of studies	5	1	7	6	7	8	4	38
Mean quality score	4.80	5.00	5.00	5.00	5.14	5.31	5.38	5.11
Standard deviation of quality score	0.97	N/A	1.00	0.77	1.18	1.00	0.63	0.91

Considering that the search included published studies until June 18th, 2019 (*q.v.*, subsection 2.3.3 - Inclusion and exclusion criteria), Table 2.9, below, indicates that the number of studies published per year, with the exception of 2014, has been quite stable. The yearly average quality score appears to be increasing. Spearman's rank coefficient ($r_s=0.24$) tells us, indeed, that there is a slight positive correlation between the two variables (year and quality). Using r_s to calculate the test statistic ($t=1.46$) and two tailed probability value ($p=0.15$) confirms, however, that this correlation is not statistically significant ($p \geq 0.05$). Thus, the change in quality cannot be attributed to the years.

2.5.2 Thematic Analysis Results

2.5.2.1 F6 - Complex topic being learnt/taught

The field F6, regarding the complex topic being learnt and/or taught, has yielded nine (9) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.1). The prevalence of each topic is presented in the chart below (*q.v.* Figure 2.3).

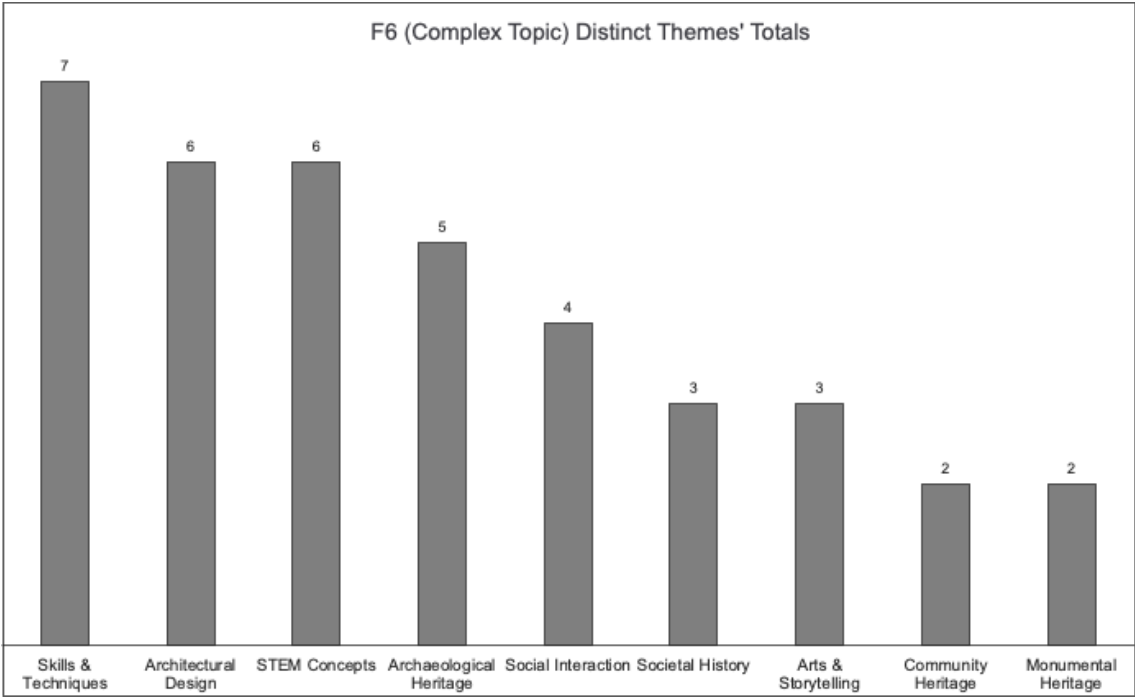


Figure 2.3 – F6 (Complex Topic) Distinct Themes' Totals.

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.10, below).

2.5.2.2 F7 - Technologies for Visual and Spatial Context

The field F7, regarding the technologies being used to provide visual and spatial context, has yielded fourteen (14) distinctive themes—all of which defined in the

Table 2.10 – Studies included in each F6 theme.

Theme	Studies
Archaeological Heritage	S2, S30, S31, S33, S36
Architectural Design	S9, S15, S16, S20, S23, S25
Arts & Storytelling	S7, S13, S22
Community Heritage	S6, S14
Monumental Heritage	S26, S29
Skills & Techniques	S3, S5, S8, S12, S18, S34, S35
Social Interaction	S4, S10, S17, S24
Societal History	S1, S21, S32
STEM Concepts	S11, S19, S27, S28, S37 (sb1), S38 (sb2)

previous section (*q.v.* Section 2.4.2).

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.11, below).

Table 2.11 – Studies included in each F7 theme.

Theme	Studies
360 cameras	S6
3D printers	S25
Fixed Cameras for Markers	S7, S12, S15
Gestural/Somatic Tracking	S1, S2, S3, S4, S5, S6, S7, S8, S10, S12, S13, S17, S18, S24, S25, S26, S33, S35, S37 (sb1), S38 (sb2)
HMDs	S2, S5, S6, S8, S18, S22, S23, S27, S34, S35, S37 (sb1), S38 (sb2)
Markers	S12, S15, S19, S23, S30
Mobile devices	S3, S5, S6, S7, S9, S10, S11, S12, S14, S16, S19, S20, S23, S24, S26, S27, S29, S30, S32, S34, S36
Monitors	S15, S31
Neurophysiologic sensing technology	S10, S38 (sb2)
Positioning and Orientation technology	S9, S10, S20, S26, S29, S30, S32, S36
Projection technologies	S1, S4, S6, S7, S12, S13, S17, S21, S24, S25, S28, S33, S34, S35, S36, S38 (sb2)
Smart objects	S24
Speakers	S1, S4, S6, S17, S24, S25, S26, S36, S37 (sb1), S38 (sb2)
Weather sensors	S19

2.5.2.3 F8 - Techniques for more visual and spatial context

The field F8, regarding the techniques being employed to provide more visual and spatial context, has yielded twenty two (**22**) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.3).

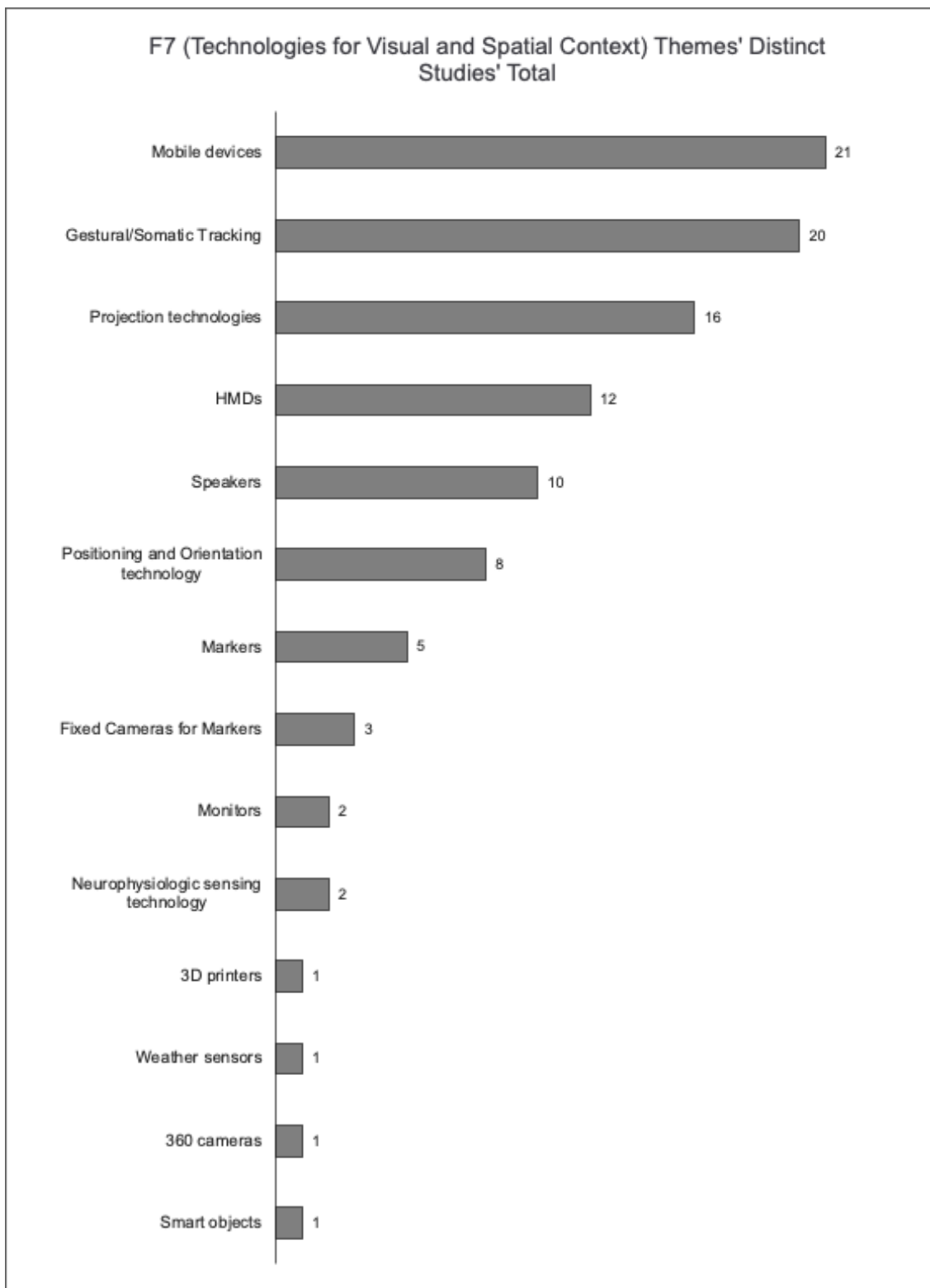


Figure 2.4 – F7 (Technologies for Visual and Spatial Context) Themes' Distinct Studies' Totals

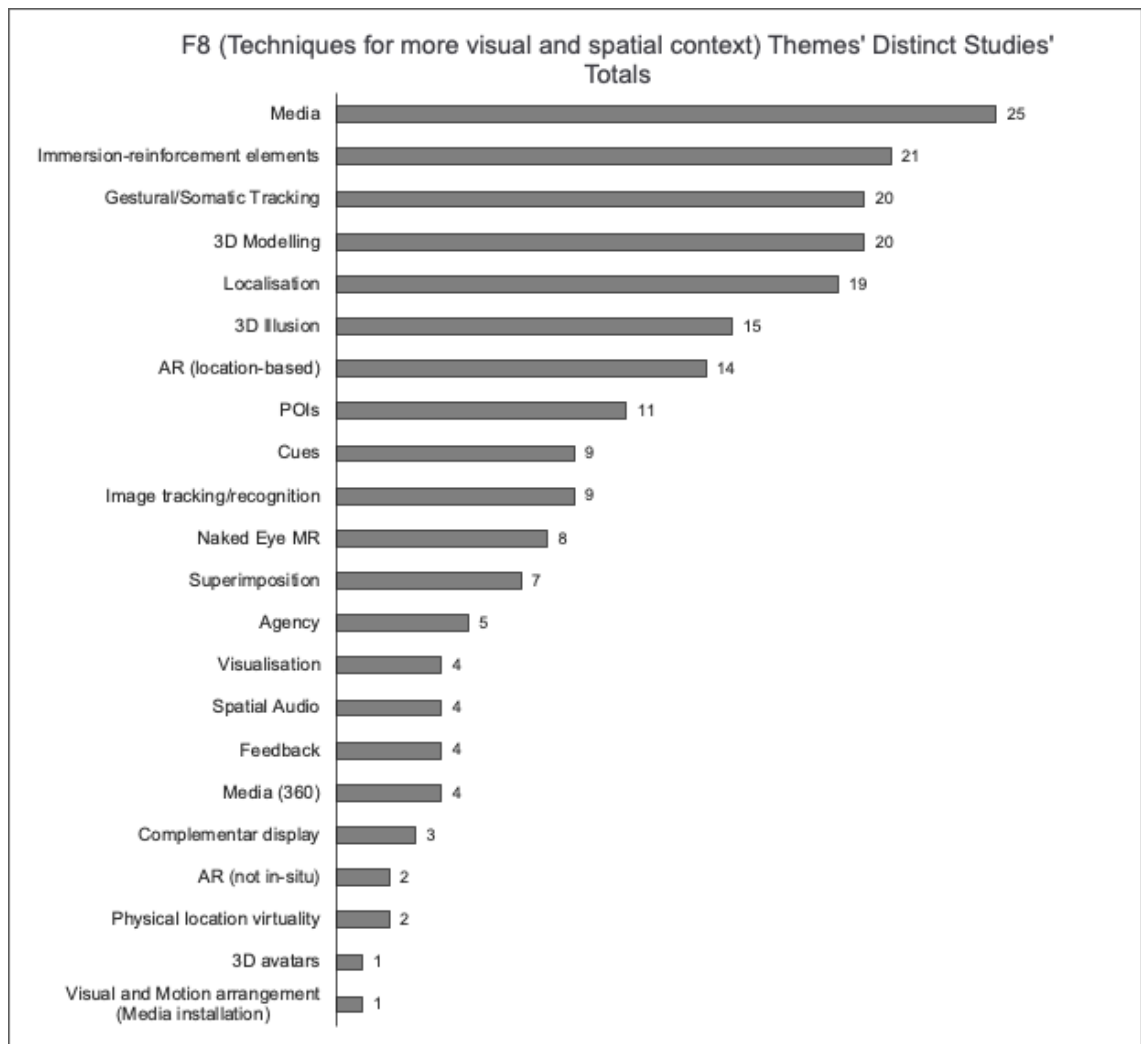


Figure 2.5 – F8 (Techniques for more Visual and Spatial Context) Themes' Distinct Studies' Totals.

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.12, below).

Table 2.12 – Studies included in each F8 theme.

Themes	Studies
3D avatars	S38 (sb2)
3D Illusion	S1, S18, S2, S22, S23, S27, S28, S31, S34, S35, S37 (sb1), S38 (sb2), S5, S6, S8
3D Modelling	S1, S15, S16, S18, S2, S20, S23, S25, S26, S29, S30, S31, S34, S35, S37 (sb1), S38 (sb2), S4, S5, S8, S9
Agency	S16, S18, S5, S8, S9
AR (location-based)	S11, S14, S15, S16, S19, S20, S26, S29, S3, S30, S32, S34, S7, S9
AR (not in-situ)	S11, S23
Complementar display	S15, S28, S31
Cues	S17, S22, S25, S26, S33, S36, S37 (sb1), S4, S6
Feedback	S10, S17, S22, S26
Gestural/Somatic Tracking	S1, S10, S12, S13, S17, S18, S2, S24, S25, S26, S3, S33, S35, S37 (sb1), S38 (sb2), S4, S5, S6, S7, S8
Image tracking/recognition	S11, S12, S15, S16, S20, S23, S26, S30, S32
Immersion-reinforcement elements	S12, S13, S14, S16, S18, S22, S23, S24, S25, S26, S27, S28, S29, S30, S33, S35, S36, S37 (sb1), S4, S5, S9
Localisation	S10, S11, S14, S15, S20, S22, S25, S26, S29, S30, S32, S33, S34, S36, S38 (sb2), S4, S5, S7, S9
Media	S1, S12, S14, S16, S19, S2, S22, S23, S24, S25, S27, S28, S29, S3, S30, S31, S32, S33, S34, S35, S36, S37 (sb1), S38 (sb2), S4, S9
Media (360)	S19, S27, S34, S6
Naked Eye MR	S10, S12, S13, S21, S24, S25, S35, S36
Physical location virtuality	S12, S24
POIs	S14, S19, S22, S23, S29, S30, S32, S33, S36, S7, S9
Spatial Audio	S17, S22, S26, S6
Superimposition	S1, S13, S20, S25, S26, S29, S30
Visual and Motion arrangement (Media installation)	S17
Visualisation	S1, S19, S35, S38 (sb2)

2.5.2.4 F9 - Technologies for culture-aware interaction

The field F9, regarding the technologies being employed to provide cultural-aware interaction, has yielded twelve (**12**) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.4).

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.13, below).

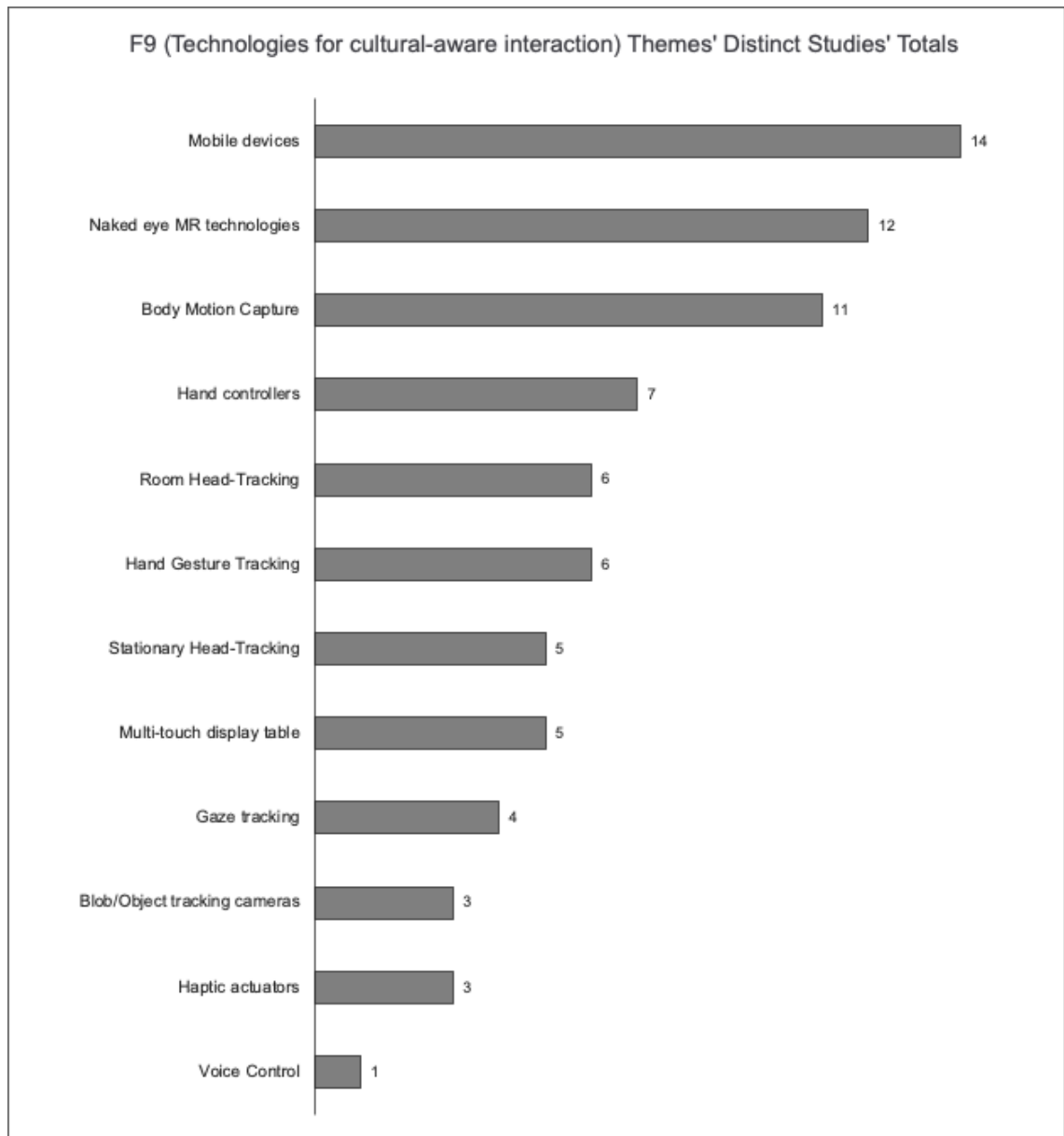


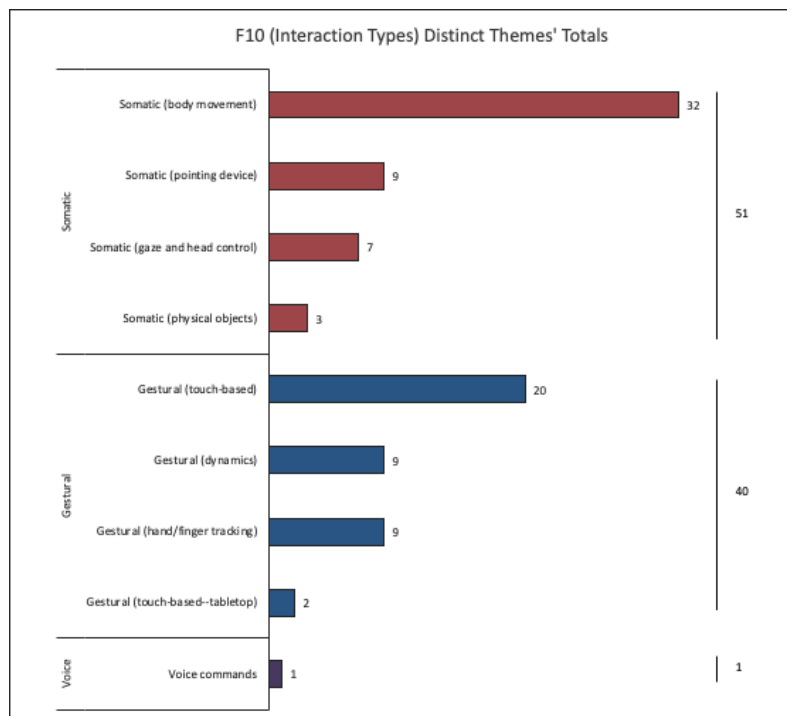
Figure 2.6 – F9 (Techniques for Cultural-aware interaction) Themes' Distinct Studies' Totals.

Table 2.13 – Studies included in each F9 theme.

Tentative Theme	Studies
Blob/Object tracking cameras	S12, S25, S37 (sb1)
Body Motion Capture	S1, S10, S12, S17, S2, S24, S3, S33, S35, S6, S7
Gaze tracking	S20, S23, S5, S6
Hand controllers	S18, S25, S33, S36, S37 (sb1), S5, S8
Hand Gesture Tracking	S10, S13, S26, S3, S38 (sb2), S4
Haptic actuators	S10, S22, S26
Mobile devices	S11, S12, S14, S15, S16, S19, S20, S23, S29, S30, S32, S5, S7, S9
Multi-touch display table	S21, S25, S28, S31, S34
Naked eye MR technologies	S1, S10, S12, S13, S17, S21, S24, S25, S34, S35, S36, S7
Room Head-Tracking	S18, S22, S37 (sb1), S38 (sb2), S5, S8
Stationary Head-Tracking	S2, S23, S27, S34, S6
Voice Control	S24

2.5.2.5 F10 - Interaction Types

The field F10, regarding the interaction type, has yielded nine (9) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.5).

**Figure 2.7** – F10 (Interaction Types) Themes' Distinct Studies' Totals.

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.14, below).

Table 2.14 – Studies included in each F10 theme.

Tentative Theme	Studies
Gestural (dynamics)	S11, S18, S24, S25, S32, S33, S37 (sb1), S5, S8
Gestural (hand/finger tracking)	S10, S13, S21, S24, S26, S3, S38 (sb2), S4, S7
Gestural (touch-based–tabletop)	S21, S25
Gestural (touch-based)	S11, S12, S14, S16, S18, S19, S20, S23, S26, S27, S28, S29, S30, S31, S32, S34, S5, S7, S8, S9
Somatic (body movement)	S1, S10, S11, S12, S14, S15, S16, S17, S18, S19, S2, S20, S22, S23, S24, S25, S26, S27, S29, S3, S30, S32, S33, S34, S35, S36, S37 (sb1), S38 (sb2), S5, S7, S8, S9
Somatic (gaze and head control)	S2, S20, S23, S27, S34, S5, S6
Somatic (physical objects)	S15, S21, S24
Somatic (pointing device)	S11, S14, S23, S26, S29, S30, S32, S36, S7
Voice commands	S24

2.5.2.6 F11 - Techniques for more cultural context

The field F11, regarding the techniques being employed to provide more cultural context, has yielded twenty one (**21**) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.6).

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.15, below).

2.5.2.7 F12 - Complex Learning techniques

The field F12, regarding the complex learning techniques being employed, has yielded twenty-three (**23**) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.7).

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.16, below).

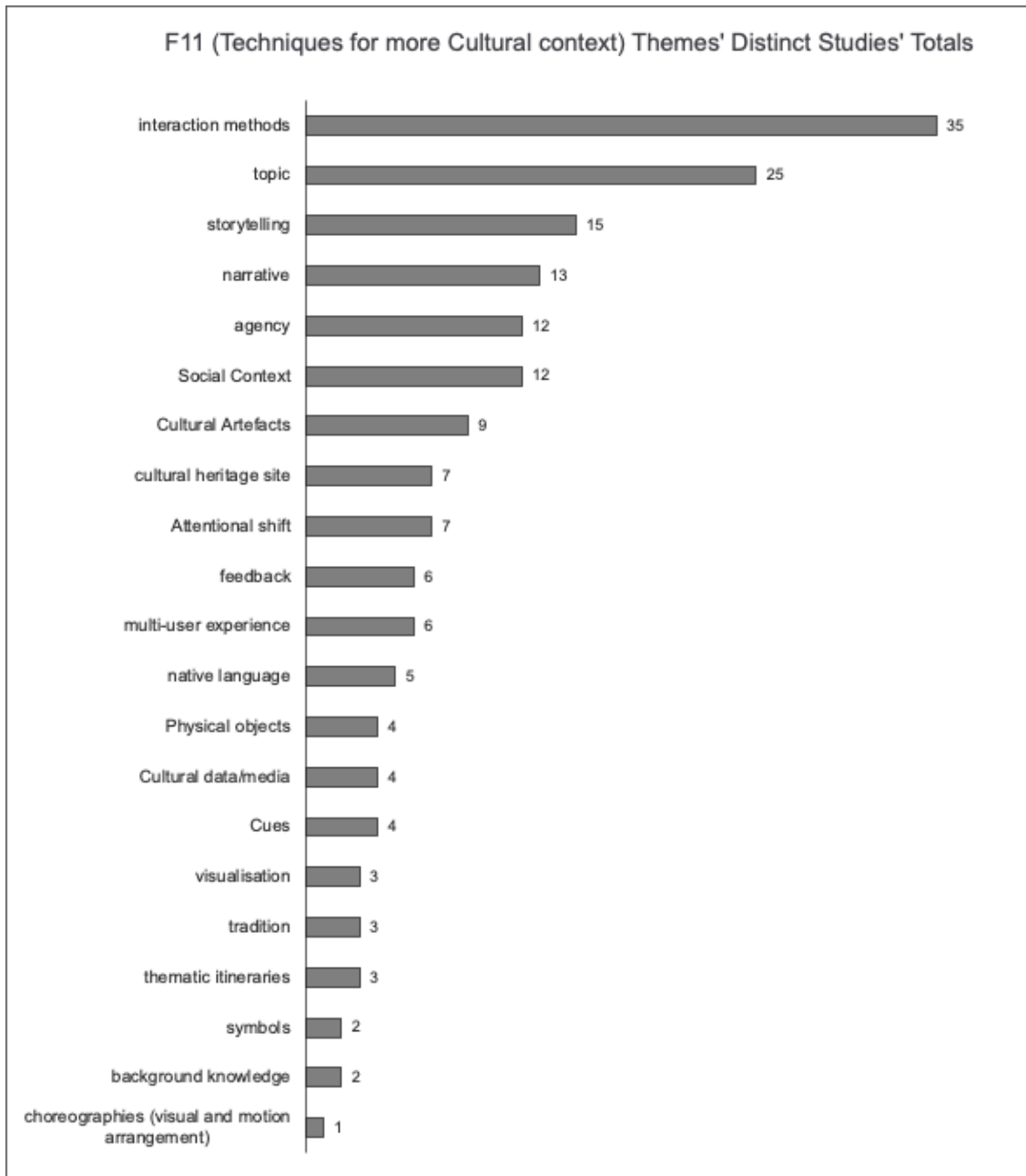


Figure 2.8 – F11 (Techniques for more Cultural Context) Themes' Distinct Studies' Totals.

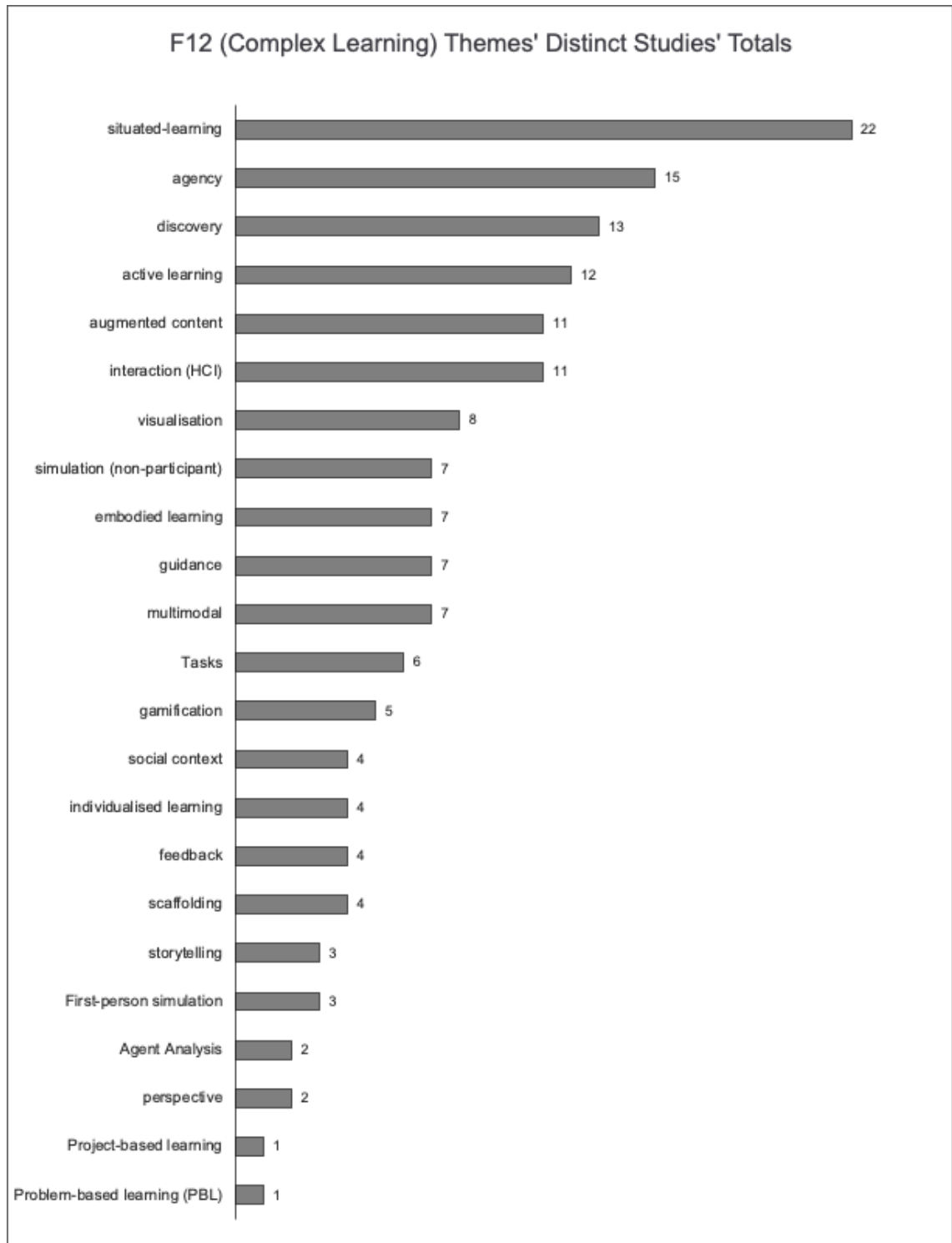


Figure 2.9 – F12 (Complex Learning) Themes' Distinct Studies' Totals.

Table 2.15 – Studies included in each F11 theme.

Tentative Theme	Studies
agency	S11, S16, S18, S22, S24, S26, S32, S36, S5, S7, S8, S9
Attentional shift	S2, S20, S23, S27, S34, S5, S6
background knowledge	S25, S4
choreographies (visual and motion arrangement)	S17
Cues	S19, S22, S26, S27
Cultural Artefacts	S1, S14, S15, S16, S19, S20, S33, S35, S7
Cultural data/media	S34, S5, S6, S7
cultural heritage site	S2, S20, S25, S29, S36, S5, S6
feedback	S10, S19, S2, S22, S26, S38 (sb2)
interaction methods	S1, S10, S11, S12, S13, S14, S15, S17, S18, S19, S2, S20, S21, S22, S23, S24, S25, S26, S27, S29, S3, S30, S31, S32, S34, S35, S36, S37 (sb1), S38 (sb2), S4, S5, S6, S7, S8, S9
multi-user experience	S21, S28, S32, S33, S34, S37 (sb1)
narrative	S1, S11, S14, S19, S27, S3, S32, S34, S35, S36, S4, S5, S6
native language	S14, S17, S19, S32, S6
Physical objects	S23, S24, S25, S36
Social Context	S10, S11, S14, S21, S28, S3, S32, S33, S34, S36, S37 (sb1), S38 (sb2)
storytelling	S1, S11, S14, S19, S22, S24, S27, S28, S3, S32, S34, S35, S36, S6, S9
symbols	S14, S19
thematic itineraries	S16, S5, S9
topic	S1, S11, S13, S14, S15, S16, S17, S18, S19, S2, S20, S23, S25, S26, S29, S30, S31, S32, S33, S34, S35, S36, S5, S6, S9
tradition	S17, S3, S6
visualisation	S19, S2, S5

Table 2.16 – Studies included in each F12 theme.

Tentative Theme	Studies
agency	S11, S13, S16, S17, S18, S22, S24, S26, S32, S36, S5, S6, S7, S8, S9
augmented content	S1, S11, S15, S16, S23, S29, S30, S35, S6, S7, S9
active learning	S14, S17, S18, S19, S25, S27, S3, S33, S35, S5, S8, S9
embodied learning	S18, S19, S29, S32, S33, S36, S8
social context	S21, S28, S32, S36
visualisation	S11, S20, S23, S28, S37 (sb1), S38 (sb2), S6, S9
gamification	S11, S20, S21, S29, S4
guidance	S12, S13, S22, S24, S28, S33, S4
situated-learning	S1, S14, S15, S16, S19, S2, S20, S23, S25, S26, S29, S3, S30, S31, S34, S35, S36, S37 (sb1), S4, S5, S7, S9
Tasks	S18, S24, S25, S27, S33, S9
storytelling	S1, S14, S24
discovery	S11, S14, S18, S20, S23, S25, S29, S30, S32, S33, S34, S35, S6
perspective	S37 (sb1), S6
feedback	S10, S25, S38 (sb2), S5
interaction (HCI)	S10, S13, S2, S24, S28, S31, S33, S37 (sb1), S38 (sb2), S6, S7
individualised learning	S12, S35, S6, S7
multimodal	S22, S24, S26, S28, S33, S35, S36
scaffolding	S12, S13, S27, S31
Problem-based learning (PBL)	S12
Agent Analysis	S10, S12
First-person simulation	S10, S12, S13
simulation (non-participant)	S11, S12, S2, S21, S3, S37 (sb1), S5
Project-based learning	S9

2.5.2.8 F13 - Target

The field F13, regarding the target public of the experiences, has yielded eleven (**11**) distinctive themes—all of which defined in the previous section (*q.v.* Section 2.4.8).

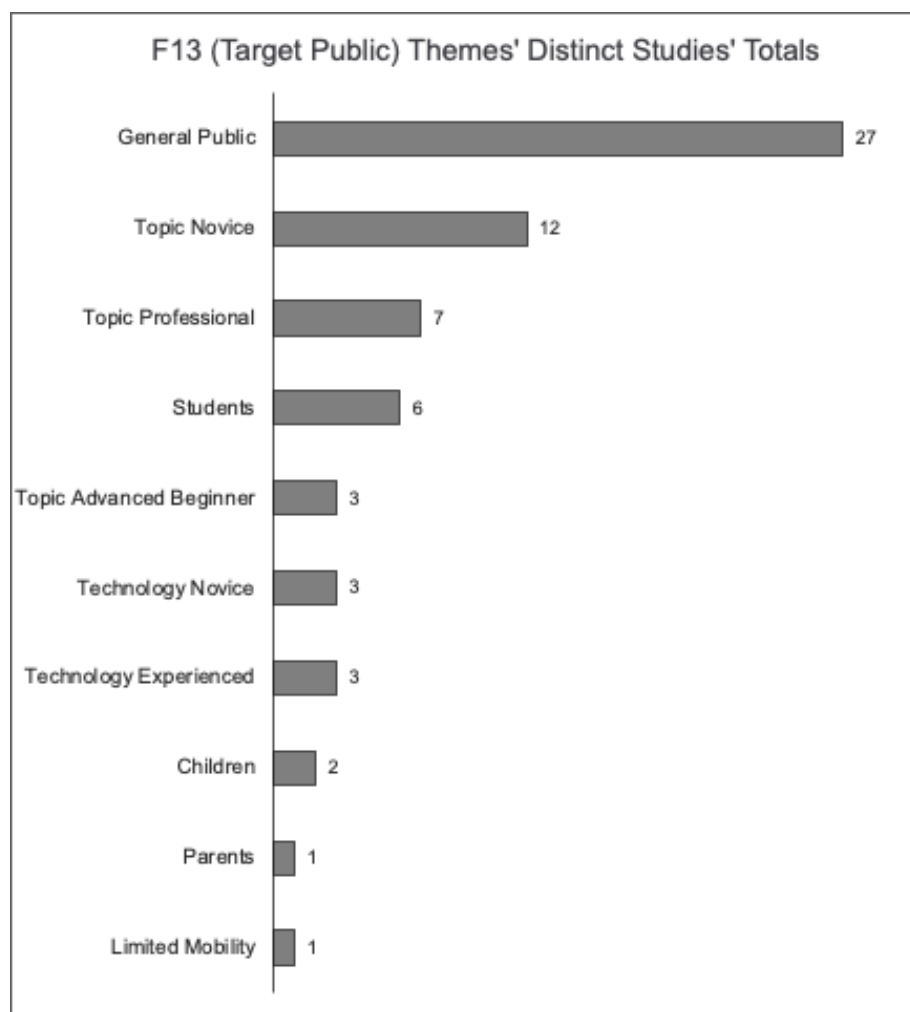


Figure 2.10 – F13 (Target Public) Themes' Distinct Studies' Totals.

The specific studies within each particular theme are listed in the table below (*q.v.* table 2.17, below).

Table 2.17 – Studies included in each F13 theme.

Tentative Theme	Studies
Children	S24, S28
General Public	S1, S11, S14, S15, S16, S17, S2, S20, S22, S23, S25, S26, S28, S29, S3, S30, S31, S32, S33, S35, S36, S4, S5, S6, S7, S8, S9
Limited Mobility	S2
Parents	S24
Students	S14, S19, S2, S21, S38 (sb2), S9
Technology Experienced	S34, S38 (sb2), S7
Technology Novice	S18, S34, S38 (sb2)
Topic Advanced Beginner	S10, S13, S8
Topic Novice	S11, S12, S13, S14, S19, S2, S21, S24, S27, S28, S38 (sb2), S9
Topic Professional	S10, S18, S2, S37 (sb1), S38 (sb2), S4, S8

2.6 Discussion

In this section, we discuss the answers to our research questions (*q.v.* Section 2.3.1), by further synthesising the results and presenting summaries and/or classifications.

2.6.1 F6 - What is the complex topic being learnt/taught?

With the distinctive theme totals of F6 (*q.v.* Section 2.5.2.1, specifically *q.v.* Figure 2.3), one can arguably join them further into a total of two (2) more general clusters already brought up in the background (*q.v.* Section 2.2.1 - Complexity):

- (Formal and natural) **Sciences** (*e.g.*, physics, mathematics, biology), **Medicine**, and **everyday life** aspects (*e.g.*, traffic jams, market crashes);
- **Humanities**, **Arts**, and **Social Sciences** (**HASS**), including emergent phenomena that resulted from the evolution of the biosphere such as life, agency, value, and meaning.

The chart below (*q.v.* figure 2.11) shows us that complexity is equally distributed between these two clusters, with 19 studies each. Despite also suggesting a higher diversity within the **HASS** cluster, one cannot forget **STEM** Concepts' **theme** is included within **Sciences**, **Medicine**, and **everyday life**, that if scrutinised could turn into 4 other themes (Science, Technology, Engineering, and Mathematics) and

therefore actually balance out the diversity aspect (*i.e.*, 6 themes within each cluster). This diversity of fields agrees with the literature in the sense that complexity surrounds us and has several possible applications across health, technology, humanities, arts, and (social) sciences (*q.v.* Section 2.2.1 - Complexity). But the particular high preponderance of HASS could be an unwanted result of including a cultural category in the search string (*q.v.* Section 2.3.2).

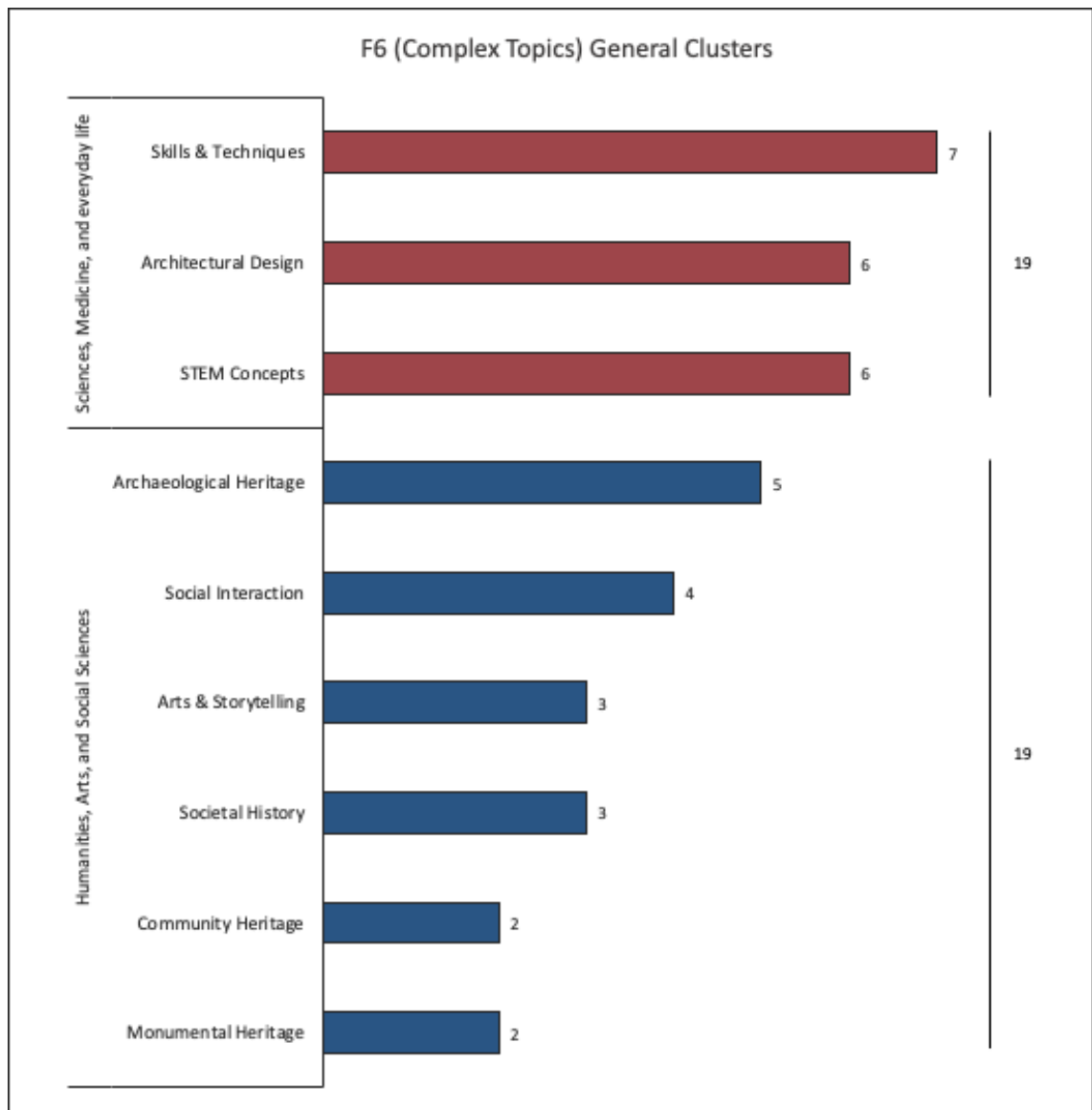


Figure 2.11 – F6 (Complex Topics) General Clusters' Totals.

2.6.2 F7 - What technologies are used to provide visual and spatial context?

With the distinctive theme totals of F7 (*q.v.* Section 2.5.2.2, specifically *q.v.* Figure 2.4), we have grouped them per purpose: for visual context, for spatial context, or for both, as clarified in their definitions (*q.v.* Section 2.4.2). All of them can, under the right circumstances, leverage both dimensions—after all, the boundary between context dimensions becomes fuzzy from a non-positivist perspective (*q.v.* Section 2.2.3 - *Learning in Context*). That is, we can envision circumstances where a technology deemed “spatial” in this classification might also be used for visual context, and vice-versa. For instance, the theme *Fixed cameras for markers* has been classified solely as spatial-context technology. In its definition, we clarify that this is due to these cameras being used by requiring that users move to a specific space (hence “spatial”). One can certainly imagine systems where fixed cameras are used for capturing visuals, such as someone’s face, and indeed that would be a case of capturing visual context, but no such instances were found in the survey papers. That being said, we tried to split the themes by their major or primary focus (*q.v.* Figure 2.12), while still acknowledging that they could perfectly be placed under both contextual dimensions (when under the right circumstances).

The first thing to point out is that the Spatial and Visual dimensions seem to be tackled more or less the same regarding the technologies being employed (totalling 44 and 41 appearances, respectively). An interesting point can be made, particularly regarding *Mobile Devices* and *Gestural/Somatic Tracking* themes (totalling 21 and 20 studies out of 38, respectively). Firstly, one can see the big incidence of mobile devices, including for the most part devices such as smartphones and tablets—which makes sense since the outburst of these consumer-wise started back in 2013 (one of the reasons that led us to set that as the initial date for the search of this SLR, *qq.v.* Sections 2.2.4 - *Immersive Environments*, and 2.3.2 - *Search Process*). Because of almost everyone owning a mobile device, it comes to no surprise that so many of these experiences and studies actually leverage them. By also taking into account the results from F9 and F10 (*qq.v.* Sections 2.5.2.4, and 2.5.2.5, respectively),

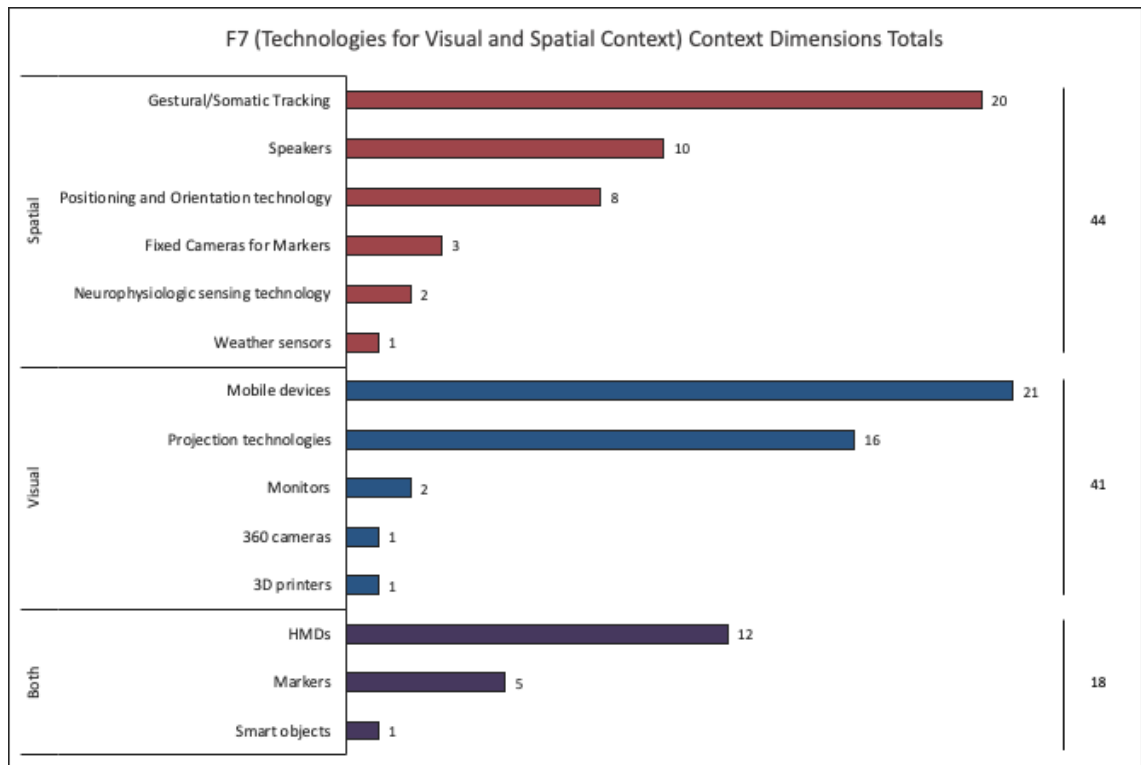


Figure 2.12 – F7 (Technologies for Visual and Spatial Context) Context Dimensions' Totals.

one can see that when it comes to Gestural and Somatic tracking, the bigger incidences are clearly **Somatic (body movement)**, and **Gestural (touch-based)** (totalling 32 and 20 studies out of 38, respectively). Though mostly implicitly, these experiences end up requiring the user to walk in museums or archaeological sites (somatic movement—sometimes resorting to **Positioning and Orientation technology**), and to use the touch-screen (*i.e.*, gestures) to interact with the device. Their ubiquity, potential low-cost, and pocket size could be part of the reason so many experiences are leveraging them rather than other, perhaps more expensive and cumbersome, alternatives (*e.g.*, **Projection technologies** and **HMDs**). This can also be part of the reason **Neurophysiologic sensing technology**, **Weather sensors**, **Smart objects**, **360 cameras**, and **3D printers** have such small incidence in these studies. We can thus conjecture that technology with more massification, cost-effectiveness, and a more user friendly experience are used the most to tackle both Spatial and Visual dimensions of context. A more in-depth look (*q.v.* figure 2.13), shows us that the

difference between the incidence of these dimensions is 2.91%.

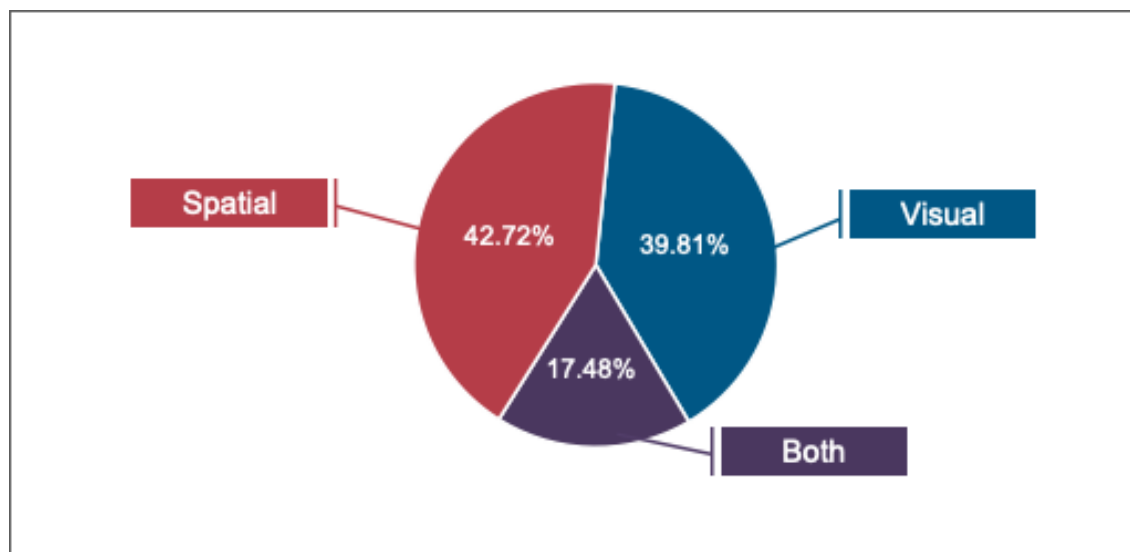


Figure 2.13 – F7's Context Dimensions Side by Side Percentages.

2.6.3 F8 - What techniques are used to provide more visual and spatial context?

With the distinctive theme totals of F8 (*q.v.* Section 2.5.2.3, specifically *q.v.* Figure 2.5), one can argue, similarly to what happens in F7, that the themes found can be split, though with some subjectivity as explained in the background section (*q.v.* Section 2.2.3 - *Learning in Context*), by the two (2) types of context (visual and spatial). For instance, the *Media's theme* is placed primarily under Visual Context, since it is more about visualising any form of media (*q.v.* Section 2.4.3). But if that media is used to pin-point the position of visitors within a museum or to describe the surrounding environment, then it could also be placed under Spatial Context. That being said, we tried to split them by their major or primary focus (*q.v.* Figure below), while still acknowledging that they could perfectly be placed under both contextual dimensions (when under the right circumstances).

The first thing to notice when analysing the chart above (*q.v.* Figure 2.14), is perhaps the division between the employment of either Visual or Spatial focused

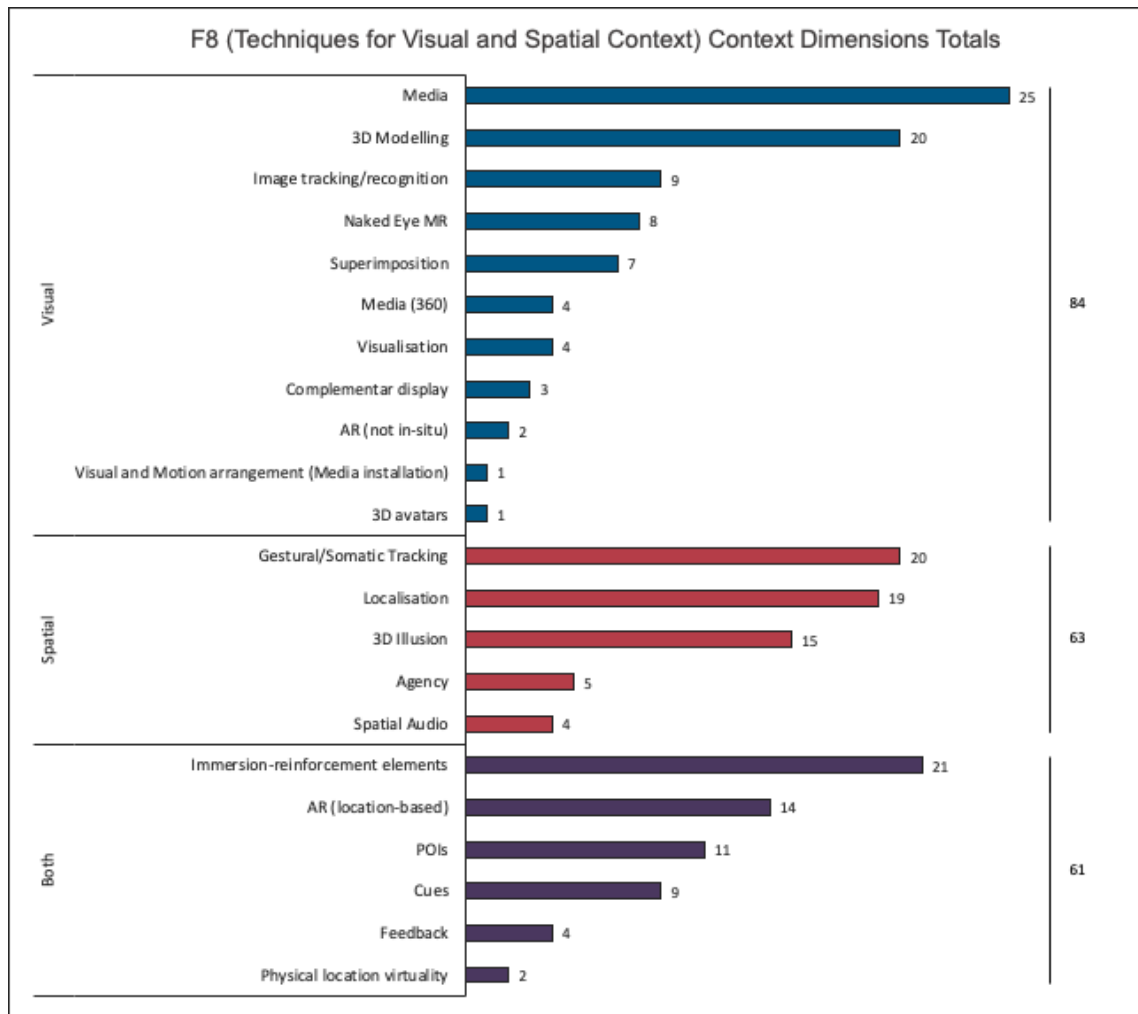


Figure 2.14 – F8 (Techniques for Visual and Spatial Context) Context Dimensions' Totals.

techniques. There is certainly a bigger gap than the one observed in F7 (2.91% vs. 10.09%, *q.v.* Figure 2.15). Of course, as mentioned earlier, this division is fuzzy and subjective. It is certainly not enough to infer anything, other than, perhaps, visual techniques being a bit easier to employ or more practical than spatial ones (*e.g.*, providing visual content through [Media](#), and [3D Modelling](#)). Before looking to investigate this further, however, let us first analyse the preponderance of the remaining techniques.

Within spatial context techniques, we found contrasting that a low-preponderance technology use (F7, [Position and Orientation technology](#)—with 8 studies out of 38)

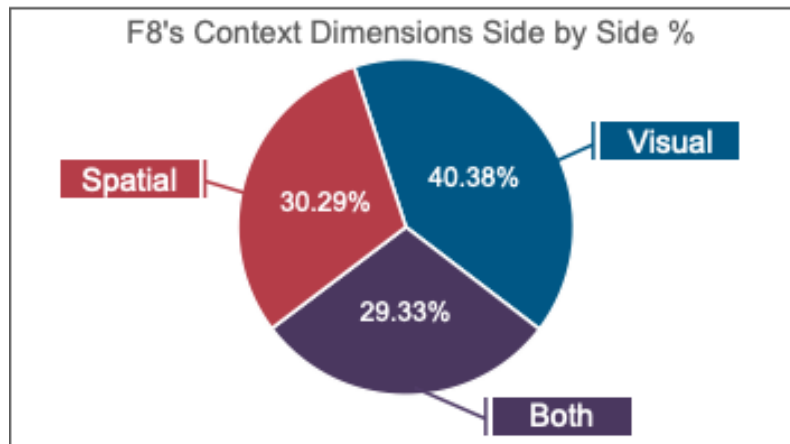


Figure 2.15 – F8's Context Dimensions Side by Side percentage.

is matched by a higher-preponderance technique here, Localisation (19 out of 38). This is because techniques employ both the positioning technology and the cognitive impact or leveraging, as mentioned in the background (*q.v.* Section 2.2.3 - *Learning in Context*). In this case the technique of Localisation is employed in more diverse manners than simply locating the user, through, for instance, providing cues to the user, as described further in its *TA* definition. One can conjecture that techniques are indeed complementing the technology, and in some cases, replacing it.

In regard to visual context techniques, an aspect that agrees with previous literature is the employment of techniques such as *image tracking/recognition*, *3D Modelling* (*e.g.*, through Photogrammetry), for things like *Superimposition*. As expected, however, modelling has more incidence (20 out of 38) than the other two themes combined (totalling 16 incidences), perhaps due to the challenges of employing such tracking/recognition and projection techniques.

The use of *AR* (both *location-based* and *not in situ*), and *3D illusion* (which includes *VR* employment) seems to be balanced (16 *vs.* 15 studies out of 38), even if there is a majority of mobile devices being used, as seen back in F7's discussion (*q.v.* Section 2.6.2). While *AR* and *VR* scenarios were clearly expected (or even *MR*), *Naked eye MR* (8 out of 38) emerged from the collected and analysed data. Unlike traditional *MR* (*e.g.*, wearing a headset), and per its definition, it

brings virtual components onto the physical world without the direct need of a display or camera (thus Naked eye). The opposite, bringing the real onto the virtual, also a bit unexpected, was named as **Physical location virtuality**, though it didn't show as much incidence (2 out of 38). It might come to a bit of a surprise, however, that these immersive environments aren't employing, at least for visual and spatial context reasons, **POIs**, **Cues**, **Agency**, and **Feedback** much (11, 9, 5, and 4 out of 38). **Spatial Audio**, as already seen in F7 through **Speakers**, isn't employed much either (4 out of 38). Despite their clear importance (*q.v.* Section 2.2.5 - **Multimodal Human-Computer Interaction**), it can be argued that they might be considered tacit, and thus not approached/reasoned explicitly when designing and describing these experiences. Though also not explicitly, immersion reinforcement elements are used 21 out of 38 possible times, which, despite being almost 50% of the studies, might also indicate a lack of background effort. This would agree with the fact that studies tackling a more holistic view on the subject are pretty recent. Up until now, there was really no agreement on the literature regarding the dimensions of immersion and their importance (*q.v.* Section 2.2.4 - **Immersive Environments**). It would then come to no surprise that, perhaps, even if these experiences tackle the narrative or challenge in some way, their authors might not realise that and thus not describe those aspects.

To further investigate the earlier impression of visual techniques being a bit easier to employ or more practical than spatial ones, and to get a more general overview of what is being employed, one can divide the themes into three (3) other clusters (*q.v.* Figure 2.16)—the types of visual and spatial techniques being employed:

- Content and content-provision techniques;
- Cueing prompting (feedback, cues, and similar low-interaction);
- User-led techniques (*e.g.*, Agency, **AR**).

Similarly to the clusters from before, it is important to note that some of these themes might make sense in more than one cluster (*e.g.*, **3D illusion**, as a content-provision technique and also cueing the sense of depth, and Immersive-reinforcement

elements, with the different dimensions tackling different immersion aspects).

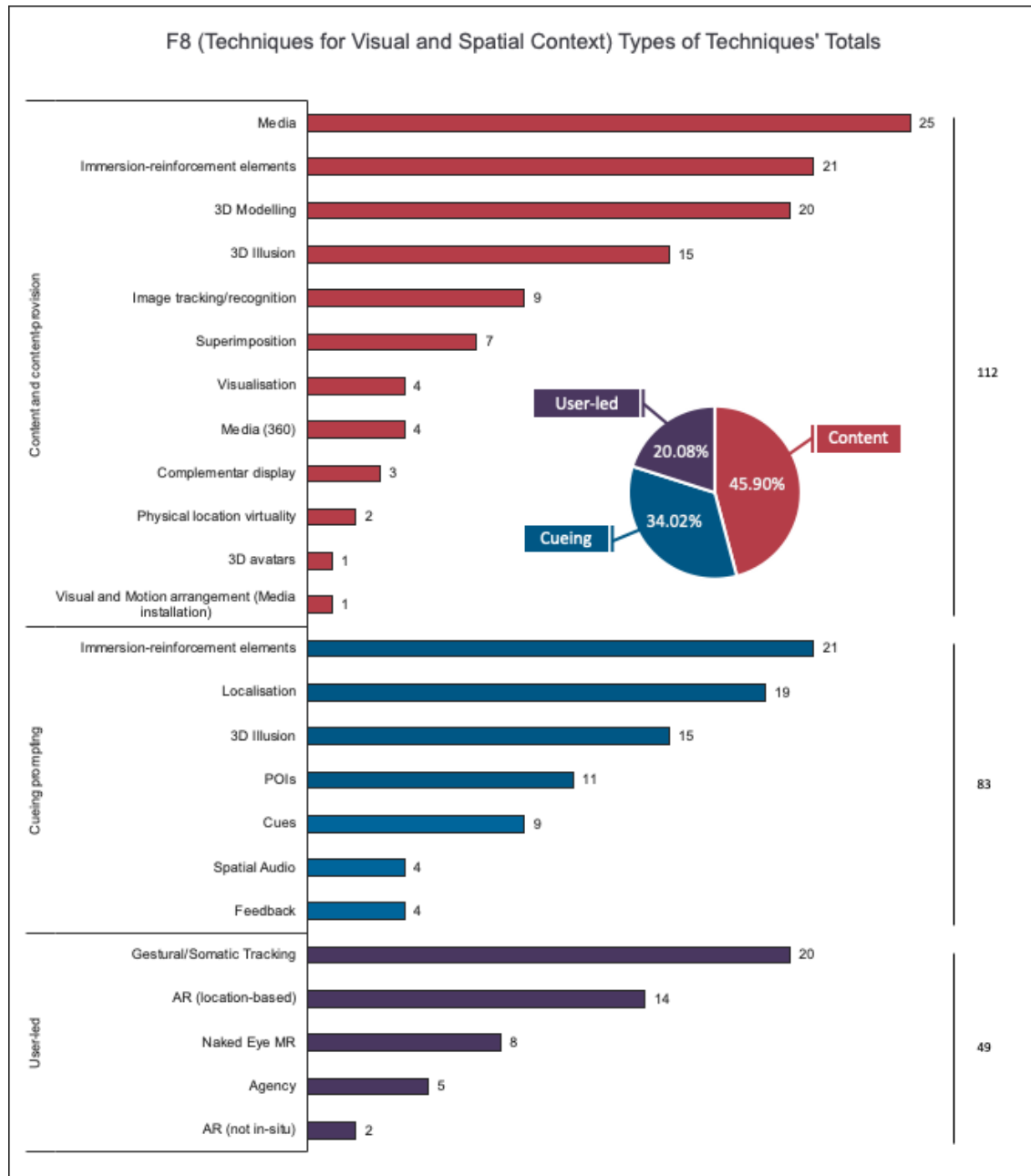


Figure 2.16 – F8 (Techniques for Visual and Spatial Context) Type of Techniques' Totals.

And as argued before, Content or content-provision techniques are indeed employed more (45.90%, *q.v.* Figure 2.16, above) than Cueing (34.02%) and User-led techniques (20.08%). This might have to do with what is easier to employ. Content

(and content-provision techniques) seems to be more straightforward to include in these experiences. Then again, the lack of incidence of the two other clusters might have other reasons behind (other than simply not being employed). They might be seen as tacit and therefore having no description/reasoning in the consequent studies; and/or these experiences are not being grounded as much in background as they could and probably should. One cannot forget, however, that this might also have to do with the fact that some of these (important) concepts are not really set into stone in the literature, which can obviously lead to less consistency, or, indeed, the not tackling of these important aspects.

2.6.4 F9 & F10 - What technologies are used to provide culture-aware interaction? And what types of interaction?

With the distinctive theme totals of F9 (*q.v.* Section 2.5.2.4, specifically *q.v.* Figure 2.6), the first thing one might notice is that similarly to F7 (*q.v.* Section 2.5.2.2), F9 has a theme named **Mobile Devices**. It is important to understand that despite the name, they are categorised according to their specific use-cases and overall field's goals. To be more clear, F7 was focused on visual and spatial context, whereas F9 focuses on the cultural dimension. Similarly to what happens in F7, **Mobile Devices** have the preponderance of incidence, followed by **Naked eye MR technologies**, and **Body Motion Capture** (14, 12, and 11 studies respectively). These three themes alone have almost the same incidence as all the other themes combined (37 *vs.* 40). It does seem again that this is related not only to the easiness of use, but also with the freedom (and agency) that is given to the user—even if implicitly. The common thing in these three themes is that the user is allowed to walk more or less freely in the physical world (*e.g.*, museums, archaeological sites)—which is someone everyone is used to, and thus being considered culture-aware interaction. A consequence of this freedom is that the user is then given the ability to interact with physical (even non-tracked) objects that are available in the surroundings. One could argue again, however, that this culture-aware interaction is implicit—perhaps not even thought of when designing the experience. The remaining themes end up splitting between

hand, and head related movements, tracking, and gestures, which usually require more hardware (*e.g.*, cameras, headsets, recognition devices). This can pose more cons, when compared to the top three themes, such as being more cumbersome and not as easy to set up. If we investigate further the type of interaction being employed we can see that the Somatic and Gestural types have the clear majority of use (55.43%, and 43.38%), mainly because of **Somatic (body movement)**, and **Gestural (touch-based)**, respectively (32, and 20 studies out of 38)–*qq.v.* Section 2.5.2.5, and Figure 2.17, below.

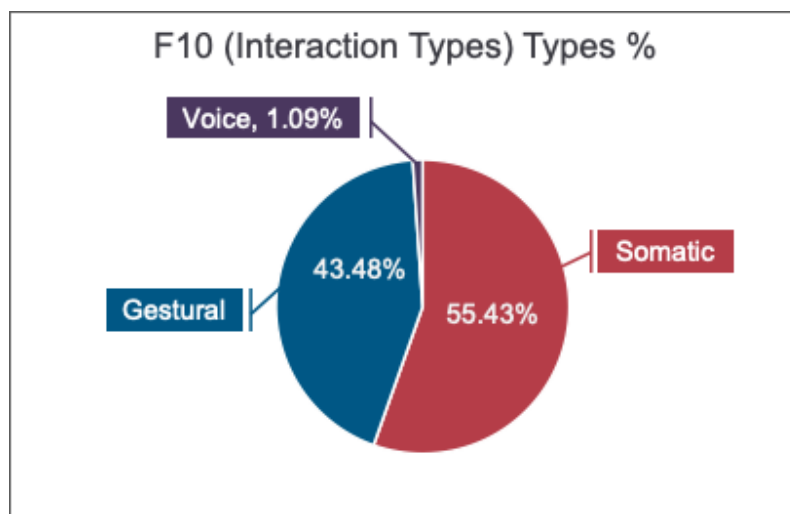


Figure 2.17 – F10 (Interaction Types) Side by Side Percentages.

We can now clearly see that the freedom and agency given to the user (*e.g.*, walking) reflects here as Somatic (body movement). Gestural (touch-based) has a higher incidence than the other subtypes in the gestural cluster because of the use of mobile devices and things like tabletops (*e.g.*, in Naked eye MR experiences). One can then say that, purposefully or not, the type of interaction used the most is arguably the easiest to employ (walking), followed by the one that everyone uses daily through their pocket-sized smartphones (touch). Perhaps surprisingly, voice control is mostly not used (only 1 out of 38), but one can argue that might have to do not only with the type of experiences (mostly outdoors, with noise, and walking), but also with the fact that voice recognition is still not mature or accessible enough.

From this discussion subsection alone, one might start wondering if the technology

used most to provide culture-aware interaction for complex learning immersive scenarios is simply the one that leverages the general spread of mobile devices (and consequently physical world experience, where one implicitly has more agency and interaction options out of the box). Not quite. Looking back to F8 (techniques for visual and spatial context, *q.v.* Section 2.6.3), we see that AR and VR are quite balanced (16 *vs.* 15 out of 38 studies), and Naked eye MR has roughly half of their incidence (8 out of 38). That leads us to argue that the ubiquity and pocket-sized devices are indeed the most employed. But that's for the Visual and Spatial dimensions of context. Looking now at the types of technology being employed in this field (*i.e.*, F9), or in other words, how do they affect the end-user in practice, one can further divide them into two other categories (*q.v.* Figure 2.18): those who detach the user from the physical world (*i.e.*, technology that helps the user immerse into a virtual environment), and those that expand it (*i.e.*, empowering the user in the physical world with more abilities and choices). There are a few cases, however, that under the right circumstances, might leverage either scenario (*e.g.*, Voice control, and Haptic actuators).

Detaching from the physical world has a slighter preponderance (50.65%) when compared to Extending the physical world (44.16%). And, contrary to one might have wonder thus far, we see that for the culture dimension of context, is actually leveraged more through the use of (typical or more known) VR-related technology (body, hand, and head tracking)—more than the ubiquitous and pocket-sized, physical world, experiences (*i.e.*, Extend the physical world).

2.6.5 F11 - What methods/techniques are used to provide cultural context (and meaning)?

With the distinctive theme totals of F11 (*q.v.* Section 2.5.2.6, specifically *q.v.* Figure 2.8), one can see the preponderance of interaction methods (35 out of 38), followed by topic (25 out of 38). Given the particular focus of this SLR on interaction and culture, one can argue that the used search keywords (*q.v.* Section 2.3.2) may have contributed to such prevalence. But one cannot forget the main research

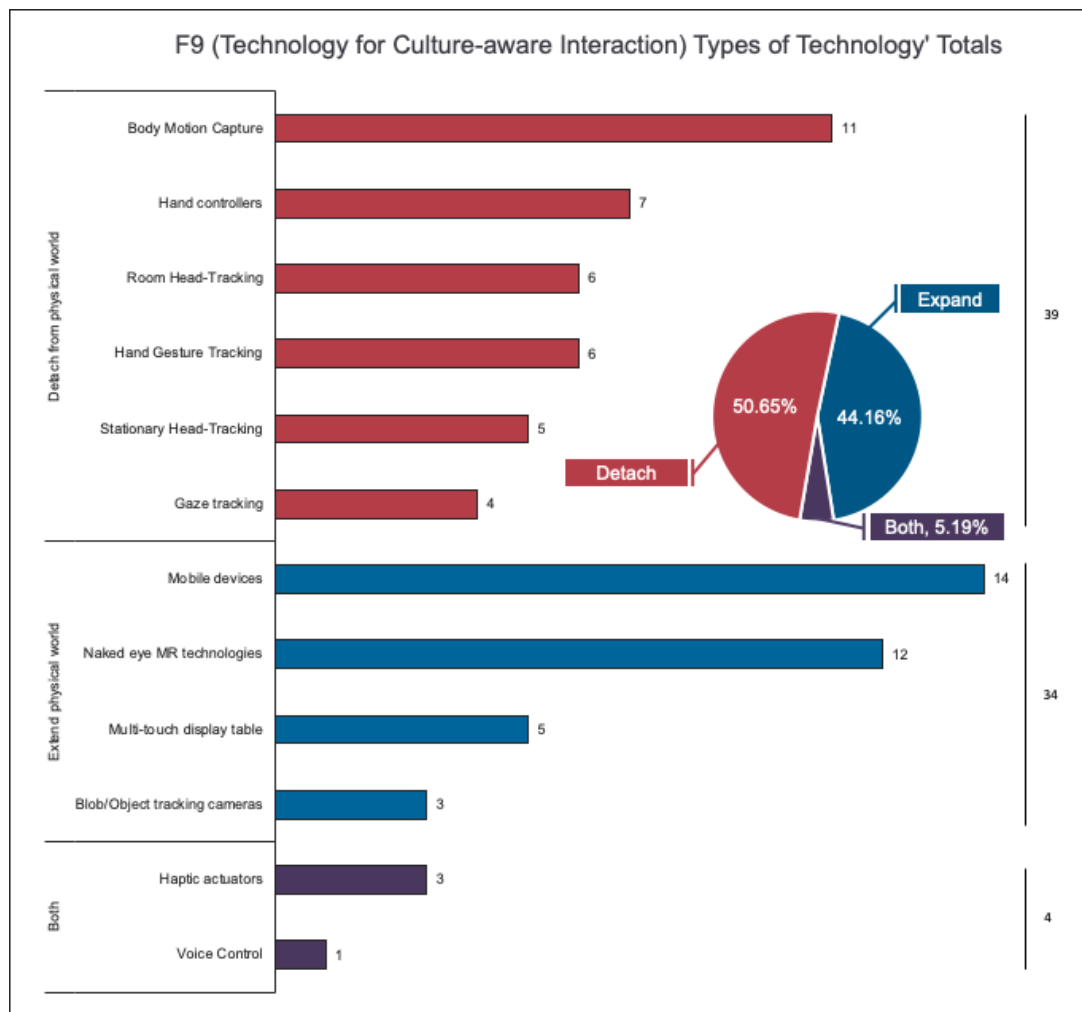


Figure 2.18 – F9 (Technology for Cultural-aware Interaction) Types of Technologies' Totals.

question of this study (*q.v.* Section 2.3.1), which requires us to do so if we expect to obtain relevant enough results. This can be an endless loop, however, as it is certainly easier to find something when actively looking for it.

Given the vast number of themes in this field, and their particular balanced incidence (other than the top two themes), it might be useful to get a general overview by applying the same clusters (*q.v.* Figure 2.19, below) we used back in the discussion of employed Spatial and Visual context techniques (*q.v.* Section 2.6.3):

- Content and content-provision techniques;

- Cueing prompting (feedback, cues, and similar low-interaction);
- User-led techniques (*e.g.*, Agency, AR).

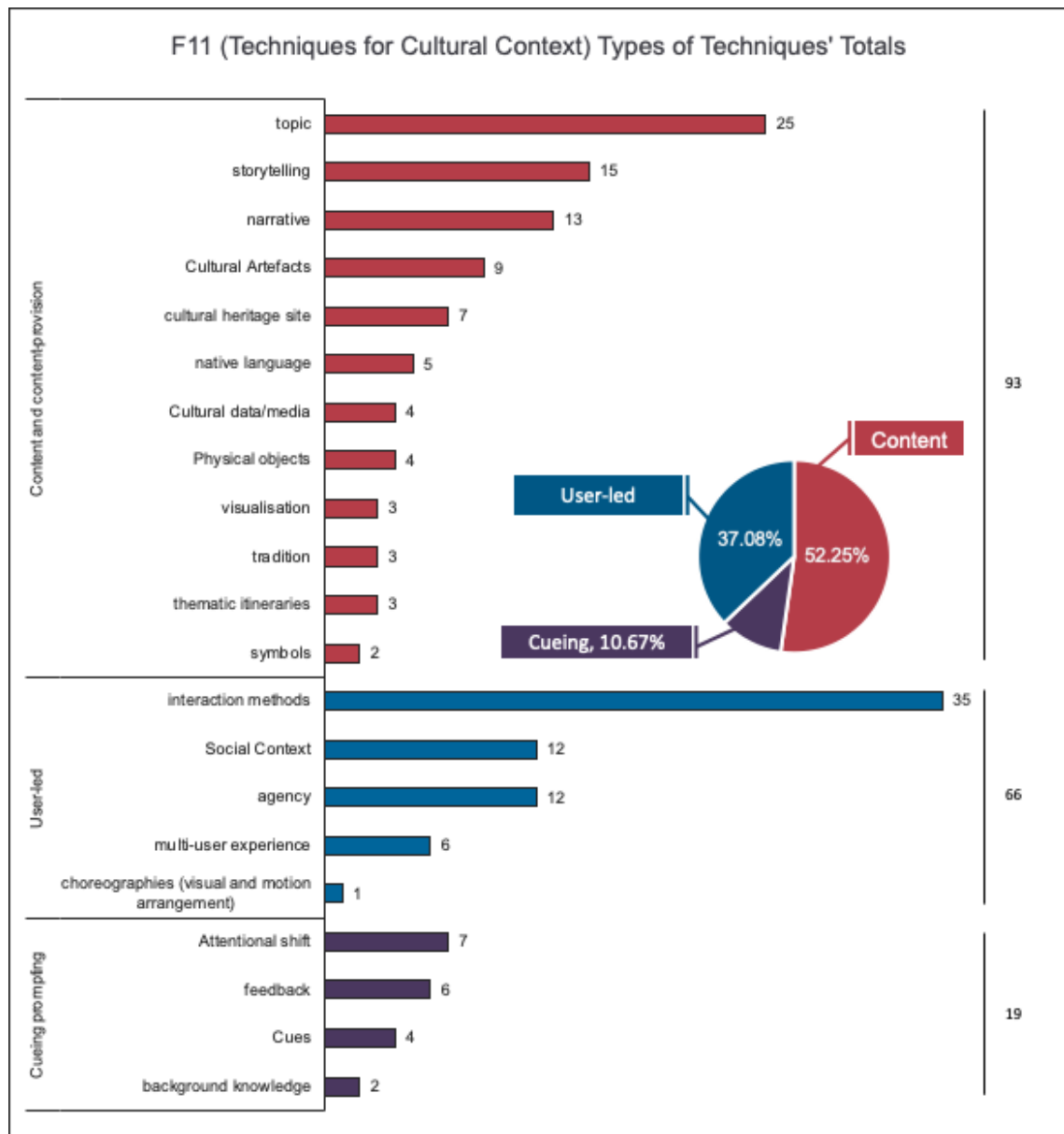


Figure 2.19 – F11 (Techniques for Cultural Context) Types of Techniques' Totals.

The first thing to notice is the increase in user-lead techniques (37.08%), and decrease of Cueing prompting (10.67%), when compared to the types of techniques

leveraged for the visual and spatial dimensions of context (17% and 23.35% difference, respectively, *q.v.* Section 2.6.3). Content and content-provision techniques keep the majority of incidence (52.25%), giving again the impression that there are more ways of showing and providing content to tackle the cultural dimension (along with visual and spatial, *q.v.* Section 2.6.3) of context.

The high incidence of **interaction methods** in user-led techniques is somewhat to be expected. The previous discussions (*qq.v.* Section 2.6.4 - F9 & F10 - **What technologies are used to provide culture-aware interaction? And what types of interaction?**) show that **body movement** (*e.g.*, walking) and **touch-based gestural** interfaces are used more, particularly when compared to **hand controllers** (7 out of 38) and **hand gesture tracking** (6 out of 38), perhaps because they are easier to employ. But if these modes of interaction (be with somatic or gestural) were to be further leveraged with **background knowledge** and meaningfully employed, then one could avoid learning a different set of inputs for every different experience (*q.v.* Section 2.2.5)—not to talk about maintaining them. In other words, an experience could be using interaction modalities that the users are already accustomed to from their day-to-day lives. Despite, many times, interfaces being referred to as “natural”, “fluent”, or “intuitive” by big technological companies, they still lack the one big aspect that would actually make them what they claim to be: leveraging the cultural dimension of context. That is, for instance, using gestures/-movements that a certain region is used to, rather than coming up with new ones that require effort to learn and are easily forgotten. Of course, this doesn’t come without challenges (detection and recognition, to name a couple). Employing **interaction methods** (implicit or not), **agency**, and designing an experience for multiple people (*i.e.*, **social context**, and **multi-user experience**) certainly poses more (technical) challenges than simply including some more or leveraging the content already present in the surrounding environment (*e.g.*, **cultural artefacts**, **data/media**, **physical objects**), though this doesn’t talk about the end-user experience (UX). Still, one cannot just ignore the challenges **storytelling** (15 out of 38) and meaningful narratives (13 out of 38) face. Though both of these are long-established traditions, used 100.000 years ago, nowadays mostly replaced by PowerPoint-like applications (*q.v.* Section 2.2.3), they still got a good incidence (when compared to other themes).

Maybe a bit surprising, is the lack of employed **tradition** (*e.g.*, dances, music, rituals), and **symbols** (*i.e.*, a semantic meaning attached to something). These are two clear and ideal themes that can and should be leveraged in these experiences, and, yet, are being forgotten or not given enough importance when designing or writing about the study—**topic** seems to be the only theme being exploited a bit more, though more can be achieved. Another theme that could use a bit more attention is the **native language**, either through vocals or subtitles, though it could be a case of simply not being reported in the studies, as we suspect happens with the use of sound and **speakers** (*q.v.* Section 2.6.2).

There is a low incidence of cueing prompting techniques (19) when compared to the other two types (10.67%), despite the known importance and incidence of **feedback** (*q.v.* Section 2.2.5 - **Multimodal Human-Computer Interaction**), **cues** (and **attentional shift**), and how they affect and influence learning (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**) in the physical world. One can again argue that it is indeed being used, but is mostly implicit, or, in other words, not purposefully employed or given enough importance. One of the reasons that this might be happening is due to a lack of grounding on previous literature. And perhaps this is also the root for the lack of purposefulness (and even low incidence) seen in other themes, such as **interaction methods**, **symbols**, and **tradition**.

2.6.6 F12 - How is the understanding/expression of a complex topic approached versus traditional methods?

With the distinctive theme totals of F12 (*q.v.* Section 2.5.2.7, specifically *q.v.* Figure 2.9), we can see that it is the field with the most themes (23). One might consider this particular field to be among the most crucial ones for this review (along with F11) as it ends up combining important aspects already seen throughout the discussion of the other fields. We cannot forget, however, how relevant the others are, and how they all ultimately influence complex learning. While we tried to separate the fields, it is important to keep in mind that the border between them can often be fuzzy—and repetitions will certainly occur. Gathering this item was an attempt

to capture the key aspects of each study that, in some way, influence the process of complex learning. To help understand what is effectively being employed to tackle this, one can group the themes further into a total of two (2) more general clusters (*q.v.* Figure 2.20, below): **Practices** (grand total of 99 incidences), and **Strategies** (grand total of 60 incidences).

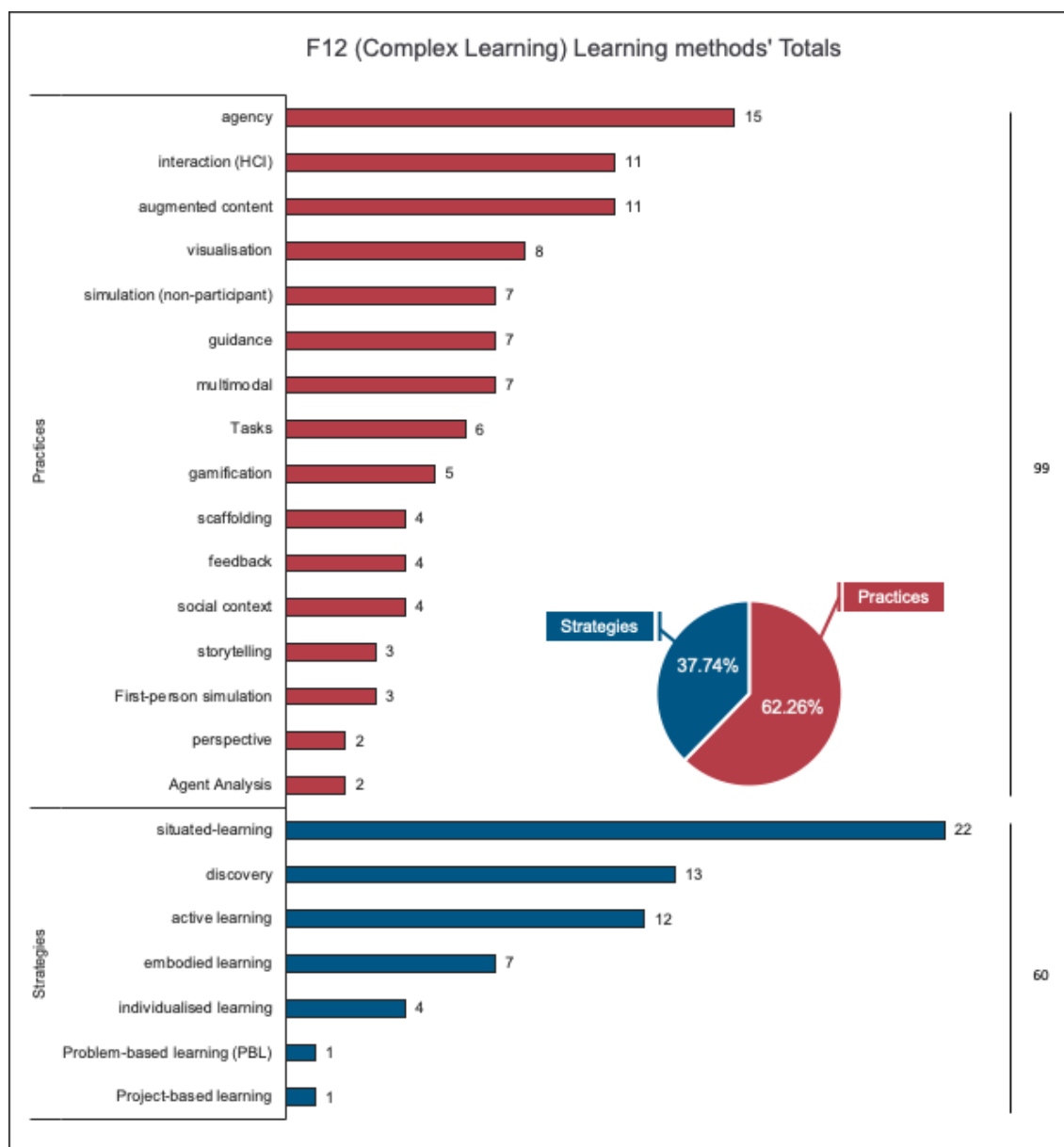


Figure 2.20 – F12 (Complex Learning) Learning Methods' Totals.

The preponderance of practices in relation to strategies (62.26% *vs.* 37.74%) might seem like a lot, but it only shows that there are more informal ways of approaching (complex) learning than formal educational strategies. If we take the same data, and retrieve the distinct studies within each cluster (*i.e.*, how many studies out of 38 do apply practices and strategies), we can see that they are more balanced (*q.v.* Figure 2.21, below). This tells us that out of 38 possible studies, practices are used in 36 of them, whereas strategies are used in 31. In other words, only 2 studies are not using any practices, and 7 are not using any formal education strategies. Given the sample size, this seems to indicate that a considerable part of the studies (~18.42%) do not define and leverage a teaching/learning approach (if any), which might indicate lack of educational focus.

Now that we have a general overview of how these studies are approaching complex learning, all that is left is to understand how these are effectively used to tackle the important aspects of this process (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**). That is, one can try to group them using the three complex learning attributes (*i.e.*, **understanding complex systems**, **developing adaptive expertise**, and **acquiring soft skills**, *q.v.* Section 2.2.2 - **Learning and Expressing Complexity**, as well as particular aspects that influence how we see the world and effectively learn (*i.e.*, through feedback, and repetition)—which are essentially part of the challenge of situativity or **simulating real life contexts**—, and finally draw a comparison with more **traditional** approaches (*e.g.*, **visualisation**, **gamification**). One last group can also include

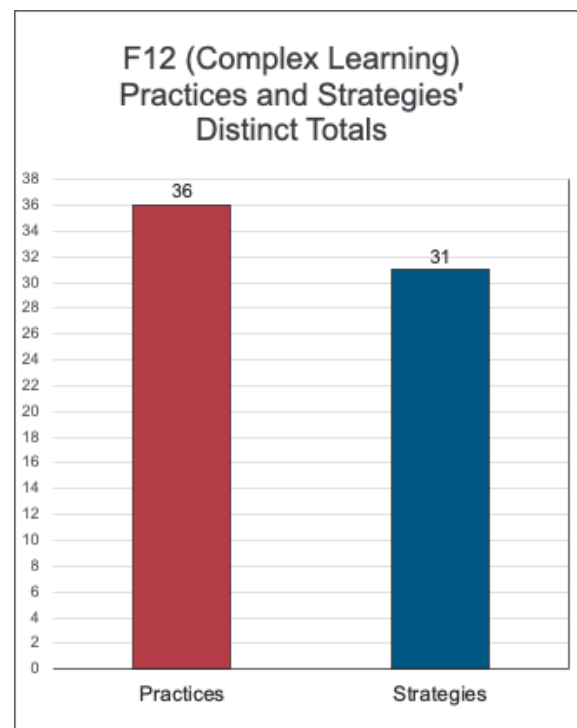


Figure 2.21 – F12 (Complex Learning) Practices and Strategies' Totals.

themes that are more of **supporting** the whole complex learning process. From this grouping attempt (*q.v.* Figure 2.22, below), which is obviously subjective, the first thing to notice is that the cluster **understanding complex systems** has the lowest incidence (18 total appearances). This seems to indicate a lack of this aspect in general. The studies are indeed approaching complex learning in some way, but they are mostly skipping/ignoring the phase of explaining how complexity and complex systems work. The main research question of this SLR is about understanding what indeed eases the process of complex learning, thus being somewhat unable to specifically assess whether this particular attribute (*q.v.* Section 2.2.2) does indeed ease that process. There is a bunch of data on aspects that more or less indirectly influence it, but there is no direct, explicit approach overall, and thus showing low incidence in the chart below. The two other complex learning attributes, **developing adaptive expertise**, and **acquiring soft-skills**, have a higher incidence, both with 36 total appearances. Both of them are the indirect result of applying other recurrent and non-recurrent skills, and actively communicating and cooperating with other individuals on tackling bigger tasks (6 out of 38)—that is the reason some of these active learning methods are common to both attributes and other clusters. **Guidance** (7 out of 38), scaffolding (4 out of 38), and individualised learning (4 out of 38), despite being more of (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**) does indeed ease that process-based, implicitly contribute to goal elaboration and balancing (*i.e.*, adaptive learning tasks), in turn contributing to (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**) does indeed ease that process (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**). Given the importance of these attributes, the extremely low incidence of **PBL** and **Project-based Learning** (both with 1 study out of 38) may be a bit surprising—but then again, we have already seen the indication that a good part of these studies (~18.42%) do not even define a formal educational approach.

When it comes to situativity, or **simulating real life contexts**, we certainly see a trend of higher incidence (101 total appearances), particularly with the themes **situated-learning** (22 out of 38), and **agency** (15 out of 38). Relatively speaking, simulating a real context implies more agency, which in turn implies more **interaction** (11 out of 38) and empowering of the user. The outgoing result is essentially (and ideally)

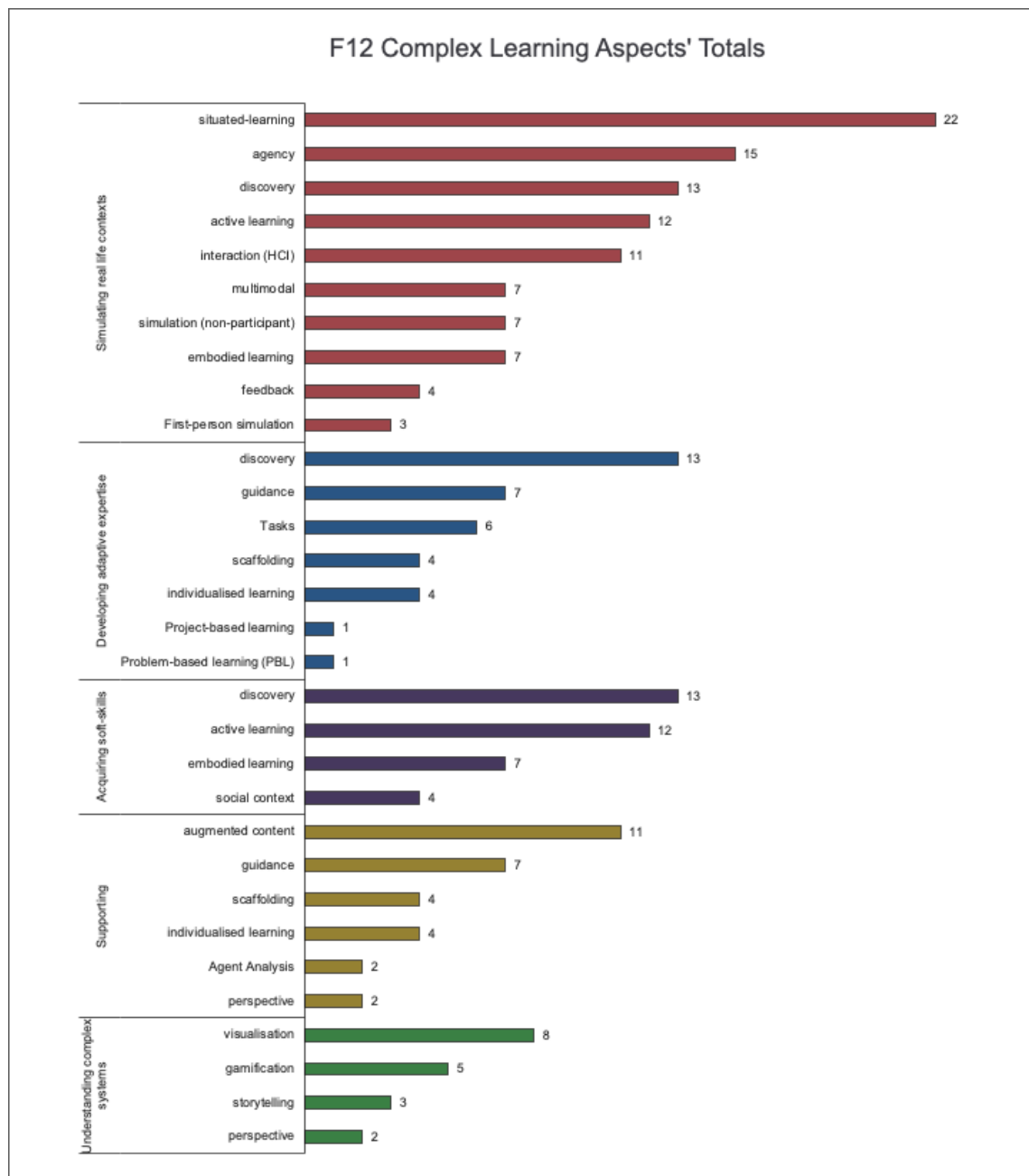


Figure 2.22 – F12's Complex Learning Aspects' Totals.

learning by doing, much like it happened before the industrial revolution (*q.v.* Section 2.2.3 - [Learning in Context](#)). The high incidence of this particular cluster can be explained if we recall the strategies and practices (*q.v.* Figure 2.22, above). By

doing that, we see that a bit over half of this cluster's incidence, 54 out of 101, are educational **strategies** (*q.v.* Figure 2.22, above), with the remaining 47 appearances being **practices**, so there is a definitive balance between both—which can be seen as a positive approach.

Supporting or leveraging the complex learning process might not be considered a formal complex learning aspect, but one can argue it is always there and can take many forms (30 total appearances). Not surprisingly, themes like **guidance**, **scaffolding**, **individualised learning**, and **agent analysis** are part of it, as they aid the users in different ways: the first three with **developing adaptive expertise** as mentioned before, whereas agent analysis is more of self management and reflective expertise focused (*i.e.*, feedback) (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**). Changing angle, or **perspective**, is one of those things that everyone knows is important, but only a few take the time to leverage it, at least explicitly—particularly in immersive environments (2 out of 38). The preponderance within this cluster belongs to augmented content, which, despite being a bit more reassuring given that, as argued by some authors, **AR** and **MR** in particular, open up new opportunities to both support and facilitate complex learning (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**), still show a lower incidence than expected (11 out of 38). If we look back at the numbers of **mobile devices** being employed (21 out of 38) in these experiences (*q.v.* Section 2.6.2), we can see that roughly half of them do not employ augmentation techniques.

The last part of this sub research question (F12) has to do with some comparisons with more **traditional** approaches of complex learning. Traditionally (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**), **visualisation** would be the method of choice when it comes to deal with complexity—through the interaction with **3D** models (*e.g.*, changing the **perspective**—only 2 out 38) in a standard **2D** screen. Although those specific cases are not included here (since the use of a **2D** screen as the main form of visualisation is an exclusion criteria, *q.v.* Section 2.3.3), they still show up in a more up-to-date manner, and employ newer technologies and techniques (8 out of 38). Positively, these immersive environments try to reach farther by making these visualisations come alive and further immerse the user in the

data, offering new perspectives, and modes of interaction, which should result in positive learning outcomes. Similarly, **gamification**, including what is known as serious games, is on what seems to be a decreasing trend (5 out of 38) when it comes to complex learning. One can argue it was more popular before the outburst of **VR** and **AR**, through the use of standard **2D** screens, or that the term is being less used given its connotation: either being too playful to learn, or being too serious. **Story-telling** has even a lower incidence (3 out of 38). If we look back at the data we got from analysing F11 (*q.v.* Section 2.6.5), we see that it is a technique used mostly to bring cultural context (15 out of 38), rather than a way to tackle complex learning explicitly. This, again, indicates the lack of awareness regarding this powerful, non-linear, teaching method that was used by our ancestors 100.000 years ago to transfer knowledge from generation to generation. It is still used today, mostly by native tribes, but this indicates, worryingly, that is not something these studies are leveraging much lately.

We can now argue that more situated and active learning approaches, employing more agency and interaction methods, have the preponderance when dealing with complex topics. This is, however, done mostly implicitly. To be more clear, most of these studies are not upfront with the educational approach used within their experiences. Their categorisation, though subjective, was done deductively, according to the available descriptions of said experiences. This indicates a lack of educational focus, which agrees with the fact that context has not been explicitly orienting pedagogical practice either (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**). We can argue, however, that this problem might be on these studies not being explicitly tackling complexity. That is, looking back at the topic being learnt and/or taught (*q.v.* Section 2.6.1), we infer that the studies (particularly in humanities, arts, and social studies cluster) are working with complexity, as the authors do not make that clear. That being said, one cannot then expect them to tackle these complex problems with formal complex learning approaches, such as Complex Constructivism, **CPS**, Connectivism, model **4C/ID** (*q.v.* Section 2.2.2 - **Learning and Expressing Complexity**). Humanities, Arts, and Social Sciences, might need more sustained work on complexity (more literature grounding) so that they can exploit

it as much as possible—otherwise it is a missed opportunity. But then again, we cannot forget this **SLR** combines many areas of research and thus being harder to find studies that tackle everything we expect them to.

2.6.7 F13 - What are the main targets of these studies?

With the distinctive theme totals of F13 (*q.v.* Section 2.5.2.8, specifically *q.v.* Figure 2.10), one can clearly see the outstanding majority of the **general public** theme (*e.g.*, visitor, tourists, citizens, *etc.*), with 27 out of 38 appearances (followed by **Topic Novices** with 12 out of 38). As mentioned in its definition (*q.v.* subsection 2.4.8), the studies included have “**no restriction in age** (*i.e.*, can go from children to adults). (...) the experience **doesn’t really have a specific target group**, so the experience within *can vary*, but is not accounted for”.

To investigate this idea further, one can take the aforementioned results and group them by **Specific** or **Non-specific**, while making sure the included studies within each cluster are distinct. In other words, to effectively understand how many studies are being non-specific (*i.e.*, targeting the **general public**), and specific (*i.e.*, anything other than the general public), duplicates should not be taken into account, as not to give a false impression. That is, if a study targets two different groups under one of the clusters, it only counts as one. Doing that (*q.v.* Figure 2.23) clearly shows that the majority (58.70%) of studies are not targeting a specific group (*i.e.*, a group other than

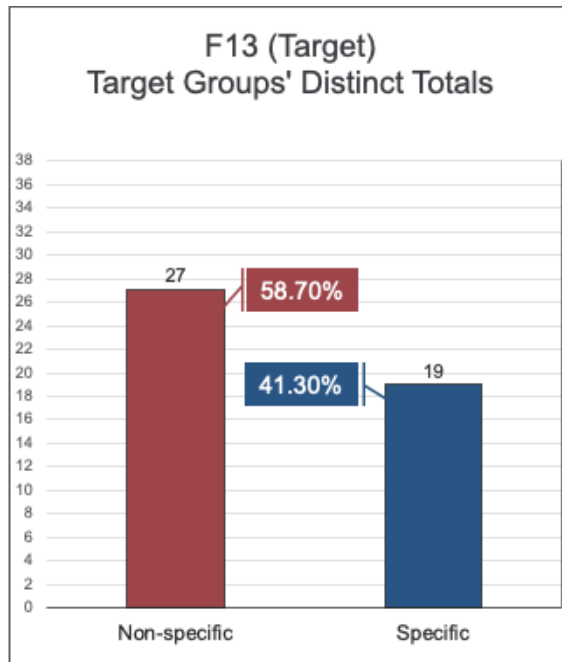


Figure 2.23 – F13 (Target) Target Groups’ Totals.

the general public) and therefore do not, and cannot, leverage the participants prior knowledge nearly to its potential (both intellectual and interaction wise, *q.v.* Section 2.2.5 - [Multimodal Human-Computer Interaction](#)). Of course, even the studies grouped as Specific (41.30%) are not necessarily leveraging the respective groups prior knowledge. On the one hand, most studies, by targeting the general public, are not targeting anyone specifically. On the other hand, even if the studies are targeting a specific age group (*e.g.*, of students), that doesn't necessarily mean they are indeed doing it when it comes to leveraging their prior knowledge and interaction methods. One can argue that although interaction methods are vastly used throughout these studies (*q.v.* Section 2.6.5), it ends up being brought indirectly and is not being accounted for for the most part (*q.v.* Section 2.2.5 - [Multimodal Human-Computer Interaction](#))—even if it indirectly affects complex learning. In other words, we argue there is still a clear lack of planning when it comes to designing more contextual-aware experiences.

2.7 Conclusions

We performed a systematic literature review to study how immersive environments leverage context, including cultural-interaction, to increase the understanding and communication of complexity and complex concepts by non-technology experienced users. We performed multiple automatic searches across several engines and databases to collect our primary studies. Results of our study suggest relevant gaps in the empirical grounding of current knowledge, thus contributing to the establishment of future goals for the research community. Humanities related fields ([HASS](#)), in particular, show the need for more grounding on complexity, so that they can exploit its traits with the already employed techniques and technology, and consequently use an adequate and explicit educational approach. Conversely, [STEM](#) related studies, particularly those of physics that are known to indeed tackle complexity using visualisation techniques, seem to be lacking the more technological and cultural aspects. Both of these cases present a missed opportunity, but show a clear path for

future research. Interdisciplinarity is not only real, but also needed to better understand our surroundings and everyday life. This should make it compelling enough to join efforts and close the rift that our results suggest is still observed between the sciences and what are often referred to as second-class citizens such as humanities and arts. Only then it will be possible to find universal or more fundamental laws that govern these complex systems.

Our results also suggest that, despite situated learning environments being used often, context continues to not explicitly orient pedagogical practice, particularly that of complexity. As a consequence, the user interaction is mostly implicit, and not reasoned. Most of these studies use somatic (body movement) and gestural (touch-based) modes of interaction. But that happens indirectly due to the use of pocket-sized and ubiquitous devices such as smartphones. Experiences are being planned and designed with some specific technology in mind, and not necessarily with the interaction or environment they imply—they are ignoring the multi-dimensionality of context, in particular the cultural diversity. Not targeting a more specific group of people, other than the general public, also suggests that the prior knowledge is not being leveraged as much as it could—if anything at all. Our main findings (*q.v.* Section 2.6 - Discussion) point towards a more clear path of research of bringing more holism and reasoning to these contextual aspects. Studies need more grounding on the literature from other fields such as context (including culture), interaction, and complexity—only then can we tackle complex learning effectively.



Software Architectures

3.1 Introduction

What began as a basic survival need and rudimentary (*e.g.*, communication, guidance, navigation) has evolved into something that no one can alienate these days, integrating naturally into society's day-to-day. As is to be expected, complexity emerges from any evolution, requiring the reflection and testing of new concepts by researchers and entrepreneurs. It is possible to draw parallels with the idealism advocated by philosophers such as Hegel and Ficht, while acknowledging some exaggeration, which at the level of science remains relevant in some ways, because “dreams should always lead the life”, always considering its efficacy and, preferably, its efficiency. Certain systems, such as those relating to evolution and culture, have always posed a challenge to science, particularly in the technological field. The use of architectures, which have proven to be essential in the analysis and description of these systems, can assist in understanding complexity (Garlan and Shaw, 1993, pp. 1–2). The concept of architectures (software-wise) was first proposed in the late 1960s, beginning with Naur's 1968 comparison of software design and (civil) architecture “software designers are in a similar position to architects and civil engineers” (Naur and Randell, 1969), with the true embryo being Dijkstra's study of

system structures (Dijkstra, 1968), followed by David Parnas (Parnas, 1978; Parnas et al., 1984). Just after that, Brooks and Iverson defined architecture as the “conceptual structure of a computer (...) as seen by the programmer” at the same time. Blaauw later expands on this concept, defining system architecture as “the functional appearance of the system to the user, its phenomenology” (Blaauw, 1972). He draws an analogy between wristwatches and clock towers, whose architecture (*i.e.*, the operation of the hours and minutes hands) is the same despite differences in implementation and realisation: “where the architecture tells what happens, the implementation describes how it is made to happen” (Blaauw, 1972). Brooks contributes to the distinction by emphasising the importance of keeping the architecture and its implementation separate. As a result, future system additions will be aligned with the defined architecture, preserving conceptual integrity (Brooks, 1975, pp. 44–45).

Only in the early 1990s was the concept of software architecture considered (Shaw, 1990; Reichtin, 1991; Perry and Wolf, 1992), presenting itself as a way to counteract the new complexity of systems, which was thought to be impossible with the techniques (*i.e.*, data structures) available at the time. Shaw argues persuasively for the need for new abstraction techniques to solve these problems, describing some of the most popular organisational systems (*e.g.*, pipe and filter, layer, blackboard), and Perry and Wolf present a model of software architecture comprised of three components, namely **elements**, **form**, and **rationale**. As a result, the following definition has been developed: “set of architectural (or, if you will, design) elements that have a particular form” (Perry and Wolf, 1992), where elements can be subdivided into processing, data, and connecting elements; form is defined in terms of their relationship and constraints; and rationale “provides the base of the architecture in terms of system constraints, which most often derive from system requirements” (Perry and Wolf, 1992).

Despite wording it differently, Perry and Wolf’s point of view is shared by Garlan and Shaw (Garlan and Shaw, 1993), who reported that **components** (**elements** in the Perry and Wolf approach) and descriptions of their **relationships** (**connectors** in the Perry and Wolf approach) establish an architecture, and that an architectural

style “determines the vocabulary of components and connectors that can be used in instances of that style, together with a set of constraints on how they can be combined”. The concept of style had already been used with the expression “software architecture level of design” (Shaw, 1990), albeit in an insipid manner.

It should thus be noted that ideas were naturally converging, and it was around this time (*ca.*, 1995) that more attention began to be paid to the true meaning of (software) architecture, in an attempt to reach a consensus (Garlan and Perry, 1995). Finally, in 1996, an increasingly universal vocabulary in this discipline began to emerge (Shaw and Garlan, 1996), with architectural styles “becoming the lingua franca of architecture-level design” (Shaw and Clements, 1997). The ANSI/IEEE Standard 1471-2000 (IAWGroup et al., 2000), which was superseded by ISO/IEC/IEEE 42010:2011 (ISO et al., 2011), also strengthened the definition of architecture, which now reads as follows: “fundamental concepts or properties of the system in its environment embodied in its elements, relationships, and the principles of its design and evolution”.

Although there should be no restrictions on definition or nomenclature, most agree that software architectures are designed to abstract the complexity of a system, reducing it and enabling a better understanding and analysis of it, where the use of a particular style of architecture, or combination of styles into a single design (Garlan and Shaw, 1993), could lead to more powerful analyses than a general architecture would, because it is focused on techniques that are effective on a particular domain (Allen, 1997).

Moving from the design of an architecture to its engineering implementation and application is also a significant step. Aside from the design principles (*q.v.*, Section 3.2) and existing styles (*q.v.*, Section 3.3) of architecture, there are also challenges with regard to enterprise systems (*q.v.*, Section 3.3.3), especially when adapted to mobile environments (*q.v.*, Section 3.3.4).

3.2 Design Principles

Some principles should be considered when designing an architecture in order to meet “best practices, minimise costs and maintenance requirements, and promote usability and extensibility” (Meier et al., 2008), namely:

- **Separation of Concerns**, whose goal is to separate the application in areas of responsibility/functionality such as the user interface from data access. The main advantage of this approach is the potential it entails, namely the ability to correct or optimise the modules independently, without affecting others, as well as being easier to understand and manage;
- **Single Responsibility Principle**, which states that each module or component should only have one responsibility as a single functionality or special feature, allowing for faster modification or optimisation;
- **Principle of least knowledge** or **Law of Demeter (LoD)**, which states that each component or object should have limited knowledge in comparison to other units or sub-components;
- **Don’t Repeat Yourself (DRY)**, whose goal is to avoid duplicating functionality that is already handled by another component or module;
- **Avoid doing a big design upfront (BDUF)**, if the application requirements are not completely clear or if there is a chance that they will change over time, implying going back and thus more labour and time;
- **Prefer composition over inheritance**, because inheritance creates dependency between parents and children, limiting its use.

3.3 Architectural Styles

Various definitions of architectural styles are presented in the literature, including the identification of components and connectors that can be used to constitute a

system based on a set of local or global constraints (Garlan and Shaw, 1993; Shaw and Clements, 1997). Architectural styles have recently been defined as a “collection of architectural design decisions” that can be applied in practice in specific contexts or problems (Taylor et al., 2009). The level of specification of this set of rules and decisions is also proportional to the spread and application of these architectures in real-world contexts. That is, the more general they are, the easier it is to adapt them to real-world scenarios, and *vice versa*, the more specific they are, the more stringent their application.

3.3.1 Benefits & Basic Properties

Perry and Wolf recognised the significance of software architecture styles some time ago (Perry and Wolf, 1992) and began drawing parallels with building architecture. They begin by emphasising the significance of a (general) architectural style from both a descriptive and prescriptive standpoint. While an architectural style specifies elements/components and their relationships from a descriptive point of view, from a prescriptive standpoint, the style constrains the type of elements/components and their relationships, and the same is considered true in software architectures.

Later, they discuss the importance of engineering principles in relation to architecture and styles, citing the perpendicular (Gothic) style of the chapel at King’s College, Cambridge, as an example of how it adds (huge) light over the Romanesque style. The lightness of the perpendicular Gothic is attributable to the application of new engineering principles other than those of the Romanesque. And, once again, a comparison with a software architecture is made, emphasising the importance of its relationship with engineering principles.

Finally, they harvest the concept of the significance of the relationship between architectural styles and materials, where these have specific properties that can be exploited in a specific style. Although they can be combined for aesthetic purposes, they are not flexible enough to allow the construction of a skyscraper out of wooden posts and beams. That is, these materials serve as the engineering foundation for constructing a specific building (architecture) while keeping a specific style in mind,

an idea that can also be applied to software architectures.

The authors conclude that it is possible to build a bridge between a building's architecture and software architectures, pointing out that styles are important forms of description and prescription, and that engineering principles and material properties are the foundation of realising a specific architecture and style.

3.3.1.1 Benefits

According to the book from [Taylor et al. \(2009\)](#), the benefits of an architectural style are:

- **Reuse**, which can be subdivided into:
 - **Design**, referring to the ability to use/apply previous knowledge to solve new problems.
 - **Code**, which can be reused in invariant areas of a particular architectural style.
- **Understandability** of a system organisation through the use of terms (*e.g.*, client-server) that are easily understood by outsiders and do not necessarily require advanced business knowledge;
- **Interoperability**, referring to the capability of features or components to work with other systems without the need for adaptation.

3.3.1.2 Basic Properties

Furthermore, according to Taylor's book ([Taylor et al., 2009](#)), architectural styles have the following fundamental properties:

- **A vocabulary of design elements**, such as component, connector types, and data elements (*e.g.*, peer, object, pipe, client);

- **A set of configuration rules**, that govern the organisation and communication of the elements that constitute the style;
- **A semantic interpretation**, because the elements and their composition (should) have well-defined meanings.

3.3.2 Common Styles

Shaw (1990) and Garlan and Shaw (1993) were the first to introduce some architectural styles and features, as well as the possibility of combining and forming hybrid styles. Soon after, Shaw (1990) name and categorise the styles presented, contributing to the field's vocabulary standardisation once more.

The subsections that follow present some of the most common architectural styles, while also taking into account more recent works (Taylor et al., 2009; Medvidovic and Taylor, 2010).

3.3.2.1 Layered

The layers in the category of hierarchical styles (layered and also known as virtual machines) are essentially a separation of concerns (Dijkstra, 1982), which involves the identification of a system's various concerns and their encapsulation in different modules. In this case, each constituent layer exposes an interface (API) or service to the layer(s) above, acting as a server, and consumes the services provided by the layer(s) below, acting as a client (*q.v.*, Figure 3.1).

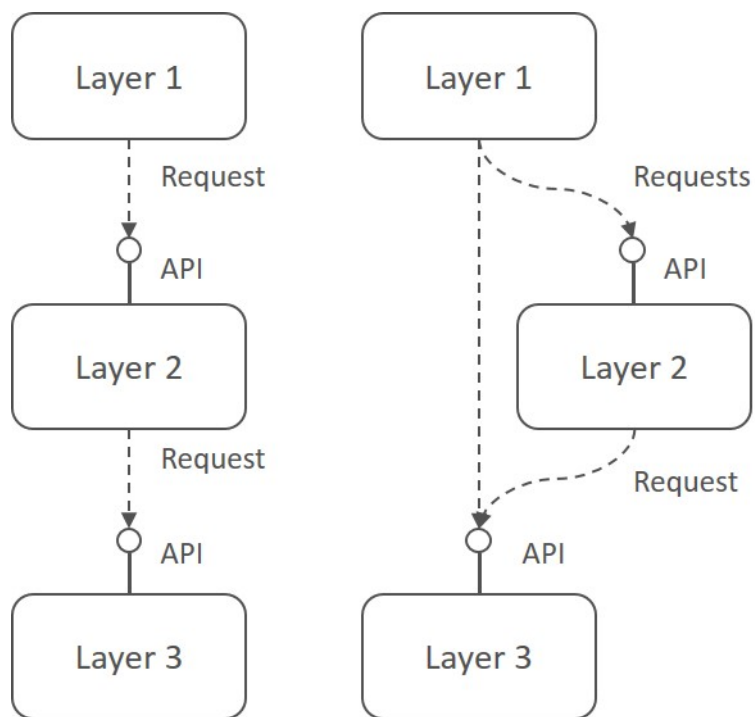


Figure 3.1 – Layer architectural style example based on Taylor et al. (2009).

The Client-Server architecture (*q.v.*, Figure 3.2 on page 123) is an example of this style, which can be understood as a two-layer style with network connections, where the server is responsible for providing the services requested by each client, which are often referred to as “thin” or “fat” depending on the amount of processing (Taylor et al., 2009).

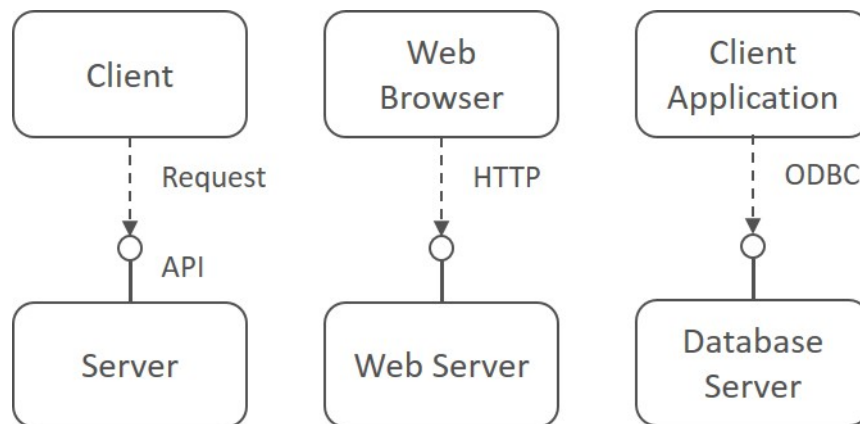


Figure 3.2 – Client-Server architectural style example based on Silva (2008).

3.3.2.2 Blackboard

The Blackboard style (Taylor et al., 2009) arose from the study of Artificial Intelligence (AI), in which various researchers frequently sat around a large problem. When one of these researchers felt confident that he could solve a portion of the problem, he would focus on it and update the blackboard with his solution to that sub-problem. From that point forward, another researcher could use what had been solved to focus on another problem, and so on, eventually solving the “big picture”.

This style essentially keeps a global/central state in the blackboard (*q.v.*, Figure 3.3 on page 124), which is then updated when one of the specialised modules around it solves a portion of the problem, ultimately solving the entire problem.

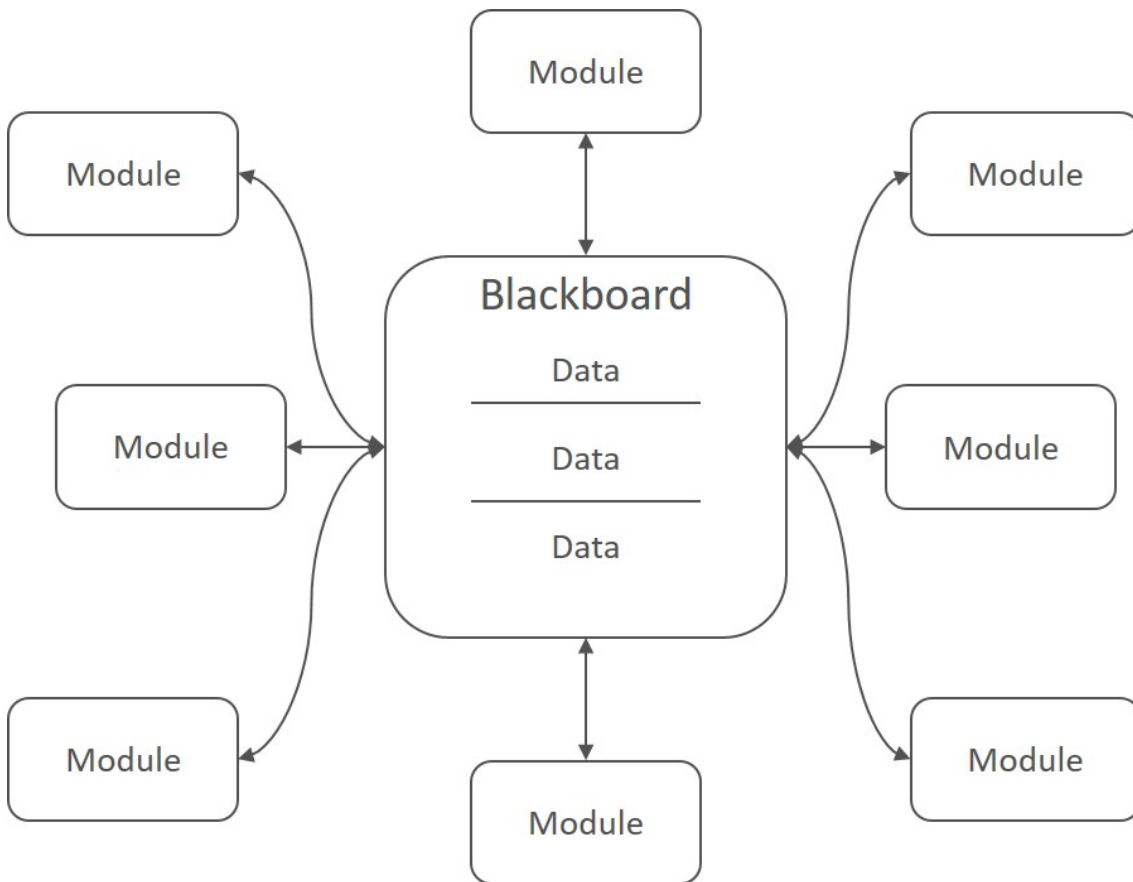


Figure 3.3 – Blackboard architectural style example based on [Houliston et al. \(2016\)](#).

3.3.2.3 Implicit Invocation

The idea behind the implicit invocation (or **Event-driven**) styles is to replace the invocation of functions (*e.g.*, [Remote Procedure Call \(RPC\)](#)) with the announcement of events, where listeners who wait for an event are subscribed to the ones that interest them, associating a function for each, which are then implicitly invoked by the system.

In this style, the one in charge of disseminating events (the broadcaster) is unaware of the consequences of such action, thereby avoiding coupling between the involved components ([Taylor et al., 2009](#)).

Publish-Subscribe

The Publish-Subscribe style is an example of this style, with its name derived from an analogy with a newspaper subscription, in which there are two fundamental elements, the publisher and the subscriber. In this relationship, the subscriber selects the newspaper or magazine to which he wishes to subscribe, and publishers send additional news or information to everyone on the subscription list on a regular basis.

This relationship is reflected in this style. It all starts with the subscribers, who sign up for the themes/specific messages they want, and the publishers are in charge of broadcasting messages to all subscribers (*q.v.* Figure 3.4). Subscriptions may be delegated to third parties depending on the size of the application. If it is a larger application, the publications can occur at system startup, on a regular basis, or on demand (Taylor et al., 2009), and subscriptions must use something like network protocols instead of callback methods (used in smaller applications), and may even require the use of proxies as intermediaries.

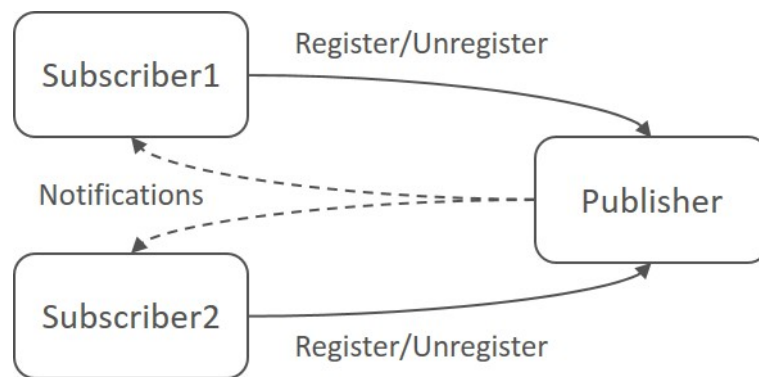


Figure 3.4 – Publish-Subscribe architectural style example based on Taylor et al. (2009).

Event-based (or Event-bus)

Communication is more transparent in an event-based style (also known as event-bus) and the distinction between publishers and subscribers disappears (Taylor et al., 2009). All of the elements or components can send and receive messages asynchronously through the use of pre-existing communication channels (*i.e.*, event

buses *q.v.*, Figure 3.5 on page 126).

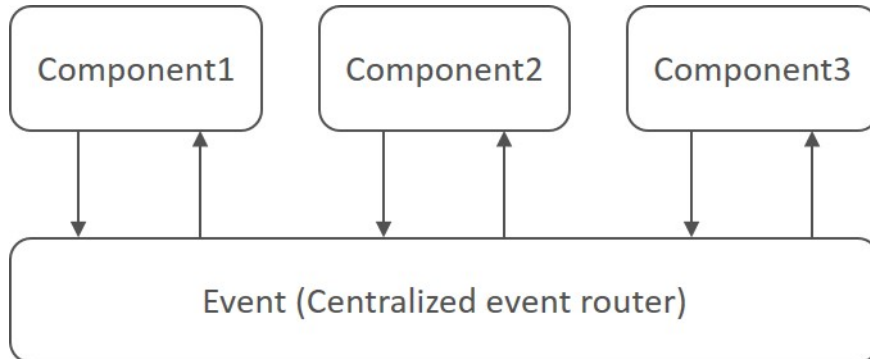


Figure 3.5 – Event-based architectural style example based on Taylor et al. (2009)

3.3.2.4 Peer-to-Peer

The peer-to-peer (P2P) style is based on the Client-Server style. The difference is that there is no longer a distinction between the client and the server because the elements or components can now perform both roles and are thus referred to as peers (*q.v.*, Figure 3.6). As a result, the information is spread across multiple servers, as opposed to being centralised in a single server verified on a Client-Server basis, which can be a significant advantage in terms of how quickly responses to requests are provided.

3.3.3 Enterprise

When discussing the development of Enterprise systems, there are several requirements and challenges that are all interrelated.

Starting with **extensibility**, the ability of a system to grow in terms of new functions or features, which should be considered so that the adaptation of existing features or the addition of a new module or component can be done easily and without affecting the rest of the system.

High Availability of a system is the ability to ensure **performance** even under

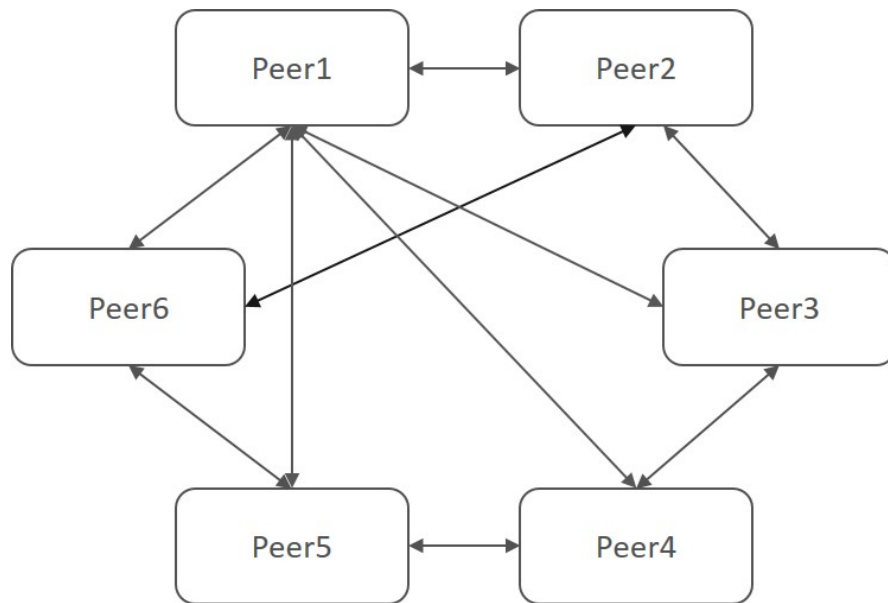


Figure 3.6 – Peer-to-Peer architectural style example based on Taylor et al. (2009)

high loads (*e.g.*, when the number of users is extremely high), where **performance** refers to the ability of a system to meet non-functional requirements such as latency, memory usage, and time response. **Load Balancing** is the process of distributing workload across clusters and other computing resources in order to meet these requirements (*i.e.*, minimise response time and memory usage, and maximise throughput).

Redundancy is the backup of critical components or modules to ensure that the system continues to operate as expected and desired in the event of a failure. It is related to **fault tolerance** because it is the ability of a system to continue to function even when some of its components or modules fail, always trying to adapt to such circumstances as best as possible.

3.3.4 Mobile Oriented

When discussing mobile, it is important to remember that smartphones are no longer the only option. There are several **HMDs** that run on Android, for example. Because there are numerous factors to consider, developing for mobile platforms is a challenge

in and of itself. Thus, in addition to the common design questions and enterprise challenges, there are mobile aspects to consider in order for the application to meet its requirements and perform efficiently and reliably in all possible scenarios.

3.3.4.1 Design Considerations

There are several factors to consider when designing a mobile application to ensure that it meets the requirements and performs efficiently (Meier et al., 2008):

- **Type of client**, depending on the amount of required local processing and the need for a server connection. Although a fat (or thick) client is more difficult to maintain, it is best suited for scenarios in which there is a significant amount of local processing and the application is occasionally connected to a server. A thin client is appropriate when the application relies almost entirely on server-side processing;
- **Target platforms**, which take into account not only the device's size, resolution, and other hardware specifications such as **Central Processing Unit (CPU)** performance, memory, and sensors such as **GPS** and compass, but also the ease with which one can develop for that platform and the implications that entails;
- **Bandwidth**, which takes into account the need for an Internet connection as well as scenarios in which this connection is limited, using strategies such as caching and selecting hardware and more appropriate protocols;
- **User Interface (UI)**, which should always be simpler given that mobile device development has limited hardware and interaction methods vary;
- **Architecture**, taking into account the principles of architecture design and styles that are already available for mobile environments, which typically use a layered architecture. (*q.v.* Section 3.3.2.1);
- **Resources**, adjusting the application to account for the fact that a mobile

device imposes additional constraints such as battery life, memory, and processing time.

4

Methodology

4.1 Introduction

This practical and (slightly) interdisciplinary reflection is an especially exciting challenge because it attempts to implement this concept of interaction by combining two factors to enable non-experts to learn and explore systemic concepts and complexity: (1) enhanced context provided by location-awareness and immersive environments, as well as the rich semantics of gestural; and (2) somatic input based on cultural gestures and social practices. With this combination, it is expected that a computational system that uses them will enable non-expert users to perform in-depth explorations of interconnected concepts and data in a contextual, location-based fashion, for a better understanding of systemic interactions. Despite the fact that the various fields required for this implementation (*e.g.*, HCI, Software Engineering, Algorithmics) are distinct, they end up complementing one another. Because a system develops emergent, yet absent, behaviour from each individual component, it necessitates a concerted effort to achieve a final result that meets the initially raised expectations: the approach's success.

There are methods described in the scientific literature for combining the above

efforts, particularly metamethods as general approaches within which specific methods of each area can be used, which is essential to ensure that a perspective does not remain isolated in their area of comfort and thus be able to interact with every field. The rest of this chapter goes over the initial set of objectives (*q.v.* Section 4.2), the metamethod used to tackle those objectives (*q.v.* Section 4.3), and how that approach was used effectively throughout (*q.v.* Section 4.4).

4.2 Objectives

The primary goal is to investigate how the conjectured cross-approach can be realised (*i.e.*, context provided by immersive environments and semantics provided by gestural and somatic sociocultural practices) for the development of HCI approaches and systems that enable end-user modelling and interaction with complex systems and systemic concepts. The following tasks, which we consider to be performance objectives, were defined at the outset:

- Identification of case studies with actors willing to participate in the research;
- Survey and characterisation of resources, as well as relevant case study characteristics (*i.e.*, link between concepts and location, cultural gestures, rituals and social habits of the participants);
- Initial model design and prototyping for combining the data and forms of interaction;
- Validation or refutation of the initial model, by testing the initial prototype;
- Model refinement through iterative prototype redesign and testing;
- Draw conclusions about the feasibility and potential of this method for creating interactive immersive environments for complex learning that realise this combined approach, using the results of successive iterations.

These tasks, or goals, can be approached in a variety of ways. This is largely determined by our worldview, particularly on the adopted philosophical paradigm.

Table 4.1 – Comparison between Positivism and Interpretativism in light of four key philosophical questions. Adapted from [Figueiredo and Afonso \(2006\)](#).

Questions	Answers	
	Positivism	Interpretativism
Ontological question what can be known?	realist hypothesis	phenomenological hypothesis
Epistemological question what is knowledge?	deterministic hypothesis	teleological hypothesis
Methodological question how can we build that knowledge?	principle of analytical modelling principle of sufficient reason	principle of complexity principle of intelligent action
Ethical question what is the worth, or value, of the knowledge we create?	neutral, independent, and value-free stance	context dependent and important among all parties involved

There are several paradigms, but Positivism and Interpretativism, often referred to as Constructivism, are the most prominent in the literature. To distinguish the two, one can consider how they respond to some philosophical questions (*q.v.* Table 4.1): ontological, epistemological, methodological, and ethical ([Guba et al., 1994](#); [Figueiredo and Afonso, 2006](#)). The ontological question is concerned with what can be known. The epistemological question asks what knowledge is and how we can obtain it. The methodological question is concerned with how we can acquire that knowledge. The ethical question asks what the worth, or value, of the knowledge we create is ([Piaget, 1967](#), p.6), and it is part of a larger axiological question about how we value things (morally and spiritually), or, more formally, what elements can contribute to the intrinsic value of a state of affairs.

The essence of our problem, complex systems and concepts, naturally leads us to a less objective and realist philosophical paradigm, in the sense that things may not be perceivable or indeed exist at simpler levels (*i.e.*, realist hypothesis, with

reality being external to the interactions). According to the phenomenological hypothesis, on the other hand, things can only be known through their interactions with us, depending on our perception of the world, our notions, which are built into our conscience. In terms of epistemology, Positivism is more objective, stating that everything happens for a reason, leading to the possibility of explaining anything in a unique and permanent way. The teleological hypothesis is used to respond to Interpretativism, which claims that knowledge is constructed for a reason and is what gets us somewhere. Methodologically speaking, Positivism, through the principles of analytical modelling and sufficient reason, is consistent with the traditional divide and conquer and reductionism approach, where there is no effect without a cause, opposing the complexity theory that defends holism instead. The Interpretativism paradigm considers this theory via the principle of complexity. This principle acknowledges the need to recognise reality as complex systems in constant change, embodying both stability and change, where the whole is greater than the sum of the parts, chaos coexists with order, and parts interact with each other in the shared construction of reality, in an emergent and largely unpredictable manner. The other part of the answer, through the principle of intelligent action, described by Schon (Schön, 1983) as "reflective action" and "reflection-in-action," leads to the typical world of practitioners, which is heavily based on trial and error, reformulating problems as they are solved, in a constant dialogue between the problem solver and the situation being solved. Finally, when it comes to the ethical question, which is part of a larger axiological question, of how we value things and how they guide us, the Positivism paradigm takes a neutral, independent, and value-free stance. The Interpretativism paradigm, on the other hand, believes that the outcomes are context dependent and important not only for those who ask the questions, but also for those who provide the answers. Both paradigms continue to foster trust and respect among all parties involved.

According to the Interpretativism philosophical paradigm, reality is made (also) of what we make of it, our worldview, and how we interact with it. How it responds to the set of philosophical questions posed above naturally leads us to the adoption of a more qualitative metamethod, Design Science (*q.v.* Section 4.3), that allows such perspectives, constant trial and error, and reformulation, while not losing sight of

the importance of the outcomes to everyone involved.

4.3 Design Science

There are metamethods described in the literature that allow specific methods of each area to be used for the combination of efforts and fields, which is essential to ensure that a perspective does not remain isolated and thus interacts with the other fields. We used a metamethod that was already well-known for this purpose. Design Science, as formalised by Hevner (2007) and systematised by Peffers et al. (2007), enables the conjugation and combination of different areas and perspectives, with sufficient detail to enable its constructions and evaluation across multiple dimensions.

The contributions of various disciplines can be brought together under the principles of Design Science and its three parallel cycles (*q.v.* Figure 4.1), or, in Hevner's words (Hevner, 2007), the three cycle view of science research design:

- **Relevance**, which analyses the application domain's environment with the aim of supporting designs that improve it;
- **Design**, which creates mechanisms and processes, evaluating them;
- **Rigour**, which provides a theoretical foundation for the previous cycles (Relevance and Design), transforming their outcomes into more knowledge—something that is always intended.

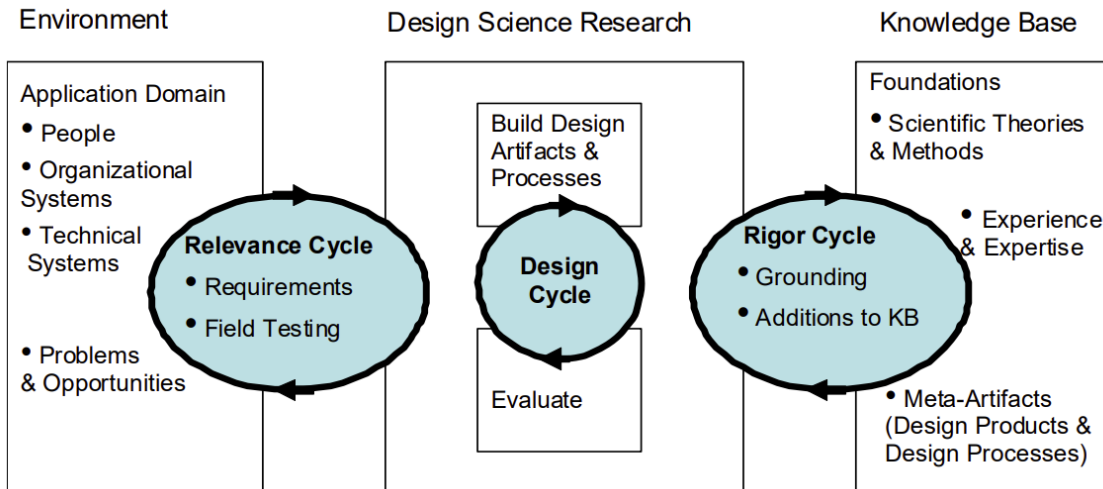


Figure 4.1 – Picture taken from [Hevner \(2007\)](#). Design Science Research Cycles.

[Peffers et al. \(2007\)](#) proposed a six-step systematisation of these cycles (*q.v.* [Figure 4.2](#)), which is used to describe the research process here. However, as its authors suggest, the process will be iterative, as the evaluation phase generates knowledge that aids in identifying improvements in the earlier stages:

1. Identify the problem Motivation

- (a) Update the state of the art based on Hevner’s three perspectives: relevance of the problem, relevance of the approach, characteristics of methods and architectures that will influence the design, and rigour—essential features to meet in the design process, evaluation methods, and analysis for the evaluation phase;
- (b) Identify a case study with actors willing to collaborate in the research, based on the most recent state of the art (*q.v.* 1.1).

2. Define Objectives of a Solution

- (a) Use the state of the art to conduct a survey and characterisation of resources and relevant case study characteristics (the link between concepts and location, cultural gestures, rituals and social habits of the participants).

3. Design and Development

- (a) Design an initial model by combining the data and interaction methods;
- (b) Develop an initial model prototype.

4. Demonstration,

- (a) Conduct tests with the initial prototype, in the identified case study, to ensure its functionality.

5. Evaluation,

- (a) Observe and measure the results of using the prototype to confirm the quality of resolution it provides to the problem;
- (b) Draw conclusions about the cross approach's feasibility and potential.

6. Communication,

- (a) Write up the findings and report them to the scientific community through peer review and publication.

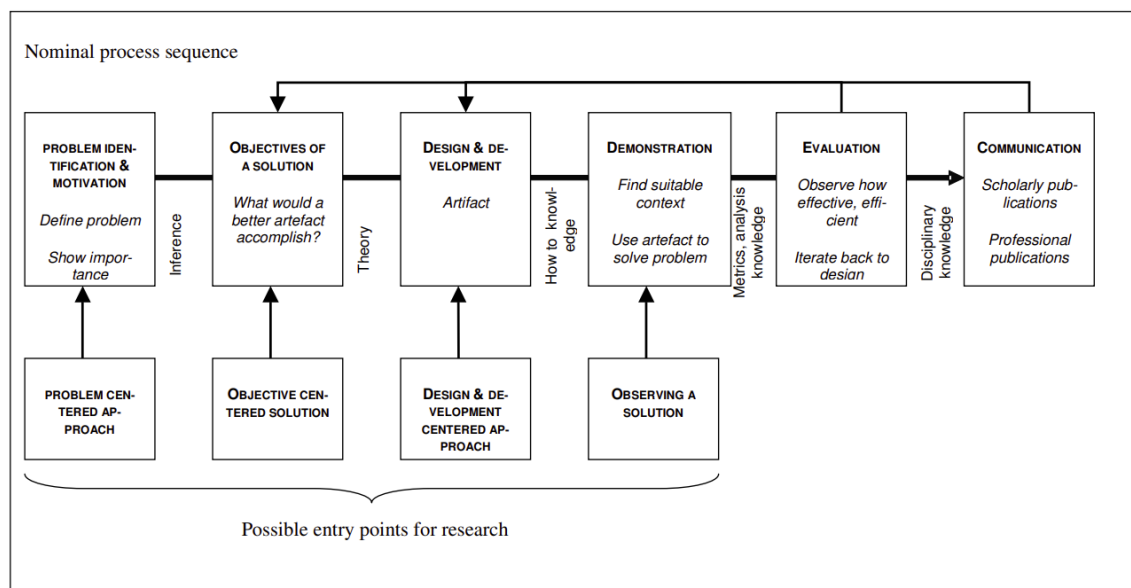


Figure 4.2 – Picture taken from [Peppers et al. \(2007\)](#). Design science research process (DSRP) model.

4.4 Practical Application

This chapter concludes with a practical application of the chosen and previously described Design Science metamethod. In other words, it goes through all of its cycles and systematised stages (*q.v.* Section 4.3), describing what was done, how it was approached, and the results.

The first year of this doctoral program required, aside from attending classes, the detailing of the problem and the establishment of conjectures on how to solve it, as well as the formal definition of the overall thesis' objectives (*q.v.* Section 4.2).

The second academic year, 2017-2018, began with the development of a broad overview of the HCI literature, terminology, and prominent research teams from around the world, as well as their main works. This resulted in the first iteration of the 1.1 phase (*q.v.* Section 4.3), a mental map (*q.v.* Figure 4.3) created with Coggle¹.

¹ <https://coggle.it/>

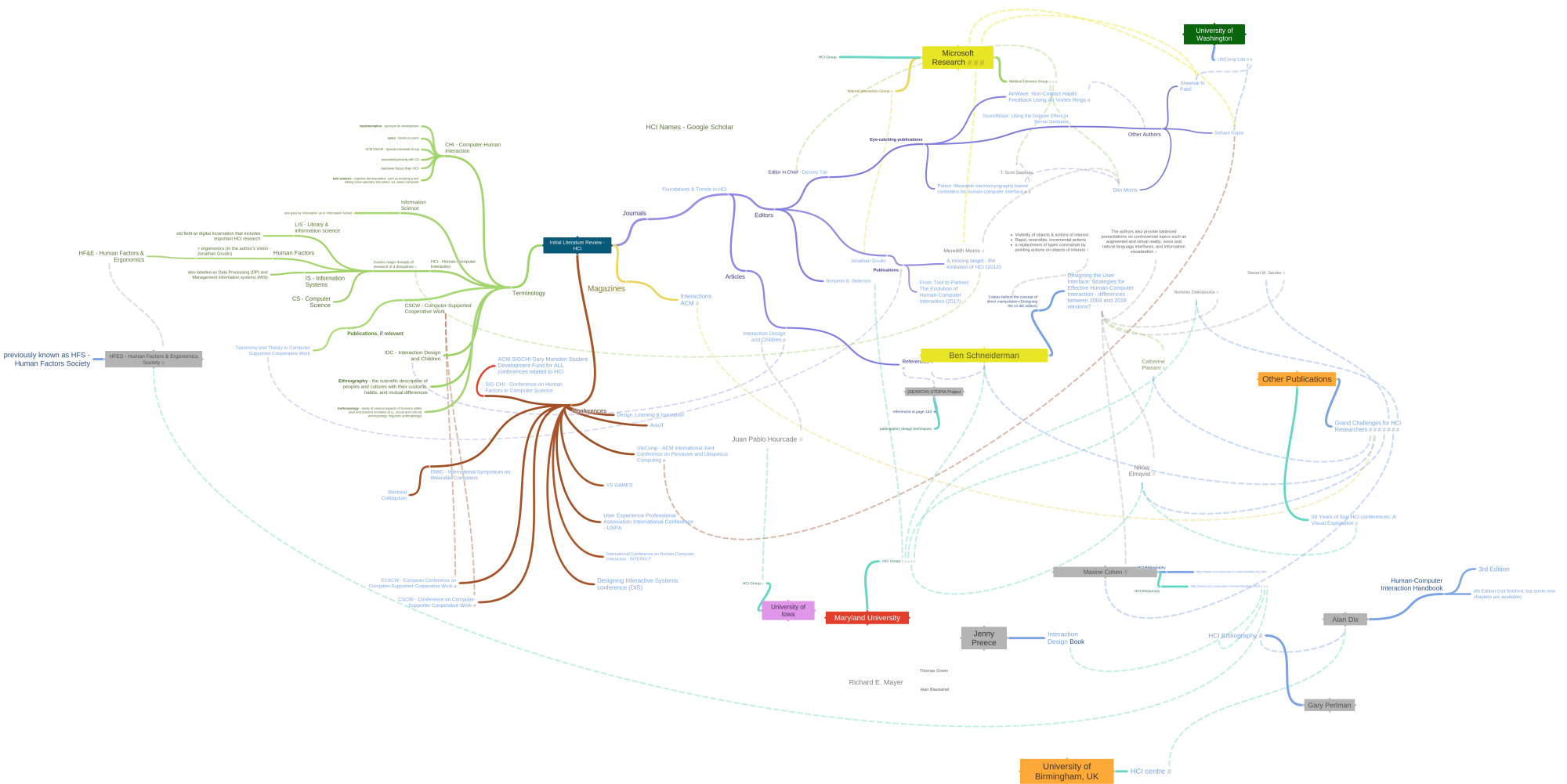


Figure 4.3 – HCI mental map made with Coggle.

In the second half of this academic year, we were given the opportunity to submit a proposal for a Fulbright scholarship, co-funded by the [Foundation for Science and Technology \(Portugal\) \(FCT\)](#), to take place in the [United States \(US\)](#) in the year after that. We had already made contact with some researchers and universities at the iLRN 2017 conference in Coimbra, Portugal, and discussed the possibility of collaborating and using the Flathead Indian Reservation in Montana, [United States of America \(USA\)](#), as a case study. We began our six-month stay in the United States after the Fulbright proposal was accepted (a first iteration of phase [1.2](#)), which formally explains the relevance of the [HCI](#) problem and how the stay at the [CSKT](#) could leverage its solution through ethnographic work.

During this time, already in the third academic year, we were looking for something deemed complex that was of particular interest to the tribes, in addition to collecting anthropologically-related information, such as particular aspects of their culture, such as their habits, customs, gestures, rituals, and mutual differences. After three months, we finally found a collection of children's books that go through (the significance of) some complex concepts, such as the cycle of life, family bonding, and the emergent nature of the Geese V-shaped formations and how the Native people gained insight from watching them migrate. We used this to improve the case study proposal (second iteration of phase [1.2](#), and first iteration of phase [2](#)), to focus the remaining time on learning more about specific aspects of these stories, and to submit a proposal (*q.v.* [Appendix B.2](#)) to the [Institutional Review Board \(IRB\)](#) and the tribal cultural committees to detail how formal interviews would be conducted (*q.v.* [Appendix B.4](#), [B.5](#), and [B.3](#)). We began a multimodal analysis of the stories and formal interviews (first iteration of phase [3.1](#), *q.v.* [Section 5.3](#)), while also exploring options for [HCI](#) interactions (using the collected gestures and rituals) and [3D Mapping](#) (first iteration of phase [3.2](#)) by contacting the Geography Departments from the [CSKT](#), and the University of Montana, which eventually led to contacts with a Computer Graphics research team at [Friedrich-Alexander University \(FAU\)](#), Erlangen, Germany, who had prominent work on the use of Photogrammetry. During the summer vacations of 2019, as a Guest Researcher at [FAU](#), some [3D Mapping](#) options were investigated in greater depth, while beginning a [SLR](#) (second iteration of phase [1.1](#), *q.v.* [Chapter 2](#)) and still exploring potential [HCI](#) techniques (*q.v.* [Section](#)

implementation examples leap motion earthquake), a second iteration of phases **2.1**, and **3**. Meanwhile, the iLRN 2019 conference was held, and the official proposal for the **CSKT** case study was published (first iteration of phase **6**).

During the fourth academic year, 2019-2020, there was also the opportunity to continue the exploration of immersive environments and **HCI** techniques as a Guest Researcher in the **Bahçeşehir University (BAU) VR First Lab** in Istanbul, Turkey, during the winter vacations. This provided access to cutting-edge technology, such as the most recent versions of **HMDs**. This occurred while the **SLR** was still ongoing (still the second iteration of the phase **1.1**). During that time, a “Interactive Virtual Museum” workshop was held at the **VR First lab**, which resulted in the collaboration of several people and the exploration of **HCI** and immersion techniques, as well as the creation of a later published **VR** application for the Oculus Rift, titled “2020: Interactive Art Collection”².

As we all know, the **COVID-19** pandemic spread to the entire world in early 2020, prompting the closure of all non-essential institutions, including universities and schools. Nonetheless, this time was spent continuing and finishing the **SLR** (third iteration of phase **1.1**, and soon to be published as per the second iteration of phase **6**), prototyping the animated concept (*q.v.* Section **8.3**), and conducting some **VR** experiments with the Oculus Quest (third iteration of the phase **3**). Because of the pandemic, we were unable to test these prototypes remotely (which would be the first official iteration of phases **4**, and **5**) at the Flathead Indian Reservation, as we had originally planned. The supporting sections of this thesis were written over the course of this time.

Despite the repeated reopenings and closures of academic institutions during the final academic year, 2020-2021, iterative prototyping from home continued, resulting in the creation of an improved version of the InMERSE Framework (*q.v.* Chapter **7**), a **VR** animated concept (*q.v.* Section **8.3**), and a **VR** application prototype (*q.v.* Section **8.4**). Official tests could not be carried out at the **Salish Kootenai College (SKC)** as planned (which would be a second iteration of phases **4**, and **5**). We chose to conduct some final tests (*q.v.* Chapter **9**, and Questionnaire **B.5**) with a

² <https://www.oculus.com/experiences/rift/3028457210604660/>

strictly limited number of people, primarily family members and close friends, for health and safety reasons (which would be a third iteration, but ended up as the only possible iteration of phases 4, and 5). Following that, the writing of this thesis was completed.



CSKT Case Study

5.1 Introduction

The CSKT Case Study can be considered this thesis' foundation goal, because it seeks meaningful sociocultural context to apply in a prototype (*q.v.* Chapter 8 - CSKT Prototype). This began with a six-month stay at the University of Montana and SKC in Montana, USA, as part of a Fulbright scholarship jointly funded with FCT. The overarching goal of conducting research in the USA was to assess the feasibility of using cultural gestures and symbols to facilitate more meaningful interaction in immersive environments (*q.v.* Section 2.2.4). The aim is to use users' cultural semantics to improve complex learning (*e.g.*, complex concepts or systems) by creating interactive stories or modelling interactivity in immersion.

The intention of this field research was to find something that not only made sense in their cultural context, but also matched our use-cases. It also included writing an official proposal for the tribes' cultural committees (*q.v.* Section 5.2) explaining the idea and process. After locating a child book story of interest, Luli's Journey (Pierre, 2010b), we conducted a multimodal analysis (*q.v.* Section 5.3) to collect facts from it in a less subjective manner. This is the skeleton of what would become the CSKT Prototype (*q.v.* Chapter 8).

5.2 Proposal

The stay in the **US** began with an understanding of the work done by various departments, as well as an attempt to learn more about the Flathead Indian Reservation tribes, specifically the Salish, Kootenai, and Pend d'Oreille (who are also known as Kalispel). The tribes have strong ties to the spiritual world as well as nature, specifically the land, wildlife, and water. Nature provides, and they return the favour by caring for it and all that it entails. Hunting is an excellent example. It is more than just an activity or a sport to them. Rather, it is a sacrifice with profound spiritual significance that they carry with them throughout their lives. They eat the meat and use the bones in exchange for taking care of the wildlife (a good example is the buffalos). To this day, only tribe members are permitted to hunt on the Flathead Indian Reservation, despite the fact that they are still not allowed to manage the National Bison Range, which was created by the government against the will of the Natives.

Elders from the tribes continue to use storytelling as the primary method of passing down knowledge (and complex ideas such as life lessons) from generation to generation, just as 100.000 years ago. And they do so in certain circumstances, such as on a specific day of the year, in a specific location, or while staying at home during a snowstorm in the winter. Some of those stories are now being published as children's books, one of which came to our attention nearly three months after we arrived. An Elder from the Pend d'Oreille tribe told this story, titled Luli's Journey ([Pierre, 2010b](#)). It's about Luli, a small Canadian goose, and her family's migration from Flathead Lake. When winter approaches, they fly to Mexico in a V-shaped formation (*i.e.*, complex flight dynamics), each taking turns in the lead, and return after winter's final blow. According to the story, the birds' knowledge of seasons is so precise that the Natives relied on them to survive and thus observed them from below. The annual migrations of Canadian Geese provided information about weather patterns as well as timelines for season changes.

The **CSKT** prototype is based on this story, with the goal of empowering users to understand the evolving nature of the complex concepts it contains, such as the

cycle of life, family bonding, geese V-shaped flight formations, and insights gained in the past from observing the birds (*e.g.*, weather patterns). It is also possible to broaden it to include more general topics, such as climate change and its effects on the food chain.

The remaining time was spent putting together a proposal for the tribes' Cultural Committees (*q.v.* Appendix B.2), which was required to conduct official interviews (*q.v.* Appendix B.5), as well as collecting symbols, gestures, and rituals of the tribes and exploring ways of 3D mapping some parts of the reservation (*e.g.*, Satellite data, Photogrammetry). For example, each tribe has its own set of constellations, but the Goose is shared by all three. Native Americans used constellations to determine a specific point in time (*i.e.*, year, month, day). Geese, in particular, appear to navigate using constellations. Other examples include specific ways of referring to the past and future, as well as signs that have some similarities to Native American Sign Language.

According to the original proposal's rationale, the CSKT storytelling prototype (*q.v.* Chapter 8) should use immersive technology and somatic interaction in the visual-spatial context of the Flathead Indian reservation, and allow interaction through socioculturally aware gestures and emblems. Changing the current season from winter to summer, for example, and reflecting that through landscape changes within the immersive environment.

5.3 Luli's Journey Multimodal Analysis

For this multimodal (text and image) analysis of Luli's Journey's book, a protocol based on Cruz et al. (2016), was used. The overarching goal of this analysis is to provide enough theoretical support for the data collected from this book through a less subjective process. In other words, to systematically justify why (*i.e.*, its relevance) and how (*i.e.*, facts) the data is collected, to leverage the subsequent VR animated concept and implementation prototype, which attempt to use the learnings of this analysis and propel them further with our main conjecture (*q.v.* Section 1.2).

5.3.1 Research Questions

The following research questions (RQs) originate from the literature on complexity and complex systems (*q.v.* Section 2.2.1 - Complexity), where two more general clusters were found:

- **RQ1:** What type of **Humanities, Arts, and Social Sciences (HASS)** related complex concepts/systems are present in the story?
- **RQ2:** What type of **Sciences, Medicine, and everyday life** related complex concepts/systems are present in the story?

5.3.2 Propositions

Each proposition is referent and can be applied to both **HASS**, and **Sciences, Medicine, and everyday life** domains, unless specifically labeled otherwise. The propositions stem from the features/characteristics that make systems/concepts deemed complex (*q.v.* Section 2.2.1 - Complexity):

- **P1:** There are references to collective phenomena that are not the simple sum of the individual behaviours or characteristics of the constituent elements;
- **P2:** There are aspects in the story that are intertwined and inter-connected not only within a system but also with the surrounding environment;
- **P3:** There are aspects in the story referring to a large number of elements, interacting at various levels, which result in the emergence of self-organisation, thus showing a hierarchical and asymmetrical structure;
- **P4:** There are aspects in the story referring to feedback loops.

The domain duality can thus be unthreaded into the complete list of specific propositions (*q.v.* Table 5.1, below).

Table 5.1 – Propositions specification

HASS specification	Sciences, Medicine, and everyday life specification
P1 HASS	P1 Sciences, Medicine, and everyday life
P2 HASS	P2 Sciences, Medicine, and everyday life
P3 HASS	P3 Sciences, Medicine, and everyday life
P4 HASS	P4 Sciences, Medicine, and everyday life

5.3.3 Units of Analysis

Units of analysis must be defined and collected in order to justify and support each of the preceding propositions. They refer to information that can be gathered to back up the propositions and thus answer the research questions. To help with the choice of units of analysis, we referred to studies on the quality analysis of text and picturebooks (Kelly and Kachorsky, 2021). They accomplish this by combining this process with social semiotic theories of multimodality, thereby taking into account “socially shaped and culturally given resources for making meaning” (Jewitt, 2011, pp. 54), such as written text, images, rituals, gestures, and music. We thus adapt their multimodal ensemble analysis tool by using the metafunctions from which they originally derived their tool, such as the compositional system (*i.e.*, how an image is composed, arranged, and framed), the representational system (*i.e.*, what is represented within the image), and the interpersonal system (*i.e.*, visual structures that position the viewer in certain ways), as well as intermodal and ideological considerations (*i.e.*, the connections between the content within the spreads and sociocultural context in which the picturebook is embedded), though we included the latter two within the intramodal considerations rather than separately:

- Intramodal considerations
 - Compositional Inventory: Textual Elements
 - * **Linguistics:** What language and words are used?
 - Representational and Compositional Inventory: Textual and Visual elements

- * **Characters:** Who is in the scene? (*e.g.*, characters, age, race, gender)
- * **Pose:** How are the characters posed?
- * **Objects:** What objects other than people are included in the image, or described in the text? (*e.g.*, fireplace, symbols)
- * **Actions:** What actions or activities are the characters performing? (*e.g.*, gestures, rituals)
- * **Setting:** What is the context of the environment? (*e.g.*, day/night, season, weather)
- * **Emotions:** What emotions do the characters show, or are described in the text? (*e.g.*, happiness, longing)
- Interpersonal Inventory
 - * **Gaze:** Do the characters look at the viewer (demand) or away (offer)?
 - * **Interpersonal Distance:** How are characters positioned in relationship to each other? (close personal, far personal, public)
- Image-Text-Design Inventory
 - **Ideational Concurrence:** Image/text present similar information (some degree of redundancy)
 - **Instantiation:** Image displays one instance of the text
 - **Exemplification:** One serves as an example of the other
 - **Ideational Complementarity:** Joint/synergistic contributions to meaning potentials
 - **Ideational Counterpoint:** Text/image present contradictory information

5.3.4 Evidence Collection

The evidence was gathered primarily from Luli's Journeys child book (Pierre, 2010b), which contains both written text and images. Aside from the cover, this book always

presents a page with a dialogue, followed by a page with a related full-sized picture. Taking this into consideration, evidence was gathered for each of the two pages. That is, written dialogues and full-sized images were always considered for evidence collection in pairs, as a whole. Table 5.2, below, compiles the evidence gathered for each unit of analysis for a single dialogue and associated image. Not all units of analysis need to be filled. Appendix B.1 contains the remaining tables.

Other books were read as an extension. *Blue Jay Brings the Chinook Wind* (Pierre, 2010a), in particular, originally written by Bon Whealdon of the Montana Writers Projects in the 1930s, has an appendix detailing the storytellers of the past and where they lived on the Flathead Indian Reservation. Given that our concept also includes telling a story, we gathered evidence from that specific appendix (*q.v.* Table 5.3). We also obtained evidence from an interview conducted on the Flathead Indian Reservation (*q.v.* Appendix B.5.1), as shown in Tables 5.4, 5.5, 5.6, 5.7.

Some evidence gathered on a specific page sometimes leads to the assumption of some information via induction and deduction. Luli's nest, for example, is located near McDonald Lake. Luli is portrayed swimming on the following page, but no explicit reference is made. We can make an educated guess that it is, in fact, McDonald Lake. While evidence is gathered factually from each pair of pages, as previously stated, the story and its narrative as a whole must not be neglected. It is interconnected, and each page loses semantic information when regarded separately. And that is why context matters.

This first image (page 3 of the book, *q.v.* Appendix B.5.2, Figure B.1) depicts some birds whose species we were unsure of. Frank Finley, a Salish tribe member and Professor at Salish Kootenai College (CSKT) on the Flathead Indian Reservation, confirmed (*q.v.* Appendix B.5.2) that the little iridescent blue is a male mountain bluebird, and the black and white with some blue traces is a black-billed magpie. He also mentioned that the presence of a peacock in the picture makes no sense, as they are not common there, other than to serve as an example for the dialogue that mentions the existence of both big and small birds.

Table 5.2 – Example of collected evidence for each unit of analysis, for a single dialogue and respective full-sized picture of the book Luli’s Journey.

Unit of Analysis	Evidence (page 2 and 3 of Luli’s Journey)
Intramodal considerations	
Linguistics	Salish; “Some are big and some are small”; “Luli is different”
Characters	Birds (peacock, snowy owl, duck, Black-billed magpie, mountain bluebird); Canadian Goose (Luli as yellow gosling)
Pose	Goose standing; Duck on the floor. Remaining standing on branches;
Objects	Several flowers (roses or tulips, Arrowleaf Balsamroot’s, lavender)
Actions	Just standing
Setting	Reservation; Day (sun); Clear sky (blue); Green mountains;
Emotions	N/D
Gaze	Luli looking at the reader
Interpersonal distance	Around Luli (not too close)
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table 5.3 – Collected evidence for each unit of analysis, for the appendix on the storytellers of the past of the book *Blue Jay brings the Chinook Wind*.

Unit of Analysis	Evidence (page 34/35 of <i>Blue Jay Brings the Chinook Wind</i>)
Intramodal considerations	
Linguistics	English; “Storytellers of the past” =>where they would tell their stories, including <i>Blue Jay</i> (even illustrated by Red Horn)
Characters	Lassaw Red Horn; Alexander Beaverhead; Eneas Michel Conko; Philip Pierre; All Pend d’Oreille elders
Pose	All standing outside of their houses
Objects	Trees, horses
Actions	N/D
Setting	Mountains on the back (Eagle Pass in sight); South Crow Creek area, south of Ronan, Montana, near the Kicking Horse Reservoir, located on the flathead indian reservation
Emotions	N/D
Gaze	N/D
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table 5.4 – Collected evidence for interview to Salish Student in CSKT – Appendix B.5.1.

Unit of Analysis	Evidence (Interview to Salish Student, Part 1)
Intramodal considerations	
Linguistics	English; “It’s the circle of life through one story”; “It solidifies the culture”; “It’s our main teaching point”; “... taught by experiencing it”; “They transmit knowledge from generation to generation”; “Other tribes have their own mischief maker”
Characters	Coyote; Snake, and bird; Owl
Pose	N/D
Objects	N/D
Actions	Teaching about mischief, strength, and growing, just to get on the right path; The coyote drawing you to veer off the right path.
Setting	After the first snowball; Cannot be told after Spring begins. There is a time period: the down-months. They are told in a way that it will stick with you; It’s the respect that comes with them, ... it’s holding your beliefs so strong, so that you don’t tell them in summer, or spring. It’s a bad omen otherwise—cultural beliefs— such as the owl being a messenger of bad news. Added in with aspects of mythicality to be a teaching point. No script. Every person has different outlooks.
Emotions	N/D
Gaze	N/D
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table 5.5 – Collected evidence for interview to Salish Student in CSKT Part 2 – Appendix B.5.1.

Unit of Analysis	Evidence (Interview to Salish Student, Part 2)
Intramodal considerations	
Linguistics	English;
Characters	Salish Bitterroot
Pose	N/D
Objects	Madison Tree
Actions	Moving
Setting	Bitterroot Valley to the Flathead Indian Reservation
Emotions	anguish/injustice
Gaze	N/D
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	N/D
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table 5.6 – Collected evidence for interview to Salish Student in CSKT Part 3– Appendix B.5.1.

Unit of Analysis	Evidence (Interview to Salish Student, Part 3)
Intramodal considerations	
Linguistics	English; “You go through the basics of colours, numbers, and then stories in salish”; “strengthens our salish”; That (birds’ migration) definitely popped through, but when you are that young you don’t really grasp the concept”;
Characters	Salish Teacher (Salish tribe); Native people; Bears;
Pose	N/D
Objects	Luli’s Journey child book; Food, shelter (e.g., teepees), blankets; Snow (as symbol of winter coming); Lake (runoff, as symbol of spring)
Actions	Teaching Salish through the story; Native people start preparing blankets, food, shelter, and settle down for winter; Bears hibernate
Setting	Not a specific day or month (for teaching); Before winter arrives, people prepare; Bears prepare for winter as well, and hibernate; Snow comes down the mountains, until it’s sort of a blanket across.
Emotions	N/D
Gaze	N/D
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table 5.7 – Collected evidence for interview to Salish Student in CSKT Part 4– Appendix B.5.1.

Unit of Analysis	Evidence (Interview to Salish Student, Part 4)
Intramodal considerations	
Linguistics	English; “I know, but I can’t, like, name them (gestures)”;
	“Not really aware of doing that (such gestures)”;
	“it (culture) fades out after a while (the awareness of such aspects)”
Characters	Salish member
Pose	Thumbs up (or fist up) for ‘yes’; Praying gesture for ‘wishing luck’ (and waving for ‘thank you’);
Objects	“Certain songs for certain times.”; “Grass songs wouldn’t be played in the winter because there is no grass”;
	“We pray for thanks, we pray for safe travel, you pray for guidance, safekeeping”
Actions	‘Growing’ gesture for plants, kids, etc; Nodding for ‘no’; ‘Sky’ going up with both hands;
	“Both hands up, going opposite ways could be ‘open’, but if looking up, also sky”;
	‘Flying’ with thumbs moving over each other; ‘Getting back down’ by pulling it to you, almost. “‘Bird’ and ‘fly’ are kinda the same, because you always connect birds to flying”;
	Turning a page for ‘changing season’;
Setting	Change of season happens gradually. During fall leaves drop. During spring trees pop. During winter, snow starts coming down the mountain.
Emotions	Nostalgia (for remembering culture)
Gaze	N/D
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

5.3.5 Propositions' Support

The evidence gathered from various sources was connected in order to support each of the enumerated propositions (*q.v.* Section 5.3.2 - Propositions). Because the overall idea of the connections can still be interpreted if left alone, explanations were added before the diagrams containing ideas we consider to be more subjective or may require some reasoning. The first two diagrams, which refer to P1 HASS and P1 Sciences, Medicine, and Everyday Life, respectively, are good examples of this. P1 HASS (*q.v.* Figure 5.1) is divided into two connections. The upper one, “Evolution of language” refers to the Salish language, specifically. When viewed from a certain angle, any language can have complex properties, and storytelling can be used to convey complex ideas. The Salish, however, do it in a non-linear fashion. The Salish language evolved in such a way that allowed it to not only be the primary means of transmitting knowledge from generation to generation, but also the tribe’s cultural values, one of which was the language itself. There is a desire among the tribes to connect with nature, the land, wildlife, and water. And that transpires to the way they describe their surroundings in their language, which they then attempt to pass on through stories. For the tribes, language is more than just a means of communication. Rather, it is how they think and feel about everything around them; it is an important part of their identity. The second P1 HASS idea, “Family & Home bonding”, could be viewed as a complex concept grounded on the biological strategy to adapt to one’s surroundings. The idea, however, is the value and significance they represent to the CSKT tribes. A location is not meaningful on its own. A place of birth or growth, on the other hand, can have sentimentality and meaning attached to it. That meaning is established by a person or a bird. The same is true for family. There are a lot of people or birds. But those who they consider to be their family, on the other hand, share a deeper and more meaningful bond.

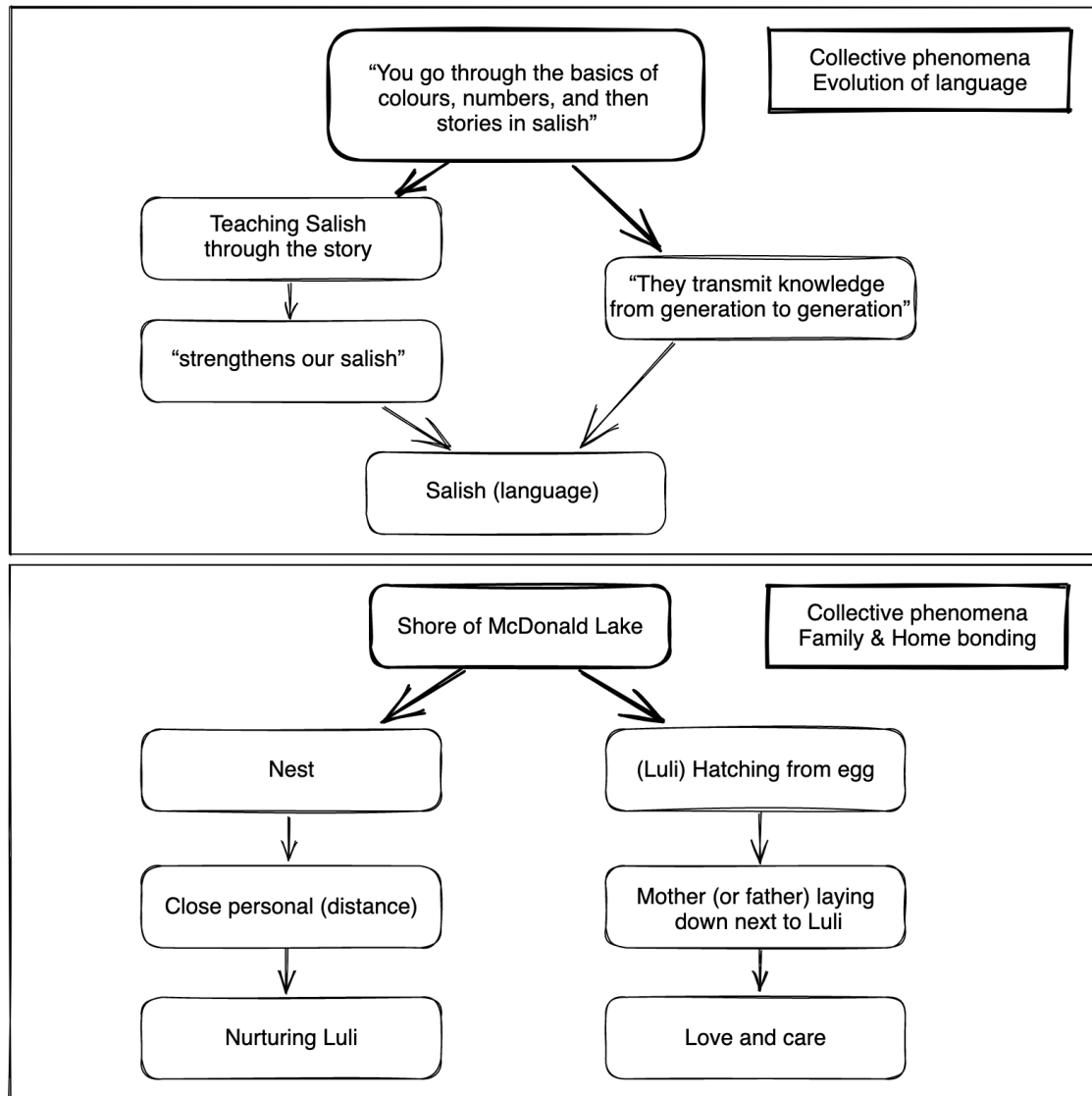


Figure 5.1 – P1 HASS' proposition support.

The concept we want to convey with P1 Science, Medicine, and Everyday Life is more related to the biological and physical aspects of the Geese, especially when flying in an emergent V or W-shaped formation. This does not appear to be a genetic trait. To put it another way, it is not embedded in the individual birds. Rather, it emerges—though not always—when they fly together. Literature suggests it reduces energy costs because the birds phase their wings in such a way that they exploit

the aerodynamic upwash from the bird in front of them (Muijres and Dickinson, 2014). Although the children's book does not focus on these aspects, possibly due to the inherent difficulty of grasping such concepts at such a young age, it is still an idea we would like to explore, even if only briefly, because we are targeting older audiences.

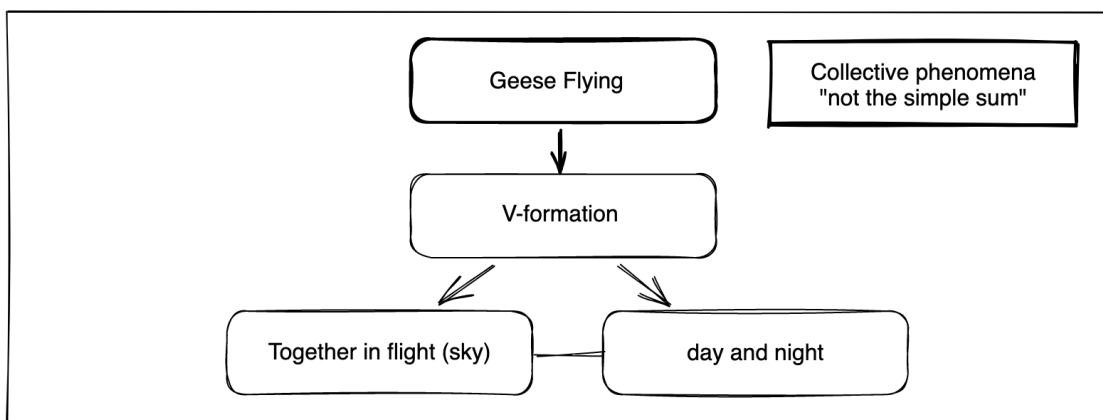


Figure 5.2 – P1 Science, Medicine, and everyday life's proposition support.

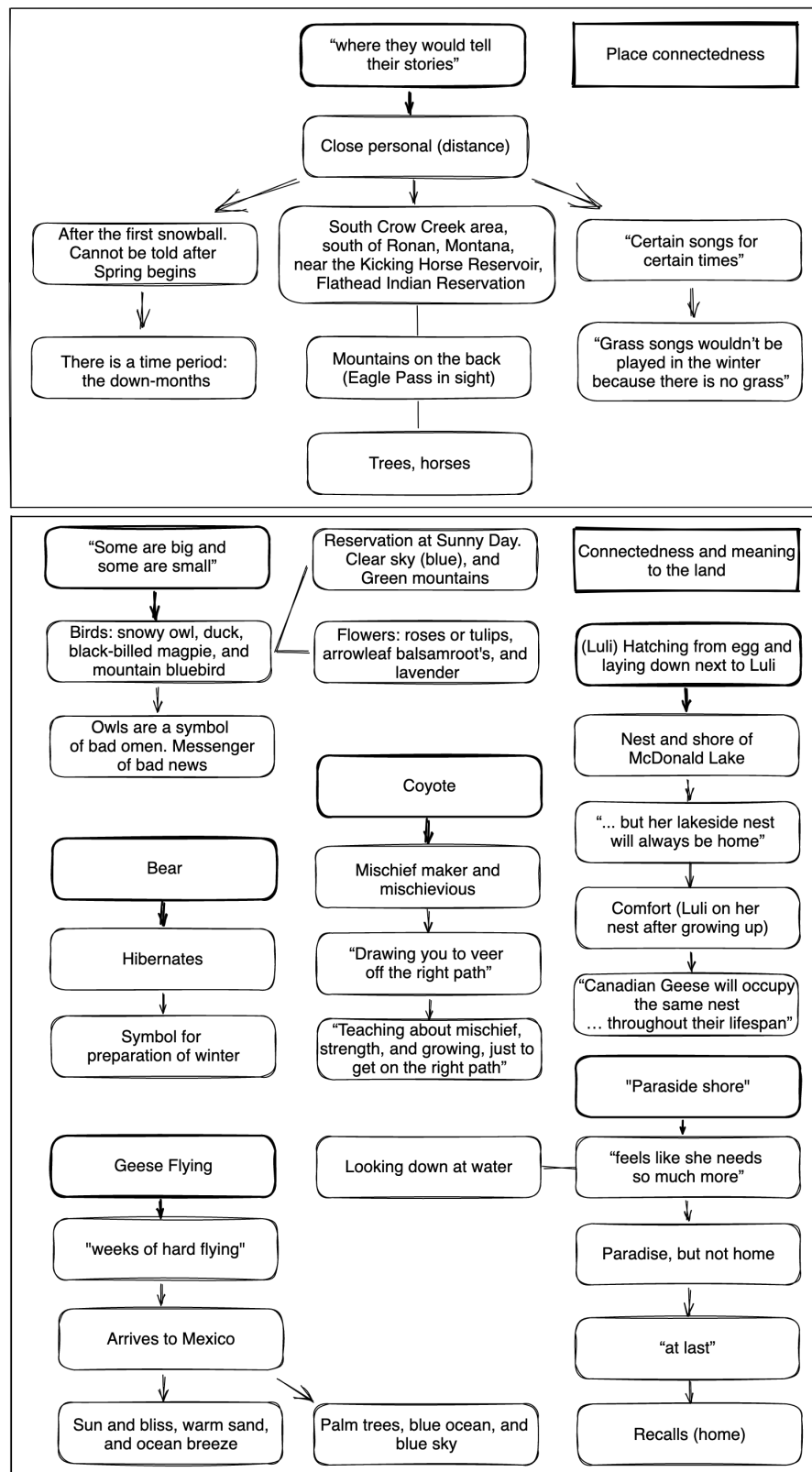


Figure 5.3 – P2 HASS' proposition support.

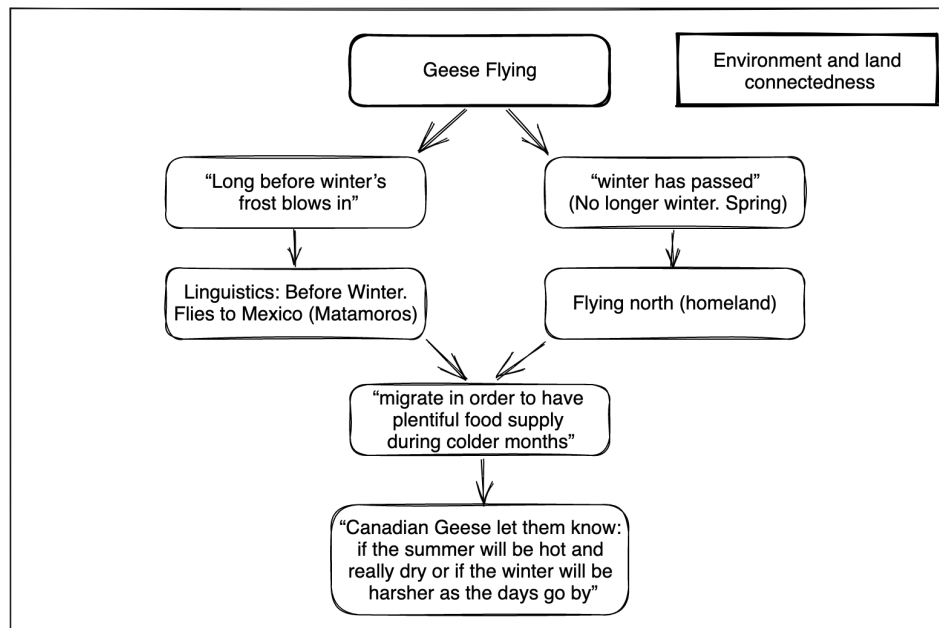


Figure 5.4 – P2 Science, Medicine, and everyday life's proposition support.

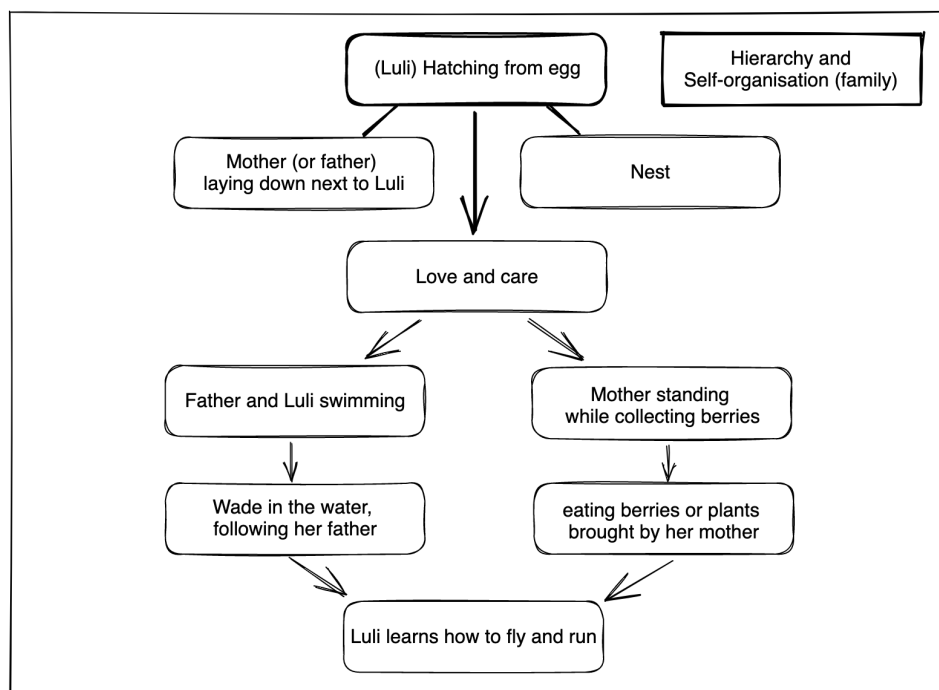


Figure 5.5 – P3 HASS' proposition support.

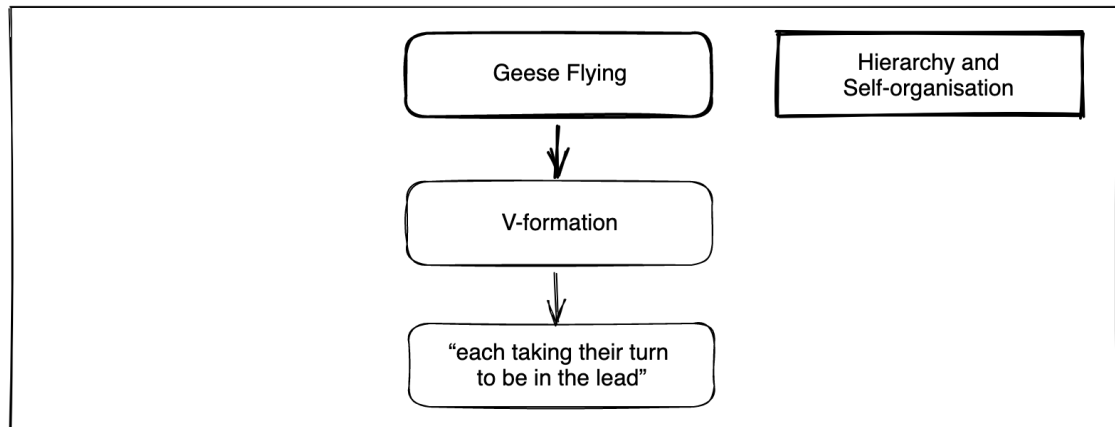


Figure 5.6 – P3 Science, Medicine, and everyday life's proposition support.

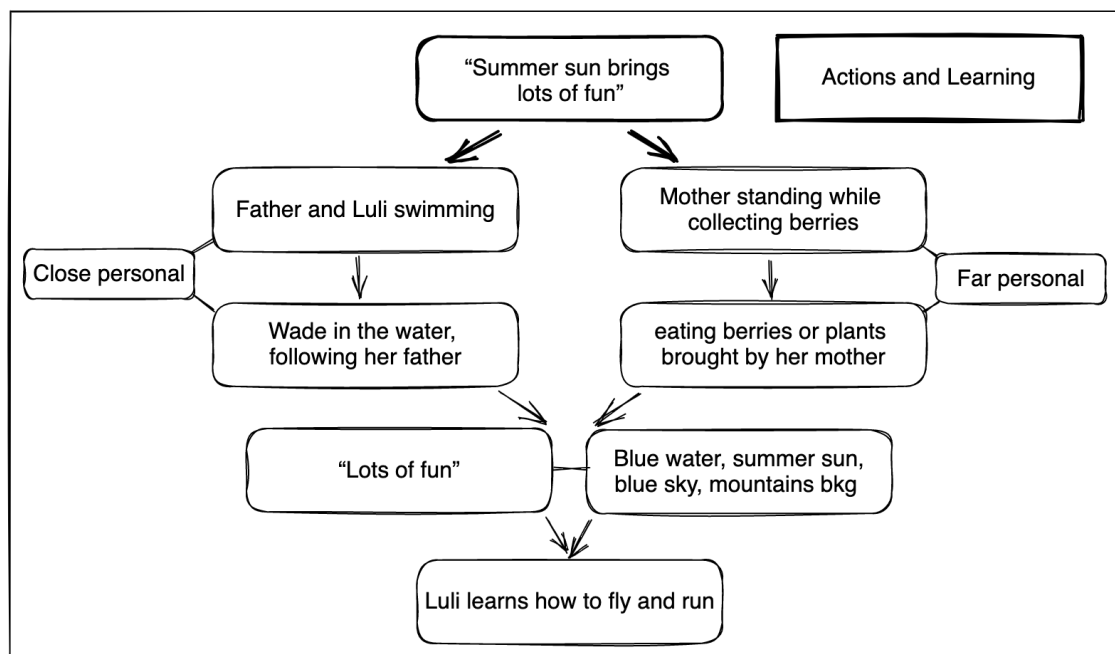


Figure 5.7 – P4 HASS' part 1/4 of proposition support.

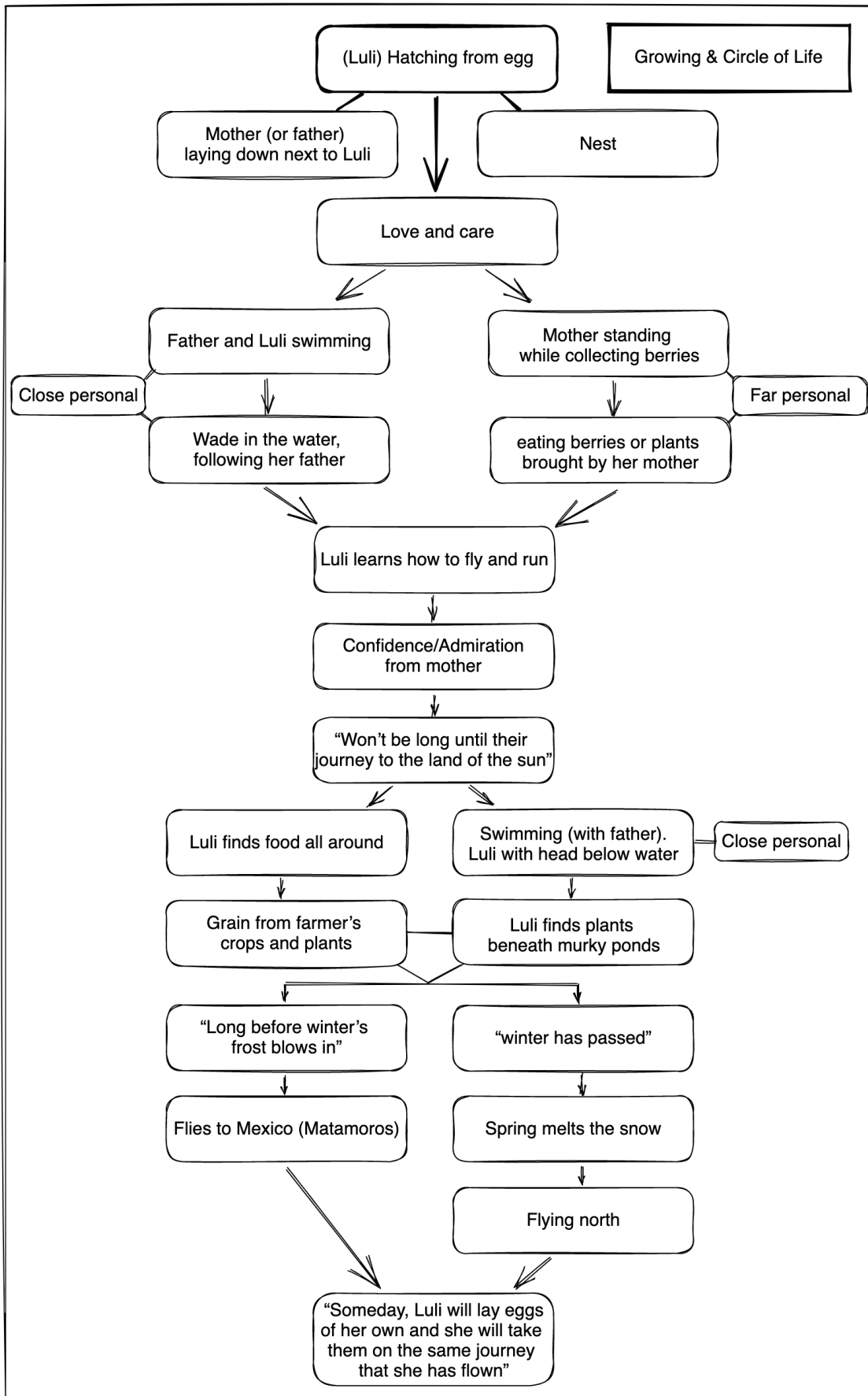


Figure 5.8 – P4 HASS' part 2/4 of proposition support.

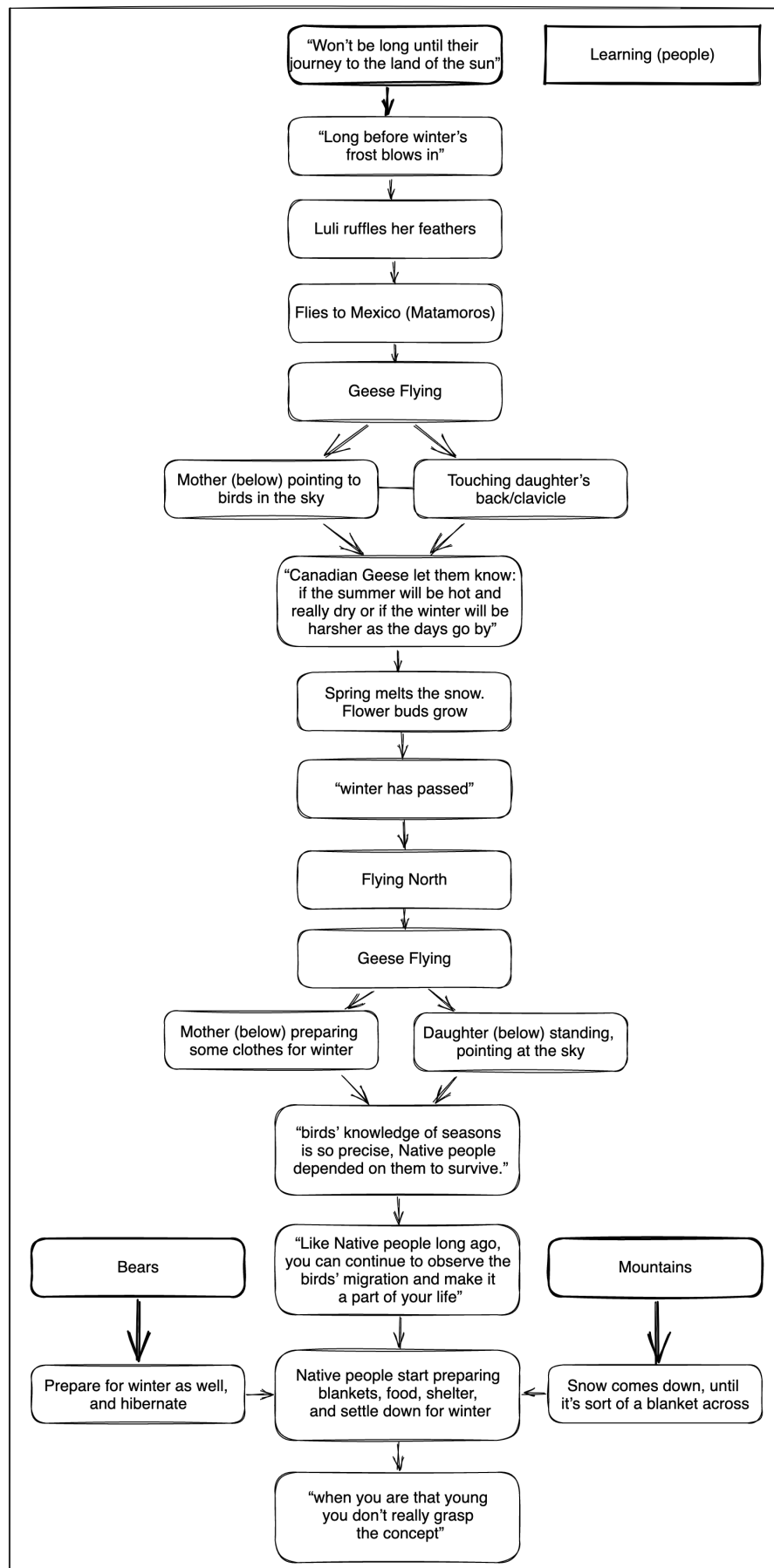


Figure 5.9 – P4 HASS' part 3/4 of proposition support.

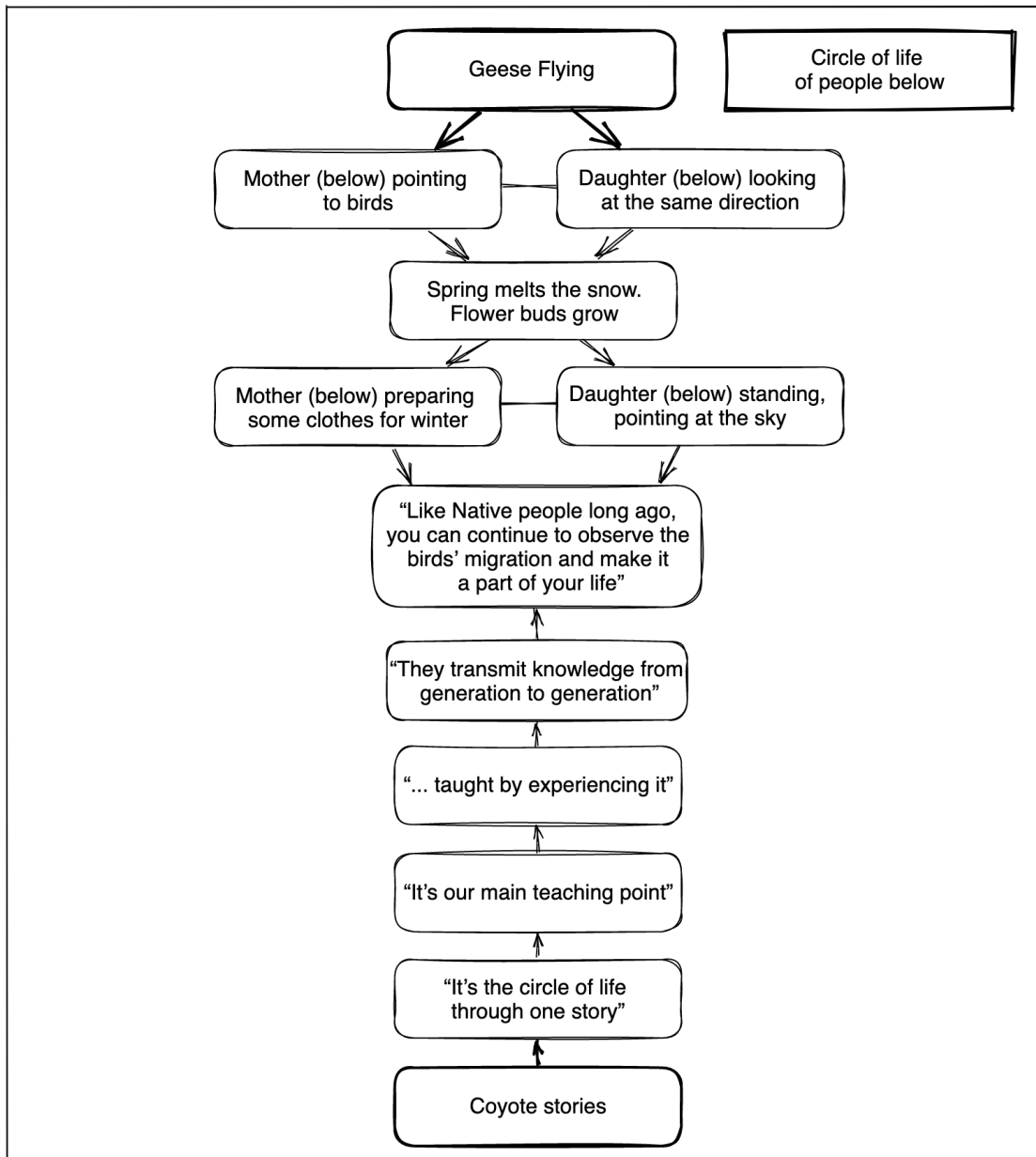


Figure 5.10 – P4 HASS' part 4/4 of proposition support.

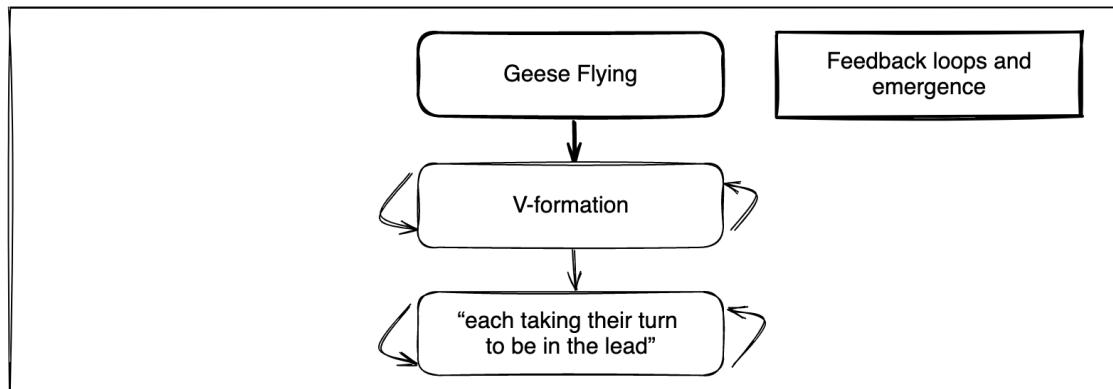


Figure 5.11 – P4 Science, Medicine, and everyday life's proposition support.

5.3.6 Conclusions

With evidence supporting each of the enumerated propositions, we can now attempt to answer the initial research questions, namely, what type of complex concepts and/or systems are present in the story, for both the Humanities, Arts, and Social Sciences (HASS), **RQ1**, and Sciences, Medicine, and Everyday Life, **RQ2**, clusters.

For **RQ1**, or HASS, the following propositions were found:

- Salish, not only as the emergence and evolution of the language and how it is used (grammar and semantics), but also as a means of transmitting knowledge from generation to generation and describing the surrounding environment;
- Family and Home bonding
 - Place connectedness;
 - Connectedness and meaning to the land.
- Growing and Circle of Life, for both people and birds;
 - Learning through experience and feedback loops.

For **RQ2**, or Sciences, Medicine, and Everyday Life, the following propositions were found:

- Canadian Geese V-shaped formation, including its emergence through feedback loops, hierarchy, and self-organisation;
- Environment and land connectedness, in the sense that migrations and other environmental cues provide insight to those who observe.

6

Multimodal HCI Architecture Specification

6.1 Introduction

Some particularly relevant HCI challenges were mentioned and explained in Chapter 2 (*q.v.* Subsection 2.2.5 - [Multimodal Human-Computer Interaction](#)). One of these, on the foundation of our work, stems from the fact that switching between [HMDs](#) such as Microsoft HoloLens, Oculus Rift and Quest, HTC Vive, and Magic Leap requires learning entirely new sets of gestures using one's hands or joysticks. Despite the fact that gesturing is natural and innate to humans, this follows the trend of traditional gestures and interfaces.

The project on which this thesis is based (*q.v.* Section 6.2, and Figure 6.1) proposed a solution to this specific challenge in 2015, but the literature review (*q.v.* Section 2.7) still shows a general lack of a semiotic perspective focusing on [HCI](#) interfaces (*e.g.*, gestural), thus ignoring the cultural and social dimensions of context. This base project is presented (*q.v.* Section 6.2), which, given its known and documented shortcomings ([Morgado et al., 2015a](#)), as well as new implications and platforms, has demonstrated the need to update its requirements and, ultimately, design (*q.v.* Section 6.4) and implement a new (*q.v.* Chapter 7 - [InMERSE Architecture Implementation](#)), improved version.



Figure 6.1 – InMERSE logo and involved institutions

6.2 Travelogue

The concept described and proposed in the InMERSE¹ project is based on what is known as the shamanic interface, or, more accurately, the anti-shamanic interface. The shamanic interface concept, originally introduced by Daniel Suarez in his Daemon book series (Suarez, 2009, 2010), was to create a universal interface that worked the same way for everyone. The anti-shamanic interface, on the other hand, proposes a different interface for each culture, taking advantage of this specific dimension of context. This anti shamanic interface is what our research group now refers to when we use the term “shamanic interface” (Morgado, 2014; Morgado et al., 2015b).

From then on, several works (Carvalho, 2019) were built on that shamanic interface concept, but InMERSE was put to the test with two separate entertaining applications made for Altice Labs (previously PT Inovação e Sistemas), namely, Digital Signage (*q.v.* Figure 6.2), developed with Gibbo2D², and First Army of India (*q.v.*

¹ <https://bitbucket.org/Apidcloud/inmerse-framework>

² <https://github.com/joafalves/Gibbo2D>

Figures 6.3, 6.4, and 6.5), developed with Unity3D³, both of which supported the use of multiple interaction modalities and devices in a more abstracted manner. As detailed further in its article (Morgado et al., 2015a), the usage and efficiency of the proposed architecture still needed to be thoroughly tested, as the results were too preliminary—only Digital Signage had been developed up to that point—but there were still open issues and topics for reflection that required additional work.



Figure 6.2 – Digital Signage screenshot showing the reproduction of a swipe right gesture.

Not long after, however, we developed and released First Army of India (Alves Fernandes et al., 2016), which brought two users into the same immersive environment via different modalities, namely, Player 1 via Oculus DK2 HMD (6DoF) and Player 2 via Google Glass. Each player could interact with the environment in different ways. The former by using an attached leap motion to track and detect the hands, and the latter by combining the use of Google Glass (for guidance through a virtual compass depicting the location of rocks to grab and the location of the ship, although the concept also included the use of a Myo Gesture bracelet, *q.v.* Figure 6.3) and a Microsoft Kinect 2 for full body tracking, allowing some movements (commands) such as throwing rocks to the ship. Although a large number of tests (Alves Fernandes et al., 2016) have revealed learning potential by facilitating entertaining, authentic,

³ <https://unity3d.com>

situated, and self-oriented experiences, the relatively high costs associated with the aforementioned technologies, as well as their limitations, suggested that adoption of such prototypes was not yet feasible or sustainable.



Figure 6.3 – Original concept for the First Army of India with Player 1, the helmsman, on the left, and Player 2, Adamastor, on the right.

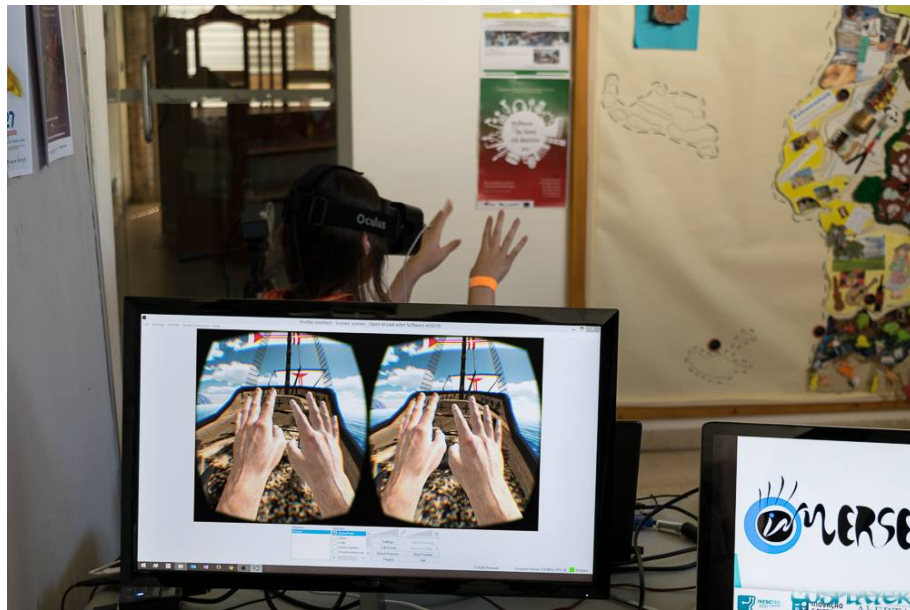


Figure 6.4 – First Army experience. Photo by Quinto Canal at the Manga Comic Event (MCE) 2015, which took place on 22nd and 23rd of August of 2015, in Faro, Portugal.



Figure 6.5 – High quality render of First Army of India's environment.

Years have passed, and immersive environments (*i.e.*, **VR**, **AR**, **MR**) have progressed further, almost to the point of democratisation—which does not appear to be far off,

given recent developments, such as Oculus Quest 2. The device, like its predecessor, Oculus Quest, runs on Android, so it is not dissimilar to a pure smartphone. More importantly, it is wireless ready. Furthermore, it can run—or rather, stream—anything that runs on a desktop, including specific desktop VR titles. At the time of writing, this was possible with Oculus Quest via Oculus Link, a USB-C cable that streams data from the desktop to the HMD. Oculus Quest 2, on the other hand, has already introduced Air Link, which does essentially the same thing as Oculus Link but wirelessly. This new version also runs at 90Hz rather than 60Hz, which greatly improves the experience. All of this is to say that developments in the field are ongoing and accelerating. It is truly approaching a point of democratization. So, what's the problem? While HMDs continue to improve, the HCI appears to be lagging behind. Haptics, as mentioned in the literature review (*q.v.* Section 2.2.5), are one of the issues. While hand gesture tracking is already possible without the use of a third-party device when using Oculus Quest or Oculus Quest 2, the interactions still have a long way to go. Attempts have been made to standardise and abstract some of the devices (*e.g.*, Microsoft MRTK, OpenXR), but the mechanics and feedback provided to the user are still lacking. UltraLeap was formed by the merger of Leap Motion and Ultra Haptics in an attempt to combine hand interaction with mid-air haptics. While this is true, the equipment they provide is still somewhat pricey, so it will take some time to truly democratise.

6.3 Requirements Analysis

The InMERSE Architecture has proven to be a viable method for developing gestural input applications by separating concerns for gesture acquisition, gesture recognition, and application commands. As stated throughout its paper (Morgado et al., 2015a), there were several open issues and topics for reflection that needed to be addressed further. For example, we did not address or achieve adequate abstraction for multiple input data formats and diverse semantic content at the time of its conception (*e.g.*, different skeleton assumptions).

One of the objectives we set for this doctoral program was to investigate the majority

of these issues and improve the InMERSE Architecture and, as a result, its implementation. This should allow for the creation of immersive environments (and other applications) that can quickly switch from one input culture to another, without the disadvantage of being completely reliant on each individual device framework to detect gestures.

Given the current goals (*q.v.* Section 1.2), and requirements for the architecture's future, the most obvious to investigate is the incorporation of machine learning in the Decoders. However, doing so has implications for the flow of and the actual Architecture, which is why it is critical to first redefine its requirements.

The Gesture Decoders must meet the following requirements in order to work with machine learning:

- The underlying input data (basic data) must remain consistent, or transformed in such a way through device-specific **Adapters**
 - This requires **ontologies** that are comprehensive enough to detect gestures, such as Hands and Body;
 - This requires not only the use of the same coordinate system, but also the data to be normalised.
- The **Decoders** should be able to handle the transformation and subsequent changes to the provided basic hand and/or body **ontologies** internally;
- The **Decoders** should be able to train their underlying a machine learning model
 - This requires another **Framework** internal state: *learning mode*, which should use the same data to train and then detect.
- Multiple **Decoders** should be supported, so that different people or different teams can focus on training a subset of gestures, which can then work in the same application
 - This requires an upstream configuration flow from the **Decoders** containing the gestures they can handle, as predicted back in 2015;

- This requires the maintenance of a centralised and combined gesture database containing unique entries within the **Framework**.
- Support for newer devices and their underlying **Software Development Kit (SDK)** input detection capabilities should still be possible (if any).
 - This requires the **Framework** continue to be able to listen to the **Adapters** via callbacks and notify the main application of their detection.
- Mapping gestures to commands should still be possible via an external configuration file.

Although this is unrelated to machine learning integration, another requirement is to abstract away the use of different adapters and decoders without requiring code changes (or even configuration files):

- To truly decouple the **Decoders** and **Adapters** from the **Framework** and avoid code changes, they should be loaded dynamically at runtime
 - This requires checking the class hierarchy at runtime to determine whether the modules are valid or not.

The architecture's outward goals, however, remain the same. We aim to minimise gesture-controlled applications' early obsolescence and make them more accessible to users with mobility impairments and from diverse cultural backgrounds. To leverage complex learning experiences, this thesis expands on these ideas, including a more robust, adaptable, and inclusive user experience.

In accordance with the lessons learned during the InMERSE project and even CE4Blind (Alves Fernandes, 2016), the objectives of this architecture converge to some enterprise requirements (*q.v.* Section 3.3.3) and should be adapted, or at least support, mobile platforms (not necessarily smartphones, as some HMDs run Android), which brings implications (*q.v.* Section 3.3.4.1) in accordance with architecture design standards (*q.v.* Section 3.2). Following that, the following enterprise requirements are considered:

- **High scalability**, which can be broken down into two aspects
 - **Stable and Robust Architecture**, allowing for the easy integration of emerging technologies;
 - **Separation of Concerns and Reuse of Code**, that is, designing an architecture with a cross-platform scenario and clear separation of concerns in mind: a solution that can work on different platforms.
- **Redundancy, High Availability and Fault Tolerance**, a term not commonly used in an enterprise environment, refers to the use of multiple servers to maintain overall system functionality while handling the response from each input device in a non-dependent manner. This increased fidelity can be achieved, for example, by employing multiple modalities, which, despite having different inputs, can be mapped onto the same command and continue the experience;
- **Performance**, which is required to process all sensor data in a timely manner;
- **Load Balancing**, which should allow for more efficient use of available resources (hardware), is regarded as critical, particularly in a mobile platform where battery life is an important consideration.

In this regard, and in light of the implications of targeting mobile platforms, the following factors should be considered:

- **Battery Life and Temperature**, which are primarily determined by screen time as well as the required hardware and processing;
- **Processing**, which is heavily influenced by the algorithms and hardware used;
- **Response Time**, because information is only relevant and reliable if it is obtained in a timely manner, and this can have a significant impact on the user experience (*i.e.*, if the feedback from the input of a user has noticeable delay).

While machine learning algorithms are becoming lighter, given their many applications in the mobile field (*e.g.*, [AR](#)), they can still severely limit battery life and

overheat the device, and should thus be considered. All of these requirements were met through iterative development of an improved version of the InMERSE framework.

6.4 Architecture Design

Despite the fact that we defined new requirements (*q.v.*, Section 6.3), they have no effect on the overall concept of the architecture, resulting in the same overview of the InMERSE Framework (*q.v.* Figure 6.6).

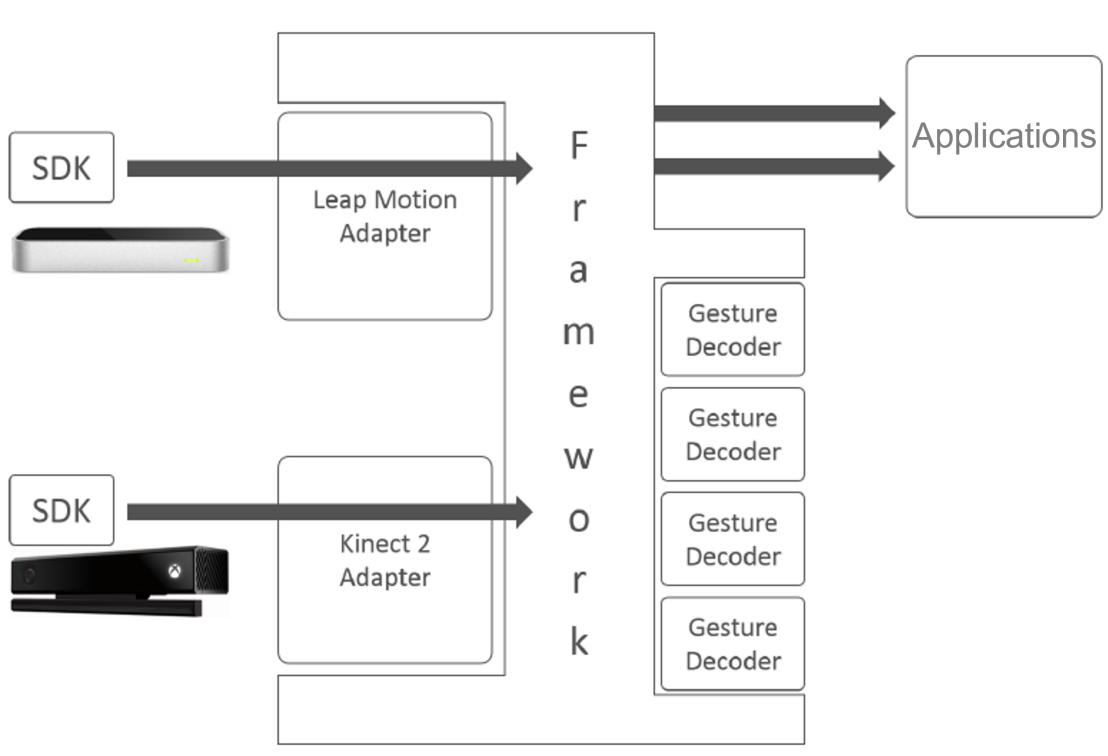


Figure 6.6 – Overview of the InMERSE Framework.

With the new set of requirements, as well as the work already done in the state of the art (*q.v.*, Chapter 3 - [Software Architectures](#)) and our own research group (*q.v.* Section 6.2), a more updated architecture has been established (*q.v.* Figure 6.7), with specific changes not only to the flow of data, but also to the specificities of each

module and how they interrelate.

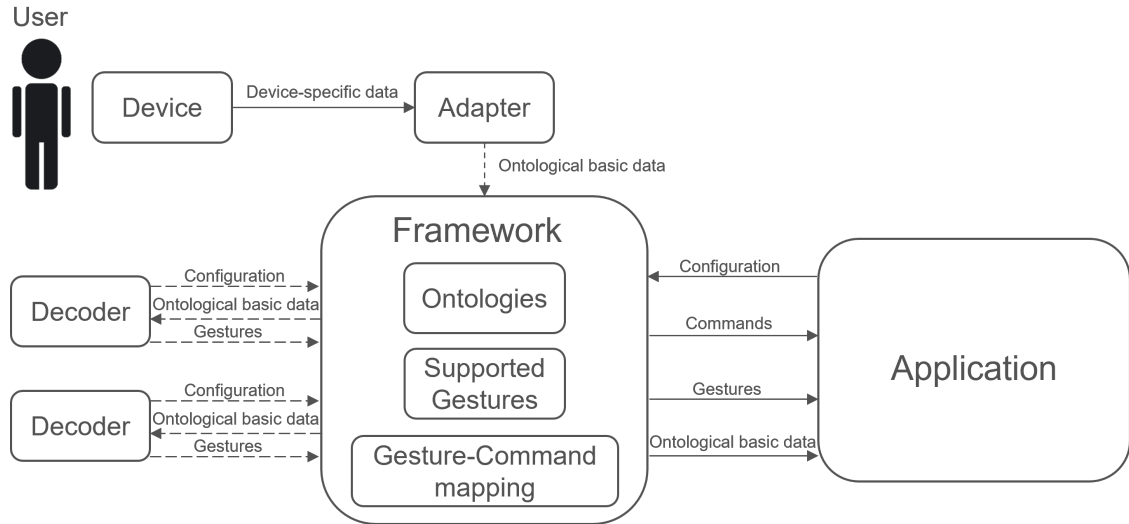


Figure 6.7 – Improved InMERSE Architecture.

The overall architecture style can be compared to a Blackboard (*q.v.* Section 3.3.2.2), with a star topology in which a central module (Framework) maintains a general state of the system and updates it as information from the surrounding specialisation modules is received. Although the main module, Framework, maintains a general, central state, the broadcasting and flow of information can be described as event-driven (*q.v.* Section 3.3.2.3), with a mix of publish-subscribe (*i.e.*, subscribing only to specific commands or gestures, *q.v.* Subsection 3.3.2.3) and an event bus (with the bus being the central module, as it abstracts away where the data comes from, *q.v.* Subsection 3.3.2.3).

This architecture allows consumer applications to access three types of data (*q.v.* Figure 6.7):

- **Commands**, which are gesture-independent information. The Framework module will map a gesture identified by a Decoder module to a command (based on an external configuration file), removing the need for the application to know what actually triggered said command (*e.g.*, if “Activate” was issued by placing both hands in the environment, pushing a virtual button with a single finger, or even looking into an object for a specific duration);

- **Gestures**, which are simply transparent access to the output of Decoders if an Application requires this level of detail at some point (for instance, for tutoring purposes);
- **Ontologic basic data**, which is the information required to represent a specific ontology, transformed to by the Adapters given device-specific data. For example, this could include an ontology for a hand and another for a body skeleton, which are then used for gesture training and detection using Decoders. Applications may require this, for example, to display continuous motion of the user's hand rather than just anticipating specific commands (*e.g.*, grab).

We have the following individual modules and their responsibilities:

- **Application** (Consumer), which begins with a configuration file that maps gestures into commands that can then be subscribed to. Access to gestures and ontological basic data is still available, which, in addition to being useful for debugging, can also be used for other purposes, such as the acquisition mode;
- **Adapter**, transforms the device-specific gestural data into the ontologies provided by the Framework. This should already account for the transformation of different coordinate systems into the framework's expected one (x , y , z , with y being the height, and z being depth);
- **Decoder**, starting with a configuration file containing supported gestures and retrieving ontological basic information only after that. This can then be transformed and normalised as needed to meet the needs of machine learning (for both training and detection). Only when the Framework is in detection mode does detection occur. When in training mode, the Decoder handles the necessary information (*e.g.*, gesture ID, subject ID, attempt number), as well as the time required to train each model entry (*i.e.*, recording time). When it finishes recording a specific entry, it saves/appendes the entry to a file and triggers a specific known and reserved event that the consumer application

doing the training can subscribe to. This allows for a separation of concerns while still allowing the final application to decide what to do when recording is complete (*e.g.*, closing);

- **Framework**, the architecture's core, provides Ontologies (and their underlying data) to all other modules and orchestrates the entire acquisition, detection, and learning process. It accomplishes this by maintaining an internal state. The Framework module is independent of all other modules, including the Adapters and Decoders (*q.v.* dashed lines in Figure 6.7). They are loaded dynamically in runtime while still ensuring their validity. This ensures that both Adapters and Decoders can be switched at any time without any additional configuration.



InMERSE Architecture Implementation

7.1 Introduction

Based on the requirements and features described in Chapter 6 ([Multimodal HCI Architecture Specification](#)), a proof of concept library was created, developing the necessary modules (*q.v.* Section 6.4) to ensure that the presented architecture, and thus the consumer solutions, functioned as intended. Because the implementation is a library, that therefore must be consumed, several examples were created, as described further below (*q.v.* Sections 7.2 and 7.3).

Finally, some of the techniques used in this implementation are mentioned and explained (*q.v.* Section 7.4), with the primary goal of ensuring higher performance, loosely coupled modules, and conceptual layers (*e.g.*, business layer) that comprise them.

7.2 Inception

The technologies and platforms that would be used to implement the architecture were one of the first decisions to be made, knowing that one of the possibilities was to enable a multiplatform solution (*i.e.*, supporting not only desktop-based applications, including **VR**, but also the most popular mobile operating systems such as iOS and Android, or even the Web).

The first version of the InMERSE¹ architecture was written in C# and used the .NET Framework². However, despite the existence and compatibility of some more general parts, which could be used when targeting iOS and Android through the use of Mono³, an open-source implementation of .NET, and the Xamarin⁴ platform, the .NET Framework was still pretty much platform specific (*i.e.*, Windows) at the time.

An implementation of the InMERSE architecture for the Web was developed in 2016-2017, already during this doctoral program. The original plan was to integrate Multimodal **HCI** into Scarlett Game Studio⁵, a 2D videogame engine and editor that runs on top of web technologies. We did, however, decide to create a more generic library that could be used freely on top of browsers, which demanded some exploratory work and eventually led to the development of a first JavaScript library prototype (*q.v.* Section 7.2.1).

Since then, and the original InMERSE implementation, the .NET Framework has branched out, namely through .NET Core, Microsoft's attempt to target other platforms while bringing some code reuse. It did, however, remain incompatible with the .NET Framework and Mono, due to the fact that they all use different **Base Class Libraries (BCLs)**. To address this issue, Microsoft merged the **BCLs** from the .NET Framework, .NET Core, and Mono into a single **BCL** known as .NET Standard. The lower its version number (1.0), the greater its compatibility. The higher it

¹ <https://bitbucket.org/Apidcloud/inmerse-framework>

² <https://microsoft.com/net>

³ <http://mono-project.com/>

⁴ <https://xamarin.com/>

⁵ <https://github.com/scarlettgamestudio>

goes, the less compatible it becomes, but the more specific its APIs become, allowing access to lower system levels.

We decided to focus on desktop platforms for the actual improved version of the InMERSE Architecture, as well as some mobile platforms, particularly Android. The latter was a decision based on the current trend of HMDs running on top of this operating system. This also enables it to be used for AR development, whether via an HMD like Vuzix⁶ or a smartphone. The library was then implemented in C# and .NET Standard, allowing it to be used on a variety of platforms, including desktop-like applications like Windows Presentation Foundation (WPF) and even Unity3D⁷—from which multiplatform is possible, albeit with more platform-specific code.

7.2.1 JavaScript Prototype

Taking into account the original implementation completed as part of the project with Altice Labs (previously PT Inovação e Sistemas), we began to consider developing the same architecture using JavaScript instead. The main reason for this choice is the constant evolution of web technologies, which now target not only traditional browsers, but also cross-platform environments via the combination of NodeJS and Electron. There are several options for JavaScript module loading, bundling, and transpiling, and it is critical to consider them before proceeding with the library. As a result, we concentrated on investigating these technologies in order to make more informed decisions and, as a result, make the library available to other developers.

7.2.1.1 Transpiling and Bundling

There are numerous options for transpiling and bundling JavaScript, owing to the various targets (*i.e.*, browsers) and supported versions. Keeping this in mind, two (boilerplate) projects were created to test the various technologies and possibilities

⁶ <https://www.vuzix.com/Products>

⁷ <https://unity3d.com/>

available in the web ecosystem (*q.v.* [detailed description below](#)). The first⁸ employs Typescript, a JavaScript superset, whereas the second⁹ employs ECMAScript 6, also known as ES6 or ES2015. Browsers cannot yet support all of ES6's features (*cf.*, [compatibility table](#)), let alone Typescript. There is an essential step to make it work/compatible with the vast majority of browsers on the market: transpiling to ES5 because it is fully supported. Transpiling is also a method of using tomorrow's JavaScript today. Both library project setups contemplate the use and combination of various web technologies and packages, namely:

- [Typescript](#), a JavaScript superset, as the language for both library **implementation** and **unit tests (only used on the first setup)**;
- [Babel](#) for transpiling ES6 (ES2015) to ES5 using [babel-preset-es2015](#), and [babel-transform-runtime](#) and [babel-stage-3](#) for async/await support;
- [Webpack](#) for bundling everything into a single file with the help of [ts-loader](#), [babel-loader](#) and [cross-env](#);
- [Webpack-dev-server](#) for running a live hot reloading server;
- [Jest](#) for testing in typescript using [ts-jest](#) and [@types/jest](#);
- [Travis-ci](#) and [appveyor-ci](#) for Continuous Integration on unix and windows systems respectively;
- [Codecov](#) for code coverage;
- [Semantic-release](#) to handle publishing on GitHub and on Node Package Manager (NPM);
- [Commitizen](#) together with [cz-conventional-changelog](#) to help with semantic commits;
- [Husky](#) to handle pre-commits and commit-msg hooks;
- [Validate-commit-msg](#) to validate commit messages.

⁸ <https://github.com/Apidcloud/TypeScript-Experiments>

⁹ <https://github.com/Apidcloud/webpack-babel>

Bundling can be interpreted as a technique of grouping the various files that comprise a library into a larger one, making only the essential functions available (*i.e.*, public *vs.* private scope). There are two popular formats: CommonJS, which is used with NodeJS and allows client applications to require objects exported from the library, and Asynchronous Module Definition (AMD), which can be used to load modules asynchronously. There is a third option for projects that must support both formats: Universal Module Definition (UMD). It enables the library to use third-party modules that target any of the module formats, as well as client applications to use the library in their preferred format. ES6 combines the best of both worlds, including compact and declarative syntax and asynchronous loading, as well as additional benefits such as improved support for cyclic dependencies. These formats, however, are not natively supported by browsers, necessitating the use of a bundler such as [Webpack](#) or [Browserify](#), which, in addition to grouping the files, also resolves the dependencies.

7.2.1.2 Web library

The InMERSE library is currently exported in the Universal Module Definition (UMD) format and can be used with AMD, CommonJS (*e.g.*, NodeJS), or through a global variable (*e.g.*, browser). For the time being, it allows client applications to listen to commands and frames, despite the fact that it is still partially tied to the Leap Motion, something we'd like to avoid entirely. A possible setup for using it in a NodeJS environment is depicted in [Figure 7.1](#), below.

```

var inmerse = require('inmerse');

inmerse.configurationPath = 'translation.xml';

inmerse.on('ready', () => {
  → this.on('navigationLeft', () => {
  →   → console.log('Navigation Left command has been evoked');
  →   → });
  → this.on('frame', (frame) => {
  →   → console.log(frame.hands.length);
  →   → });
});

inmerse.init();

```

Figure 7.1 – InMERSE library setup example.

After registering the desired events (*e.g.*, when the navigate left command is triggered, *q.v.* Figure 7.1), the process can be initiated by calling the function `inmerse.init`.

7.2.1.3 Interaction Demos

Throughout this doctoral programme, some forms of interaction have been explored and developed, primarily through the use of the Leap Motion device with some well-known online services. With these examples, it is also possible to extrapolate some commands (underlines below) that, when combined with the architecture, can make them gesture and device-independent, which is the entire point of the InMERSE architecture:

- **Prezi**¹⁰ (Leaprezi¹¹): a left or right swipe can be used to navigate to the next or previous slide, respectively;
- **Reel**¹² (LeapReel¹³): the image can be dragged along the horizontal axis by

¹⁰ <https://prezi.com/>

¹¹ <https://github.com/Apidcloud/Leaprezi>

¹² <http://reel360.org/reel>

¹³ <https://github.com/Apidcloud/leapreel>

moving the hand to the right or left. Moving it faster produces a throw effect;

- **Google Street (LeapStreet):** the view heading is controlled by the right hand, while the pitch is controlled by the left. If the movement is fast enough with either hand, the direction is blocked, allowing the same movement to be repeated and thus full 360 degrees rotations.

Some of these forms of interaction began to show that hand-controlling a 3D object could be difficult. While pinching (joining the thumb and index fingers) appears to be the standard nowadays for a grasping interaction, which then blocks the objects and allows almost full control of it, it lacks some cultural-awareness and intuitiveness, particularly with larger objects that naturally, and physically, require both hands. Consider picking up a football with both hands as an exercise. How would you rotate it? Because of its peculiarities, this question was on my mind for some time. One appears to rotate not with both hands at once, but one at a time, gradually. This resulted in the development of two prototypes displaying exactly this behaviour, which, unsurprisingly, feels very intuitive. A first prototype for the Web, dubbed **Leap Motion Earthquake Visualisation** (*q.v.* Figure 7.2), allows two-handed control of a globe within a web application. Architecturally, it is accomplished through the acquisition-mode, in which the consumer application requires direct and real-time access to the input device's basic data.

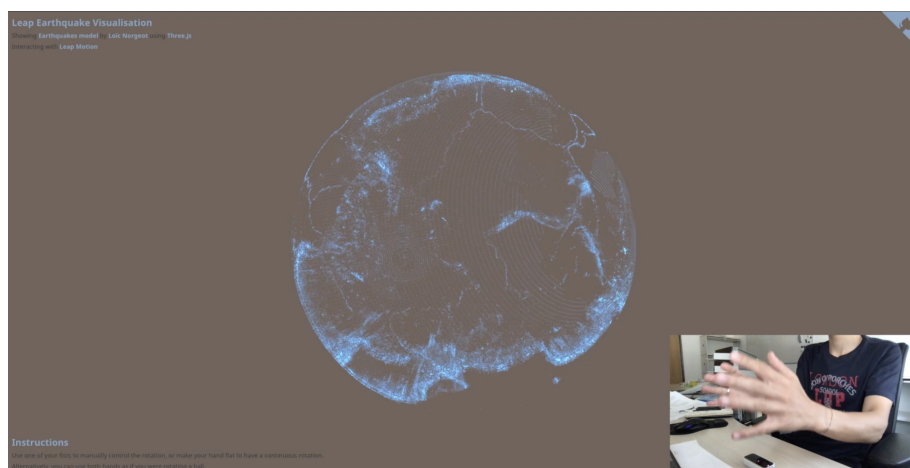


Figure 7.2 – Leap Motion Earthquake Visualisation showcasing a two-handed interaction with a globe within a web application.

Given our intention to focus on immersive environments, we also created a VR prototype of the same behaviour, showcasing a full two-hand rotational interaction system, dubbed *Leap Motion 360 Interaction* in VR (*q.v.* Figure 7.3).

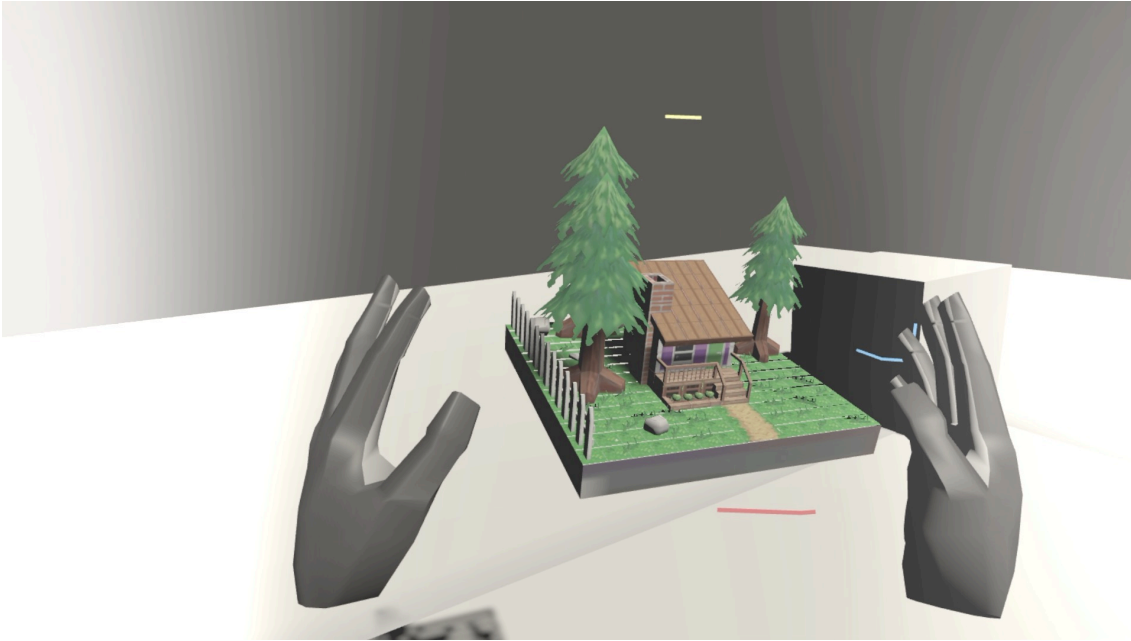


Figure 7.3 – Leap Motion 360 degrees Interaction showcasing a full two-hand rotational interaction system with an object in VR.

7.3 Design

The new architecture was designed and refined continuously to improve the **separation of concerns** and, as a result, to meet the requirements and objectives outlined in the preceding sections (*q.v.* Section 6.3). The development was done on top of the previous version¹⁴ to make this process more feasible and acute. After all, as previously stated, the idea was to expand it by incorporating machine learning in the Decoders, and the implications of doing so on the overall Architecture are not insurmountable, proving that its original, and sufficient, separation of concerns was effectively a way of supporting further updates.

¹⁴ <https://bitbucket.org/Apidcloud/innerse-framework/>

A first proof of concept for a WPF application was implemented and is also included in the repository as a base-case. This was enough to put it through its paces with another desktop application, Digital Signage, which already made use of the original version of the InMERSE framework. As previously stated, Digital Signage was created with Gibbo2D¹⁵ and is designed for use with the Windows operating system. Regardless, it was a good scenario to essentially update the consumer application while still ensuring that the new version could run it with good performance. It was more a case of updating the dependency as well as its use and configuration. Although our goal was to actively target immersive environments, such as VR, this was the first fully functional test of the new version of the framework.

7.3.1 Adapter Module

The Adapter module, as previously described (*q.v.* Section 6.4), is in charge of converting device-specific data (*e.g.*, gestural), from an input device, into data that the Framework module and others can understand. This is accomplished by extending one of the Framework module's abstract classes (*e.g.*, **OntologicalHandsFrame**, **OntologicalBodyFrame**) and implementing the necessary abstract methods (*e.g.*, **GetOntologicalHands**, **GetOntologicalBody**). This extension and implementation not only transforms data from a specific input (*e.g.*, **gestural**, **somatic**) into data that the Framework expects, but also in the form that the Framework expects. There are multiple 3D coordinate systems, including one that considers the *y*-axis to be the height and the *z*-axis to be the depth (*i.e.*, right-hand coordinate system), or vice versa (*i.e.*, left-hand coordinate system). Because the framework uses the former, right-hand coordinate system, it is necessary to swap axes if the input device uses a different one. While ignoring this may still work with a specific Decoder, especially when using machine learning, it may cause some unwanted behaviour and bugs to emerge with others. Some Decoders may further normalise or transform the data, and not always in the same way for all axes. To ensure that an adapter, and thus the consumer application, continues to function regardless of the Decoder(s) in use, this right-hand 3D coordinate system must be followed.

¹⁵ <https://github.com/joafalves/Gibbo2D>

If at all possible, the Adapter module should also implement some basic framework-specific gestures that do not require Decoders. This is useful for devices that already have some device-specific mechanism for doing so. A simple example is the gesture **Activate**, which can, for instance, be implemented as the detection of the presence of both hands. This can be accomplished by extending and implementing the abstract class **GestureListener**.

Finally, the actual connection to the Framework is established by extending and implementing the provided Controller abstract class and feeding all necessary information to the Framework. The Adapter module should accomplish a few things in this implementation:

- Establish a connection with the desired input device (peripheral or not);
 - Get information from the device continuously, in a loop or pool-based fashion;
 - Transform the data into one or more of the supported ontologies, such as **OntologicalHandsFrame**, through their extension (*e.g.*, **FrameAdapter**);
 - Fill the provided Framework’s **FrameBuffer** with the transformed and supported frames.
- Instantiate the implementation of the abstract class **GestureListener** (*e.g.*, **GestureAdapter**);
- Retrieve configuration files, including device-specific gestures (if any) and the gesture-command mapping that the **consumer application** should provide;
- Instantiate the provided Framework’s **CommandDetector** class, and feed it with all the information above (*e.g.*, **LeapMotionHandManager**).

Each **Adapter** will instantiate its own instance of the **Framework**, which, in turn, will support multiple **Decoders**. This may initially raise the pertinent question of whether it is preferable for the **consumer application** to instantiate a single instance of the **Framework** that, in turn, handles all the **Adapters** (one-to-many),

rather than instantiating multiple **Adapters** which handle their own instance of the Framework (many-to-one). To be clear, a single consumer application can support multiple Adapters and Decoders. The reason why, for the time being, the starting point is made through the **Adapter** has to do with the gesture-command mapping. It's a way of ensuring that multiple gesture-command mappings can still be created independently without having to deal with them all at once. If the need arises in the future to handle multiple devices within the same Framework instance, at the same time, we may consider implementing it in this manner. For the time being, however, it would make it unnecessarily complicated.

7.3.2 Decoder Module

The Decoder module, as previously described (*q.v.* Section 6.4), is in charge of, as the name implies, decoding data regardless of the original input. This is ensured by the use of **Ontologies**, minimal basic data sets used to represent a hand or skeleton. However, in order for this to work, the Decoder module must extend and implement the provided Decoder abstract class. It begins by supplying the Framework with a configuration file containing a list of the gestures that are supported. The module can then further transform and normalise the incoming data to meet the needs of the underlying machine learning model (or any other method of decoding). Although the current implementation uses the right-hand 3D coordinate system (*q.v.* Section 7.3.1), a decoder can rearrange it as needed. Rather than using a simple array with multiple values and possible interpretations, the incoming data is still labelled as x , y , and z . The Adapter is still in charge of swapping the input modality's axes to the right-hand 3D coordinate system, though they could be labelled right-left, up-down, back-front instead. One can argue for or against it, so for the time being, we've decided to stick with the axes' naming convention while still relying on the Adapter to do the swapping if necessary.

The Decoder relies on the current mode of the Framework to function, which is determined by the consumer application. When in *detection* or *recording mode*, it receives ontological frames that can be cast into more specific ontological frames

like **OntologicalHandsFrame** or **OntologicalBodyFrame**. When in *recording mode*, the Decoder module should take care of its own state. While this is true, the module should notify the consumer applications when the recording is complete. This can be accomplished by reserving a unique gesture or command name (*e.g.*, **RecordingEnded**), which the consumer application can then use to decide what to do next. The persistence of recorded data (*i.e.*, data entries) can be accomplished directly in this module by defining a recording duration and some logic that keeps track of what is being recorded (*e.g.*, **GestureID**) and by whom (*e.g.*, **SubjectID**). That will all be dependent on the underlying detection logic (*i.e.*, whether or not machine learning is used, the used features and their types, and so on), and it may even be delegated to a new module.

Because our current implementation is centred on machine learning models (*q.v.* Section 8.4.3), we prefer to keep all of this logic in the Decoder module. When in detection mode, the module should use whatever internal decoding logic it employs to notify the Framework of its results (*e.g.*, **Gestures**) so that they can be mapped into commands.

7.3.3 Framework Module

As previously described (*q.v.* Section 6.4), the Framework module, the architecture's core, is responsible for providing **Ontologies** and their underlying data to all other modules while orchestrating the entire *acquisition*, *detection*, and *recording* process. It achieves this by maintaining an internal state that includes not only the current mode but also the current gestures being detected. An Adapter instantiates the core, **CommandDetector**, and is responsible for feeding all the information it needs to function. This includes configuration files, as well as an extension and implementation of **GestureListener** and a **FrameBuffer** containing Ontological basic data via more specific types of frames (*e.g.*, **OntologicalHandsFrame**, **OntologicalBodyFrame**). The adapted frames from the incoming Adapter are broadcasted to Decoders and the consumer application. Decoders, and even Adapters, can then notify the Framework that a particular Gesture has been detected, which is then

mapped into commands if possible. **Commands**, as previously defined (*q.v.* Section 6.4), are a collection of one or more gestures and/or commands. A queuing technique is used to convert gesture poses into commands, which is constantly used to try and match a predefined command's gesture (or sub-commands) list. When the consumer application first launches, all commands, sub-gestures, and sub-commands provided by the Adapter, Decoders, and the default Framework's set are merged into a unique name list. Each gesture is then uniquely coded, with a two-digit hexadecimal code corresponding to it. The command they form is then the union of the hexadecimal codes of the sub-gestures—not their sum, which would not guarantee uniqueness.

This module is also responsible for ensuring decoupling from both Adapters and Decoders. This is achieved in a few ways:

- Decoders and Adapters have a distinctive naming convention (by default, it must end in the word **Decoder** through a glob pattern `"*Decoder.dll"`) that allows the Framework to find them. Although the Framework manages this, the pattern can still be changed by the consumer application;
- Decoders are dynamically loaded and instantiated in runtime within the Framework, while ensuring that the actual Decoder is valid. This is accomplished by determining whether the runtime assembly's exported types are subclasses of the Framework's abstract classes;
- Adapters are not a Framework dependency or directly used by it;
- The consumer application dynamically loads and instantiates adapters during runtime. This is done by the consumer application, which uses Framework helper methods, because this is where the application's flow begins, as described further in the Adapter module explanation (*q.v.* adapter module subsection). However, in order for this to happen, the Adapters must also adhere to a naming convention (by default, it must end in the word **Adapter** through a glob pattern `"*Adapter.dll"`) that allows them to be discovered. This pattern, however, is still configurable because it is ultimately managed by the consumer application.

The solution is also aimed at mobile platforms, where battery and temperature are important issues that are heavily influenced by the processing and resources used. These issues were considered when designing the architecture, allowing Adapters and Decoders to implement the **IConsumptionMode** interface (*q.v.* Figure 7.4 on page 194), enabling each module to act in accordance with the current system mode (*i.e.*, **Low**, **Balanced**, or **High Performance**), turning components on/off and adjusting processing as needed (*e.g.*, number of frames per second).

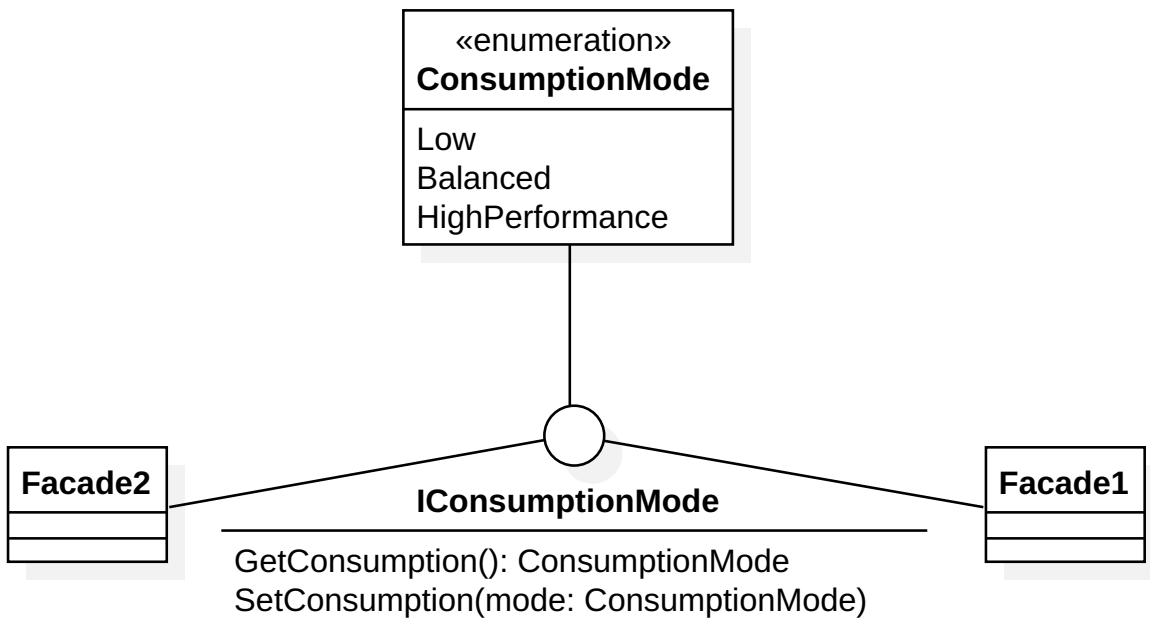


Figure 7.4 – Consumption Mode Class Diagram.

7.4 Culmination

Although the architecture specifies certain general rules that must be followed in its implementation, there are a number of techniques that can be used at a lower level. That is, while some of the classes and their behaviour are already somewhat defined in terms of their goals, the code itself can take many forms and be implemented in various ways. It is possible to draw an analogy with building construction, where the method of fixing, for example, a plasterboard plate in a wall or ceiling, according to

a work project (architecture), could be the use of nails, screws, cement, or even glue, leaving always the architect to choose the best suitable solution to each context, not forgetting that what matters to the customer in the end, ends up being a performant, long-term solution.

Among various techniques, this section comes down to describe **concurrency** (*q.v.* Subsection 7.4.1), **communication** (*q.v.* Subsection 7.4.2) and **design patterns** (*q.v.* Subsection 7.4.3).

7.4.1 Asynchronous Programming

There are several types of **Concurrency** (Cleary, 2014), such as Multithreading, which uses multiple threads of execution, and Parallel processing (a type of multithreading) which allows doing a lot of work by splitting it up among multiple threads that run concurrently.

Asynchronous and reactive programming are two other types of concurrency. Although similar, reactive programming is based on asynchronous events such as user input rather than asynchronous operations. Asynchronous programming enables responsiveness in an end-user **Graphical User Interface (GUI)** application, whereas it provides greater scalability in a server-side application.

Asynchronous programming is a type of concurrency that uses promises (or futures) or callbacks to avoid using unnecessary threads. In a broader sense, a promise is something that will be completed in the future, and an asynchronous operation is something that is started and will be completed later. The heart of these operations is that, as a result, they do not block the thread that invoked them (usually the user interface thread), leaving it free to perform other tasks, and upon completion, they notify their promise or invoke their completion callback event.

In modern .NET, the keywords assigned to asynchronous operations are both **async** and **await**, where an asynchronous method must return task types (**Task<T>** or **Task**) that represent promises and notify the calling code when they complete.

7.4.2 Communication

Concerning the communication between modules, which must allow loose coupling, it can be accomplished using the architectural style Implicit Invocation described in Section 3.3.2.3, which allows the system to do method invocation implicitly. Although using a single communication channel, such as a Message Broker, is preferred, it is not feasible for two reasons:

- **Enterprise**, as existing solutions are only geared toward business environments, necessitating the use of servers, which is not feasible given that the solution should work offline;
- **Multiplatform**, because even if a one-time solution is possible, it would necessitate native implementations for all platforms, which is something to avoid.

Thus, while the Publish-Subscribe architectural style is not as transparent and scalable as the event-based one, it is more effective and practical in terms of using platform-independent—the expected behaviour.

7.4.3 Design Patterns

The design patterns are dispersed throughout the implementation and should not be regarded as dogma (Esposito and Saltarello, 2014, p. 76), but rather used to make a genuine difference. They can frequently backfire (become an anti-pattern), be counterproductive, and add unnecessary complexity to the project.

The following patterns are used throughout the solution:

- **Dependency Injection**, in which the instances of the objects are controlled and provided by an **Inversion of Control (IoC)** container (*e.g.*, Unity¹⁶, Autofac¹⁷). Those instances are then placed and made available via an interface

¹⁶ <https://nuget.org/packages/Unity/>

¹⁷ <https://nuget.org/packages/Autofac/>

that allows modules to implement a bootstrap sub-module (*q.v.* Figure 7.5 on page 197), which is then in charge of filling their **IoC** container with the appropriate objects;

- **Singleton**, so that it is guaranteed to be working with the same instances of objects and respective data/information that are accessed by multiple modules. If the modules in question use Dependency Injection, the **IoC** Container should control the life cycle of the objects instead, because it usually also supports Singleton, eliminating the need to implement one;
- **Facade**, so that (especially) modules only show/offer what is absolutely necessary, and these are the access points that could implement the **IConsumptionMode** interface (*q.v.* Section 7.3.3).

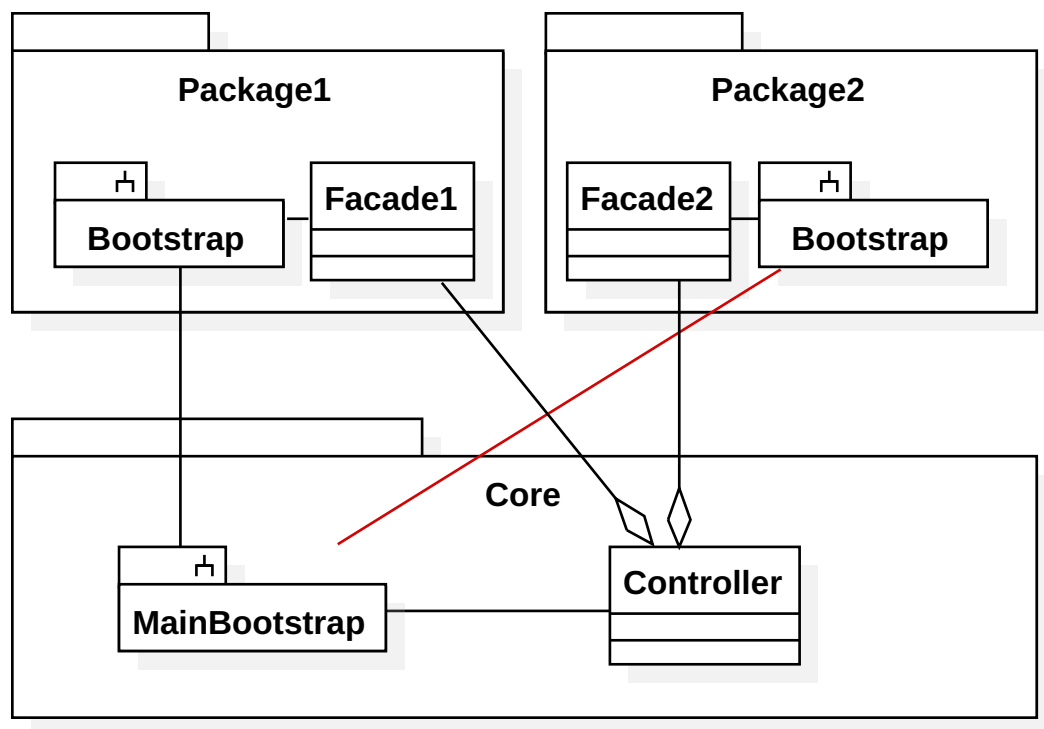


Figure 7.5 – Bootstrap Example Class Diagram.

The **Composite** design pattern is used in the **Framework** module to allow Commands to have no depth limit. As previously stated (*q.v.* Section 7.3.3), this enables

commands to be composed of sub-commands and/or sub-gestures.

Instead of recreating the necessary information, such as the command list, every time the Framework runs, a case can be made for using databases, possibly even through design patterns that are more common in enterprise environments. True, such an option could be useful in some situations, particularly as the consumer application and the number of Adapters and Decoders grow in size. Such scenarios were considered, albeit only theoretically for the time being, with, still, the following design patterns in mind:

- **Repository**, that provides a clean separation of concerns between logic layers (often referred to as business layers) and data access layers, thereby isolating the application database access, which can then be transparently changed to another **Object Relational Mapping (ORM)**. This pattern also aids in the reduction of code duplication, such as queries that are used repeatedly. It can be implemented in a variety of ways, enabling basic functions such as **Create, Read, Update, and Delete (CRUD)** or more sophisticated or common throughout the solution operations, with a repository for each **entity** that needs to be accessed;
- **Unit of Work**, that, unlike repositories used to build an abstraction layer between the **logical layers** and **data access**, coordinates the work of multiple available repositories that should share the same database connection. It is described by Martin Fowler as keeping “track of everything you do during a business transaction that can affect the database. When you’re done, it figures out everything that needs to be done to alter the database as a result of your work” (Fowler, 2002, p. 184). Thus, changes in one repository are reflected in others, allowing them to work with the most up-to-date information without having to worry about anything else.

All of these options are viable and ideal in their own right, but they are still dependent on the use-case. As previously stated, if not used correctly, design patterns can quickly become anti-patterns. The information required for the current InMERSE Framework to function is still very lightweight, and there is no real need to use even

a simple database, such as SQLite.NET¹⁸, let alone full ORM solutions like the Entity Framework¹⁹ or NHibernate²⁰. Again, such solutions were considered, but were deemed unsuitable for the InMERSE Framework's near-term future. Furthermore, one of the goals of this Framework is to target mobile solutions (*e.g.*, Android, iOS), not only for smartphone AR applications, but also for HMDs running on top of Android. Taking into account the solution specification (*q.v.* Section 6.3) and design considerations (*q.v.* Section 6.4), battery and performance, as well as processing and response time, could all have a negative impact on the experience. And accessing any kind of database would undoubtedly introduce some additional delay and have a negative impact on these mobile aspects.

¹⁸ <https://nuget.org/packages/SQLite.Net-PCL/>

¹⁹ <https://nuget.org/packages/EntityFramework/>

²⁰ <https://nuget.org/packages/NHibernate/>



CSKT Prototype

8.1 Introduction

The **CSKT** prototype can be considered the thesis's ultimate goal because it showcases a little bit of everything we've been working on for so long. It all starts with what we consider to be a necessary cultural foundation (*q.v.* Chapter 5 - **CSKT Case Study**), which we attempted to obtain with a six-month stay at the University of Montana and **SKC** in Montana, **US**, thanks to a Fulbright scholarship co-funded by **FCT**. As previously stated (*q.v.* Section 1.2), the outward goal of conducting research in the United States was to assess the feasibility of employing cultural gestures and symbols to facilitate more meaningful interaction in immersive environments (*q.v.* Section 2.2.4). The idea is to use users' cultural semantics to improve learning and teaching of complex concepts and systems by creating interactive stories or modelling interactivity in immersion, for example.

Following the logic of the original proposal ([Alves Fernandes et al., 2019](#)), and the overall goal of exploring the conjecture, as explained in Section 1.2, this **CSKT** storytelling prototype should employ immersive technology and somatic interaction in the visual-spatial context of the Flathead Indian reservation, enabling interaction via sociocultural-aware gestures and emblems. The previous work in this area (*e.g.*,

Pinto et al. (2015); Carvalho (2019)), provided examples of interaction with cultural gestures, but provided no guidelines on how one can design the immersive environment for exploring complexity with cultural gestures. This chapter clarifies this gap, by presenting the method devised for designing such an environment, based on the complex concepts identified in the cultural exemplar (the case study). This resulted, for instance, in aspects such as changing the current season from winter to summer or from day to night, and reflecting that through landscape changes within the immersive environment.

The CSKT case study (*q.v.* Chapter 5) serves as the prototype's foundation, in which we identified its complex concepts. For designing the immersive environment with which people can interact to explore them, we first decided that it should include a narrative, thus the design process starts with storytelling design (linear or nonlinear). Since the complex concepts emerged from a child story that piqued our interest, while in the US (*q.v.* Section 5.3), we decided to design the storytelling based on this story, albeit other avenues could have been pursued. These aspects were then used to establish a narrative (*q.v.* Section 8.2). Everything we include in the narrative, including the interaction and place, must have a valid reason, whether it is to add context or to serve as a link between two ideas. It should not be made up without purpose, because that would not serve the purpose of exploring our conjecture (*q.v.* Section 1.2). That is why we established the foundation for the entire narrative (*q.v.* Section 8.2.1). It sets the stage for the subsequent prototypes, which include the (VR) Animated Concept (*q.v.* Section 8.3), and a VR Prototype (*q.v.* Section 8.4).

8.2 Narrative

A narrative (and a script, *q.v.* Appendix C) was established as the foundation for the subsequent prototypes. Although it is heavily based on the children's book Luli's Journey (Pierre, 2010b), which we discovered during our visit to the US, we gathered facts from it and other cultural elements on the Flathead Indian Reservation (*q.v.* Section 5.3), which we then used to support propositions of our interest (*q.v.*

Section 5.3.5). After that, we define a script and iteratively update it, always with appropriate reasoning (*q.v.* Section 8.2.1). This seeks to ensure that both the story and its narrative follow some logic that allows us to explore our conjecture (*q.v.* Section 1.2).

8.2.1 Grounding

The entire script (*q.v.* Appendix C) is based not only on existing literature (*q.v.* Chapter 2), but also on previously collected and organised facts that support a set of propositions (*q.v.* Section 5.3). The full script is dissected in Table 8.1, below, with each part linked to one or more ideas that support the propositions.

Table 8.1 – Full narrative's script dissection.

Script subset	Grounding
South Crow Creek Area, south of Ronan, Montana, near Kicking Horse Reservoir, located on the Flathead Indian Reservation	P2 HASS - Place connectedness
Elder or someone from the tribe eligible to tell the story (even teachers of the Salish language tell them, so it's not something only elders can do)	P2 HASS - Place connectedness
Narration in Salish	P1 HASS - Collective phenomena - Evolution of language
Scenario starts in broad daylight. And changes accordingly, so that the story can be told. The environment can also change as the story progresses.	P2 HASS - Place connectedness

Table 8.1 continued from previous page

Script subset	Grounding
Stational experience/tracking.	Technical requirement. Target audience: young adults, mid-age, and elderly
Elder introduces himself and the tribe: “Hello! My name is Michel, and I am going to tell you about the Confederated Salish and Kootenai Tribes. This includes the Salish, Kootenai, and Pend d’Oreille tribes. Pend d’Oreille are also known as Kalispel”	Context
Asks the user’s name	Agency, and user customisation
Answers: “Nice name!”	Feedback
Elder says: “Before going any further, let me just tell you a bit about complex systems.”	Connection point
Go over complex systems (removed after first iteration)	Literature
Elder says: “I hope that wasn’t too boring for you. What is important to understand is that the traditional divide and conquer approach is not always the answer. There are many things that cannot be explained in such a way.”	Literature; Connection point
Elder says: “Now a bit more about the tribes.”	Connection point

Table 8.1 continued from previous page

Script subset	Grounding
Elder tells the origin of Salish and the location of the experience	Context (subsection on context)
Tell the user to look into the mountains (Eagle's pass)	Agency
Elder shows some pictures of the reservation and says "Great. Now take a look into these pictures to see what our land actually looks like!"	Context
Show interactive ring menu	Agency
"We take from the land, wildlife, and water, so it's only right to give it back in some shape or form."	P2 HASS
Show the example of structures to help bears cross a road (bridge or tunnel) while saying: "An example of that is trying to not disrupt the wildlife. We build bridges and tunnels so that bears and others can cross roads safely."	Context (example)
Elder mentions the importance of art, such as paintings/drawings and colours, music, etc: "We also give utmost importance to drawings, paintings, colours, and music. They are a big part of our culture. They show our vision of the world and our surroundings."	Connection point (Unit of analysis - Evidence (page 16 and 17))

Table 8.1 continued from previous page

Script subset	Grounding
Show in sort of an interactive ring menu with some photos of paintings, etc., taken there.	Agency, and context (example)
Elder explains the importance of story-telling in Salish (and other tribes) culture	P4 HASS part 4 - Circle of Life of people below
“Stories always tell about mischief, strength, and growing (e.g., a tree).”	P2 HASS - Connectedness and meaning to the land
“Firstly, they were (and are) a way of conveying an idea or teaching, often in a nonlinear way.”	P4 HASS part 4 - Circle of Life of people below
Can perform certain gestures here to represent these. Closing hand with palm facing upwards.	Context (example)
“A single story often tells about the circle of life. Today is no different!”	P4 HASS part 4 - Circle of Life of people below
“They solidify the culture. They are a way of transmitting knowledge from generation to generation. There were no schools back then; you would learn by experiencing it.”	P1 HASS - Collective phenomena - Evolution of Language; and P4 HASS part 4 - Circle of Life of people below
“An important remark to make is that these stories are not necessarily scripted. Each tells them their own way, with their own words.”	Context

Table 8.1 continued from previous page

Script subset	Grounding
“Another is that there might be specific scenarios that define when and where these stories can/should be told.”	P2 HASS - Place connectedness
“A good example of this are the Coyote stories. They are about life lessons.”	P4 HASS part 4 - Circle of Life of people below
The coyote can be represented as a real animal in the scene, or through a spirit-like animation floating around.	Context (example)
“The coyote, for the Salish—other tribes might have different animals—, is not necessarily a bad character (he can be). He’s more of a mischievous character.”	P2 HASS - Connectedness and meaning to the land
“They are a teaching point to get in the right path.”	P2 HASS - Connectedness and meaning to the land
“They can only be told in a certain time period, particularly the “down months”. That is, during winter (with snowfalls). They cannot be told after the first bloom of Spring.”	P2 HASS - Place connectedness
“This is because of Cultural beliefs. Salish, and other tribes, have signs of bad omens. One of which is the Owl—sign of misfortune, death. They are messaging birds.”	P2 HASS - Connectedness and meaning to the land

Table 8.1 continued from previous page

Script subset	Grounding
“Other stories might be told in some specific days, depending on the memories they bring. Context matters!”	Context (interview)
“Now that we went through a bit of the Salish culture and the importance of storytelling, it’s time to set an appropriate environment to go through a story ourselves!”	Connection point; and P2 HASS - Place connectedness
Gesture to change from broad day light to night time. 1. (Environment changes) 2. Certain background song starts	Context (example); and P2 HASS - Place connectedness
“This is based on the children’s book story ‘Luli’s Journey’ about a little Canadian Goose called Luli, her family, and their journeys.” 1. Show picture of book while talking	Context (authorship)
“It was written by Alyssa Kelly and Jess Nenemay, and translated by Pat Pierre. It was partly funded by the Administration for Native Americans.”	Context (authorship)
“There are many birds on the reservation. Some are big and some are small, but a Canadian goose named Luli is different from them all.”	P2 HASS - Connectedness and meaning to the land

Table 8.1 continued from previous page

Script subset	Grounding
Ask the user to draw a bird, by giving her/him a frame to do so: “Can you draw a bird?” The drawing in mid-air can be done with pinch gestures.	Agency
Show a hologram of a Gosling, when the drawing is detected as a bird.	Feedback (literature)
“Luli hatched from an egg. Her nest is on the shore of McDonald Lake. Her mother and father were also born there. They raised Luli with love and care.”	P1 HASS - Family & Home bonding
Show a hologram with father, mother, and Luli as gosling.	Context (example)
“Summer brings a lot of fun as Luli learns to fly and run. She wades in the water as she follows her father and eats berries or plants her mother brought her”	P3 HASS - Hierarchy and self-organisation (family); and P4 HASS part 1 - Actions and Learning
Show a hologram with father, mother, and Luli a bit older.	Context (example)
“Luli has grown so big and so strong, her mother says it won’t be long until their journey to the land of the sun.”	P4 HASS part 2 - Growing & Circle of Life

Table 8.1 continued from previous page

Script subset	Grounding
Ask the user to make the goose bigger: “Can you make the goose bigger?” * General gesture would be zoom in/- zoom out or using both hands * The cultural one for Salish would be the ‘growing’ gestures shown before by the elder.	Agency (cultural interaction)
* Show an hologram of a bigger goose	Feedback
“Luli’s father recalls the autumn days, before Kerr Dam was built, when he and his friends played on the muddy flats of Flathead Lake.”	P1 HASS - Family & Home bonding
* Leaves falling (autumn like) * Sound of leaves and wind.	Context
“Long before winter’s frost blows in, Luli prepares to leave her friends. She ruffles her feathers and away she goes. Maybe she will travel to Mexico.”	P2 Science - Environment and land connectedness; and P4 HASS part 2 - Growing & Circle of Life
* Show map of North America with POIs being Montana and Mexico	Context

Table 8.1 continued from previous page

Script subset	Grounding
“Can you make Luli fly?” * hologram of Luli appears, with some indication pointing upwards * User performs cultural gesture of lifting the bird onto the sky * General gesture could be grabbing the bird in direction to the sky	Agency
* Elder points/looks to the sky as animation goes from the fire to the sky	Cue
* User needs to look to a specific location, too, as the Geese start showing up in a V formation * this happens by showing some sort of arrow to show where to look at, etc. * User needs to look back to fire/elder	Agency
“Luli’s family stays together in flight, they fly in a ‘V’ through the day and the night. People watch from the earth below. Canadian Geese let them know: if the summer will be hot and really dry or if the winter will be harsher as the days go by.”	P1 Science - Collective phenomena - “not the simple sum”; P2 Science - Environment and land connectedness; and P4 HASS part 3 - Learning (people)
Snow starts falling (next to the user and fire); the environment starts changing.	Context, and cues

Table 8.1 continued from previous page

Script subset	Grounding
“Weeks of hard flying have come to this: Luli arrives to sun and bliss. Mexico has warm sand and an ocean breeze, but Luli was excited most to see the palm trees.”	P2 HASS - Connectedness and meaning to the land
* Show hologram of Palm tree / beach diorama	Context, and cues
* Play some sound of waves	
“Months have passed on the paradise shore, but Luli feels like she needs so much more. She remembers her nest near McDonald Lake; the mountains and trees were her favorite place.”	P2 HASS - Connectedness and meaning to the land
* Environment starts changing to Spring (not as much snow; trees will start popping)	Context, and cues
* Change to a more appropriate song for spring; birds chirping, etc.	
“Luli’s parents tell her winter has passed and it’s time to return to their home at last. So they ruffle their feathers and fly away north in a ‘V’, each taking their turn to be in the lead.”	P2 Science - Environment and land connectedness; P3 Science - Hierarchy and Self-organisation; and P4 Science - Feedback loops and emergence

Table 8.1 continued from previous page

Script subset	Grounding
* Show globe/map with an animation from Mexico to the reservation (a simple line suffices). * Or again timeline, where the user needs to drag goose from one location to another.	Agency
* Elder looks to the sky again	Cue
* User needs to look as well	Agency
* Birds show up flying in V again	Feedback
“Spring melts the snow and flower buds grow. Luli’s family arrives after winter’s last blow. The time when birds arrive after a late winter snowstorm is called Bird’s Sock.”	Context; P4 HASS part 2 - Growing & Circle of Life; and P4 HASS part 4 - Circle of life of people below
* No more snow in the environment.	Context, and cues
“Wild flowers bloom on the mountain-side and blue skies make the earth come alive. Luli and her friends have all grown, but her lakeside nest will always be home. Someday, Luli will lay eggs of her own and she will take them on the same journey that she has flown.”	P1 HASS - Family & Home bonding; P2 HASS - Connectedness and meaning to the land; P4 HASS part 2 - Growing & Circle of Life;

Table 8.1 continued from previous page

Script subset	Grounding
“The birds’ knowledge of seasons is so precise that the Native people depended on them to survive. The Canadian Geese’ annual migrations gave the people an idea of weather patterns as well as timelines for season changes.” “Tribes look forward to birds as harbingers of Spring.”	P4 HASS part 3 - Learning (people)
“Would you like to know more about the information the Native people gathered from observing the birds and the environment?”	Context; and P4 HASS part 3 - Learning (people)
* User needs to answer with a gesture for yes or no * Salish would be thumbs up (or fist up) * General would be a ‘ring’ (done with thumb and index fingers)	Agency
If the answer was no; Move to next question.	Feedback
If the answer is yes	Feedback
“Cold weather triggers migrations.”	P2 Science - Environment and land connectedness
* Wind blowing (with sound effect, too)	Context, and cues

Table 8.1 continued from previous page

Script subset	Grounding
“When the birds fly to the South, that would mean the tribes needs to gather food, shelter, and settle down for Winter.”	P2 Science - Environment and land connectedness; and P4 HASS - Learning (people)
“They can use the Sun, and the stars as guidance.”	Context
* Show hologram of Sun and stars	Context
“They can also use landmarks, such as coastlines, rivers, or mountain ranges.” “They can even use Earth’s magnetic field!” “They can go as far as Brazil or Argentina. But most get to Texas or Mexico, which will already be considerably warmer.”	Context
“Tribes also used other information available in the environment.” “When bears start disappearing; when they start hibernating, it’s because they are also preparing for winter.” “You can also tell it is getting colder by watching the mountains. The peaks will start to get white first, and it will slowly go down the mountain, like a blanket.”	P4 HASS - Learning (people)

Table 8.1 continued from previous page

Script subset	Grounding
“Hopefully this will give you some insight into using the land, and the wildlife, as part of your life.”	Connection point; and P4 HASS - Circle of life of people below
Would you like to know more about the emergent nature of V-shaped flights?” (Yes/No)	Context; Agency; P1 Science - Collective phenomena - “not the simple sum”
If the answer was no; Move to next question.	Feedback
If the answer is yes	Feedback
* Show the birds in a V-formation as a hologram next to the fire.	Feedback, and context
“As each bird flaps its wings, it creates an uplift for others behind him. There is 71 percent more flying range in V-formation than flying alone.”	Context; and P1 Science - Collective phenomena - “not the simple sum”
“Whenever a goose flies out of formation, it quickly feels the drag and tries to get back into position”	P4 Science - Feedback loops and emergence
“When the lead goose gets tired, it rotates back into the formation and another goose flies at the head”	P3 Science - Hierarchy and Self-organisation
“The geese in formation honk from behind to encourage those up front to keep up their speed”	Connection point

Table 8.1 continued from previous page

Script subset	Grounding
“When a goose gets sick or wounded and falls, two geese fall out and stay with it until it revives or dies. Then they catch up or join another flock.”	P1 HASS - Family & Home bonding
“That is it for now! Would you like to explore this environment freely?” (Yes/No)	Agency
If the answer was no; Move to end point.	Feedback
If the answer is yes	Feedback
Elder says: “Hope we can meet again soon! Goodbye.”	Connection point

8.3 VR Animated Concept

Any type of development is ultimately an iterative process. Whether you want them to or not, mistakes or last-minute changes will undoubtedly appear and force a project to be updated in some way. Early prototyping is a useful method for developing an idea. Low-level prototyping, such as sketches, and paper or cardboard prototyping, can help achieve these goals (Kelley, 2001). Fast prototyping enables an idea to develop much more quickly. Creating machines to aid in the current COVID-19 pandemic is no exception (Cooper, 2021). Elon Musk’s SpaceX is also based on this concept. When developing a new rocket, it appears that they conduct a flight test every week until it is perfected (Vance, 2015, p.64). It’s no different with a storytelling prototype. Trying to develop the final, perfect project from the start is not a good idea—it is even utopian. It is a good idea to iterate quickly, or at least have

a way of doing so, rather than developing too much and having to change everything at the end, especially when it comes to the development of immersive environments, which ends up being more cumbersome than normal. After iterating on paper with the narrative, we decided to create an animated concept of what the ideal UX would be in our minds. We knew that doing so would allow us to keep iterating quickly, allowing us to spot errors or come up with better ideas for specific parts. The plan was to start with a 2D video created in Blender¹ using 3D models, which led to the use of Tvor² to create a VR animated concept instead. It is an excellent tool for both AR and VR experiences, especially for quick prototyping and animation beginners. After porting an early 2D concept from Blender to Tvor, users could be placed on the actual experience—as if it were a fully fledged VR experience (6DoF). The difference is that it does not support interactions; they were still animated, but the user could see them in real time. This resulted in the development of an actual VR prototype (*q.v.* Section 8.4), which included the missing user interactions.

8.3.1 Environment and 3D Modelling

A narrative tends to occur in some meaningful place with some predefined characters. Ideally it should leverage not only the visual and spatial context, but also its cultural meaning. In our particular example, we knew it should have some resemblance to the Flathead Indian Reservation, but we also knew we needed a quick version of it to begin prototyping the animated concept. Although we investigated and planned to use several 3D mapping techniques, including through Satellite data, and Photogrammetry, the current COVID-19 pandemic forced the SKC college to close for an unknown amount of time. We were also aware of the drawbacks of working in such high-quality environments, as people begin to look for flaws rather than focusing on the experience (*q.v.* Section 2.2.4). As a result, we chose a low-poly environment. Simple enough to construct while remaining viable (*q.v.* Chapter 2). It was actually created within Unity3D³ by stacking several free Creative Commons BY (CC-BY) assets (the individual licences and respective attributions are placed

¹ <https://www.blender.org/>

² <https://tvori.co/>

³ <https://unity.com/>

next to the animated concept itself). It was then exported to Blender and updated with new materials, as well as new elements such as mountains, trees, and houses, among others. The environment's design was inspired by children's book stories we read when in the US. We specifically based ourselves on the Blue Jay Brings the Chinook Wind (Pierre, 2010a), where they explain not only the use of storytelling but also depict the Ronan area, located within the Flathead Indian reservation in Montana, US (Pierre, 2010a, p.35). We made certain landmarks, such as the Eagle's Pass, a V-shaped mountain peak, and the number of houses nearby, to be included in the low poly environment (*q.v.* Figure 8.1). These kinds of details are what end up bringing context and some familiarity to the people who live there—after all, that is the whole point.



Figure 8.1 – CSKT Prototype low poly environment.

8.3.2 Dialogue

Although one could theoretically have a fully exploratory experience, not having clear limits of what is possible and what is expected from the user could negatively affect the outcome. To avoid this, one can introduce dialogues, which also entice more engagement and create more immersion. In our particular example, we decided to record the dialogues in English ourselves, even if they were far from perfect. For the time being, they are only placeholders. This is due not only to the current pandemic, as it is not feasible to have some elders from the tribes record themselves in Salish, but also to the fact that they are still subject to change. As a result,

they serve as placeholders for rapid prototyping. When we have the opportunity to show it to them, to the tribes, and they approve, we will be able to replace the dialogues with Salish ones. Having said that, for the time being, the user tests (*q.v.* Chapter 9) are being conducted with Portuguese people, some of whom do not speak or understand English, so we decided to include Portuguese subtitles. It is critical, and one of our pillars, to try to tailor the experience to the end-user rather than the other way around.

Throughout the first 2D animated concept, created in Blender, various elements appear and disappear at various points in the story. In a related vein, all of these extra elements were created, or at the very least published, by Google (CC-BY license) via the now-defunct Google Poly. Despite the fact that it no longer exists, a community option called Poly Pizza⁴ began to appear and imported the publicly available assets from Google Poly. This 2D animated concept was narrated entirely in English. That is, there was no visual character to which the voices could be attributed (*q.v.* Figure 8.2, above). Ambient sounds, including music, were also included, all of which were royalty free.



Figure 8.2 – 2D animated concept without avatar.

Following this initial concept, we realised it would make more sense if there was an

⁴ <https://poly.pizza/>

avatar that appeared to be talking, as if it were an elder or someone knowledgeable about the CSKT. Animating a humanoid, on the other hand, is a difficult task, and we weren't looking for a professional solution either. Still, this is the point at which we began experimenting with Tvor⁵, which allows the use of **Forward Kinematics (FK)** and **Inverse Kinematics (IK)**, as well as some mimicry via real-time animation. There is a (good) but, however. Tvor is a **VR** application that requires the use of a **HMD** and joysticks. Given the different joystick buttons and movement combinations (which we are actually looking to replace with more meaningful interactions), it takes a little time to get used to, but it takes less time than learning how to use Blender or any other professional animation tool. Furthermore, it already has some ready assets and **Visual Effects (VFX)** that we assume are all created by them, including the avatar we chose to use (although we customised it slightly), *q.v.* Figure 8.3.



Figure 8.3 – CSKT Animated Concept with Narrator's Avatar.

Tvor, as previously stated, is an excellent tool, particularly for inexperienced animators. Real-time animation allowed us to pretend to be the avatar by controlling it with puppetry (*q.v.* Figure 8.4).

⁵ <https://tvori.co/>



Figure 8.4 – Tтвори’s Puppetry showcasing strings attached to the puppet’s head and elbows.

Although it isn’t perfect—it’s not meant to be—it allowed us to quickly perform those movements without any looping or repetition, which could degrade the experience and provide no meaningful context at all. We try to emphasise certain parts of the dialogue by gesticulating while they are being said. It took a few iterations to make it look less robotic and to try to time it correctly, but it works. The ultimate goal was to create a **VR** immersive experience, so it appears that using **VR** directly to do so is the way to go. The alternative would be to hire a (semi) professional animator or rely on extra equipment with depth cameras, such as the old Microsoft Kinect 2 or the newer Intel RealSense. While the movements would be better captured, the setup would undoubtedly take longer, and we were looking for a quick prototyping method instead. Of course, it is dependent on computer hardware. Tтвори has a public version for **Personal Computer (PC) VR**, although they are currently testing a lightweight version for mobile platforms called Shapes (*e.g.*, Oculus Quest). Having said that, the development of this **VR** animated concept with Tтвори was made possible by the use of an Oculus Quest and an Oculus Link, as previously explained (*q.v.* Section 6.2). All processing and rendering is then performed on the PC before being streamed via USB-C and properly synced with the **HMD**. The computer’s

hardware specifications are as follows:

- **CPU**: Intel® Core™ i7-6700K @ 4.0Ghz (8 CPUs)
- **Graphics Processing Unit (GPU)**: NVIDIA GeForce GTX 1070 8GB VRAM
- **Random Access Memory (RAM)**: 16GB 2400MHz
- **Solid-state Drive (SSD)**: Samsung 850 EVO 250GB
- **Hark Disk Drive (HDD)**: Western Digital Black 2TB

8.3.3 Interactions

Defining a narrative and the environment content might not suffice if not thought together with the user interaction, and vice-versa. Introducing the narrator's avatar into the environment brought some new challenges. For example, the user is asked for his or her name early in the script. When done with audio, the user may be able to simply respond normally. It might be a good idea, however, to cue the user with the actual time and action that is expected with the avatar of the person asking the name. We devised a solution for this by providing the user with a virtual microphone given by the narrator's avatar, which can then be used to say the name. It does bring some more agency to it. Even if it is only a concept intended for watching, we believe it is important to demonstrate how gestural and somatic interactions would work in a full-fledged VR experience. For example, when presenting the microphone to the user by moving it closer to the camera, it is desirable to show and grab the microphone with a hand rather than having it levitate.

This implies adding more detail is added to this concept, however. Not because adding an arm to the animation is difficult, but because the rigging and finger movements we want to show are particularly difficult. We believe it is still worthwhile to devote time to this issue, so we contacted a professional animator⁶ who was

⁶ Paulo Cristóvão, Invited Professor at the University of Algarve, provided immense help through suggestions and guidance.

a part of the original InMERSE project. He had rigged hands for the Digital Signage project back then, which he could now repurpose. To our dismay, it is not possible to import animations into Tvorì at all, so we couldn't just ask him to do the animations for us again, and then have them appear and disappear at the appropriate times. Despite the fact that the animations had to be done entirely within Tvorì, he was still kind enough to prepare some images for us to use as a reference when attempting to perform some actions, and even to suggest some additional changes to the concept. As a way of progressing, we decided to use this type of cueing and interaction form throughout the narration portion of the VR concept. It allows the user to participate, albeit in a scripted and linear manner, which we consider an improvement over simply listening.

One of the changes suggested by the professional animator, was to rig and animate the other characters that appear throughout the story, such as the geese, coyote, and bear. Even if it is just a concept, making them appear and change position without actually moving their legs or wings does not feel right. We then iterated once more and spent some time in Blender rigging all of the characters in the story (*q.v.* Figures 8.5, 8.6, 8.7, 8.8, and 8.9).

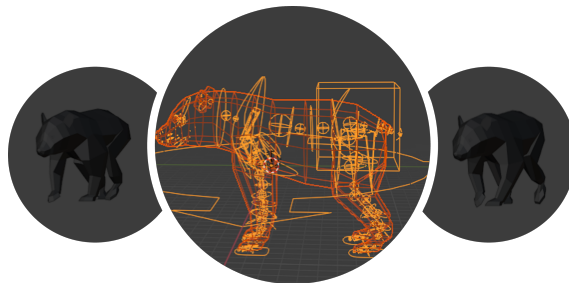


Figure 8.5 – Bear rigging and animation.

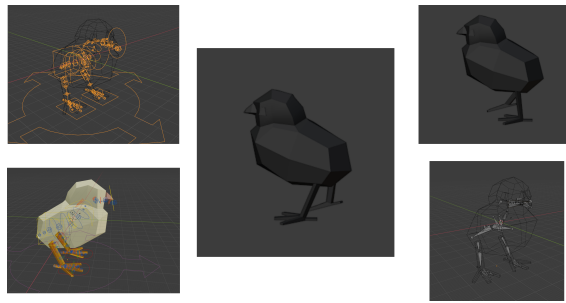


Figure 8.6 – Chick rigging and animation.

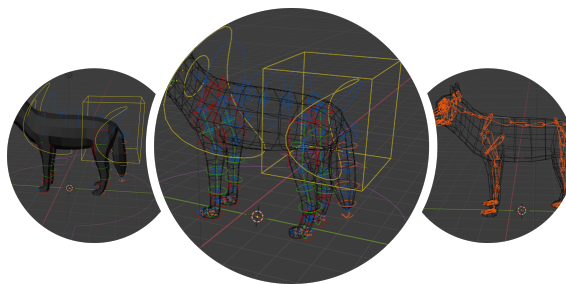


Figure 8.7 – Coyote rigging and animation.

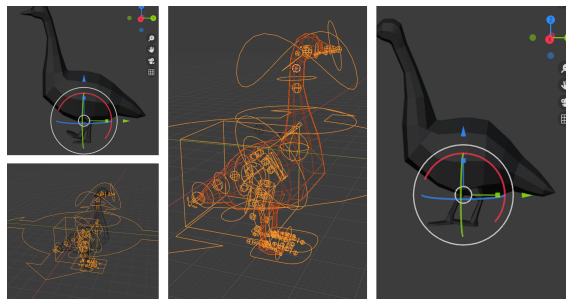


Figure 8.8 – Goose rigging and animation.

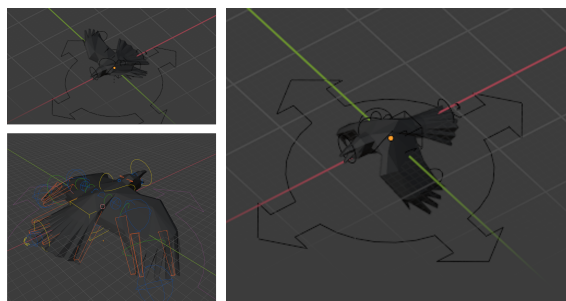


Figure 8.9 – Bird rigging and animation.

When we finally got a VR animated concept that felt right, we consulted with some more people, including this thesis' supervisors, on some of the dialogues before moving on to the actual VR Prototype. Aside from some translation corrections and volume changes to some of the ambient sounds, it was brought to our attention that a section of the script, particularly the type of long, early explanation on complex and chaotic systems, could undermine the overall goal of the experience, which is to explain complex concepts and/or systems to non technical users. Understanding complex systems is one of the requirements for effective complex learning, according to the literature (*q.v.* Section 2.2.2), particularly Ifenthaler. While this may lead to the assumption that one must understand all of the mechanics of complex systems before understanding a complex system, this is not true. We can get around this by explaining the concept or system through the story itself, giving the learners some intuition on the subject. That is the entire point of our conjecture (*q.v.* Section 1.2). Furthermore, because the user appears to only advance the story and not interact freely with it, it was suggested that some exploration be added to it. The user should then be able to interact with his or her surroundings more freely. This required us to revise the script once more before incorporating the changes into the VR concept. The story can then serve as the scaffolding required to convey the idea to the user, but the user should still be able to explore the concept in some way—that is the whole point of immersive environments.

8.4 VR Prototype

8.4.1 Foundation

The previously described, VR animated concept serves as the foundation for the actual VR prototype (*q.v.* Section 8.3). The animated concept has the advantage of being more than just a concept in that everything on it, including the narrative, content, and all of the animations, can be reused. What differs is the actual HCI component. In addition to a VR 6DoF experience, users should be able to see and interact with their own hands. This is part of the thesis's culmination, built

on top of Unity3D⁷. The other component is user testing (*q.v.* Chapter 9), which tells us whether the conjecture is promising or not. The former involves selecting appropriate culturally meaningful gestures, and one of the ways of doing so, as put forward by Carvalho (2019), is to directly ask the users what they would employ in a given situation. Although we selected defaults that we deemed appropriate for the Portuguese Culture, the set was expanded (*q.v.* Table 9.3) with data collected just before the actual User Testing (*q.v.* Sections 9.4 and 9.3).

The improved InMERSE Framework, which we discussed in previous chapters (*q.v.* Chapters 6 and 7), is another foundation for this VR prototype. The plan is then to integrate the entire VR animated concept and leverage Multimodal HCI by utilising the improved InMERSE Multimodal Framework. Although it appears to be a simple task, there were a few issues to work out, particularly with the gestural detection. Although we didn't go into great detail before, Decoders are indeed ready to use machine learning (*q.v.* Decoders). However, one of the observations we made was the use of the same coordinate system. Although it was thoroughly tested with a Leap Motion device, we performed the training and detection using its desktop mode. Leap Motion is a fantastic little device that was released in 2013, and it continues to outperform any other type of visual hand tracking available, including those now enabled by Facebook with Oculus Quest, which is why we have continued to use it all these years. The constant evolution of Leap Motion's device through firmware updates allows it to work extremely well. They began by supporting a desktop mode in which the Leap Motion was oriented upwards and the user used the hands on top of it. Some time later, VR began to resurface, so they began working on a VR mode in which the Leap Motion would be attached to an HMD and face what your eyes would see. However, switching to this mode has a peculiarity. As the device's position changes, so does the coordinate system. All testing for the improved InMERSE Framework was done through a WPF application and Digital Signage, both of which are desktop applications, so desktop mode was used. However, for the VR Prototype, it makes perfect sense to change that and use the VR mode instead.

⁷ <https://unity.com/>

Changing it to **VR** mode, however, means that the Decoder we previously had is no longer suitable for our new use-case. We had to retrain it while wearing the **HMD** with the Leap Motion device attached. When we did that, however, we began to consider making this setup completely wireless. Oculus Quest, which we had previously used to develop with Tvorì, now works wirelessly. The **VR** prototype can be designed to target the Android platform and thus operate in the same wireless fashion. Leap Motion, on the other hand, has only publicly supported desktop platforms, despite the fact that they have been working with Android for a few years now. One possible solution would be to completely replace Leap Motion with the Oculus Quest hand tracking system, thereby avoiding any issues. However, as previously stated, we have been working with Leap Motion almost since its initial release in 2013. Its performance is exceptional, and it truly makes a difference, as users feel as if they are immersed in the actual experience (Alves Fernandes et al., 2016). We were aware of a few options, one of which was particularly recent (Qian et al., 2019). They were able to attach the Leap Motion to the back of an Android smartphone in their research. They make it work by wearing a pocket Windows computer and streaming data from the Leap Motion sensor to a smartphone. However, because they use web sockets, there is some significant delay. While web sockets are a viable option, and are even used by Leap Motion to make it work for web platforms, using an **User Datagram Protocol (UDP)** based protocol to establish low latency connections with fault tolerance could be significantly faster. **Open Sound Control (OSC)** is one with which we are already familiar. Although this protocol was originally designed to exchange audio data, it has high interoperability, rigour, and flexibility and is now used to exchange other types of data (Schmeder and Freed, 2008). **OSC** is based on a client-server architecture. In our case, the server would be a pocket computer, and the client would be the Oculus Quest, which would receive data from the Leap Motion. Portalble, on the other hand, was capable of achieving high accuracy gesture detection (what percentage), which piqued our interest. The study describes the features used in their machine learning model, emphasising the significance of scale and rotation independent features. Even though the goal of this thesis is not to create or use the best gesture detection model available, we believe it has the potential to significantly improve user experience. After contacting the

Portalble study's authors over the last few years, they were incredibly helpful in understanding their gesture pose detection solution, as well as their motion detection solution, which is based on queueing and is incredibly similar to the one used by InMERSE to detect commands. The Portalble solution is primarily open-source. The machine learning model and script, on the other hand, are not. It is known that they used Accord.NET⁸, which is compatible with iOS and Android. We were able to create our own version of it by exchanging emails with Portalble's authors⁹. We created a model using Google Colaboratory and its Jupyter Notebooks¹⁰, Tensorflow¹¹, and Keras¹², similar to the Drawing detection model we use (*q.v.* Section 8.4.2), despite not knowing about compatibility issues. This meant that our model could not be loaded by Accord.NET, even if it was in the industry-standard **Open Neural Network Exchange (ONNX)** format. To work with Accord.NET, it had to be completely created with that library and actually trained through the experience itself, which defeats the purpose here, as we want different teams to work independently on different parts of it. That is, one team can concentrate on the immersive environment and the overall experience, whereas another team can concentrate on developing machine learning models that can be loaded and used seamlessly within that experience. And that scenario is exactly what we attempted to replicate here, as further explained throughout this chapter.

8.4.2 Drawing Detection

Among many other forms of expression, the **CSKT** tribes have a special affinity for painting and drawing. It is a way for them to describe their worldview. We then considered that incorporating such interaction would be a good way to bring a bit of their culture into the experience, even if it was simplified. Although this is not the primary focus of this thesis, we did spend some time investigating our options for this type of interaction. Although it is clear that the InMERSE Framework

⁸ <http://accord-framework.net/>

⁹ Special thanks to the authors of Portalble ([Qian et al., 2019](#)) Jing Qian, and Xiangyu Li, who provided immense help.

¹⁰ <https://colab.research.google.com/>

¹¹ <https://www.tensorflow.org/>

¹² <https://keras.io/>

can still support such a thing via the acquisition-mode and thus feeding the basic ontological frames to the consumer (in this case, a VR application), it is still a different modality of detection. Drawing becomes more about the visuals than the movements. This is due to the fact that there is no single way to begin and end a drawing, nor is there a single way to draw an object or symbol. It all comes down to visual and cultural perception. A few years ago, we became aware of another Google project called QuickDraw¹³, in which people would draw some doodles that Google requested while also guessing what others, drawn by other people, were. They did this to create a massive dataset¹⁴ with multiculturalism in mind, moving closer to the shamanic-interface definition—an interface that adapts to the person’s culture or leverages it in some way. That is, they use the dataset to detect what an object is despite being drawn in various ways all over the world. Given the availability of a dataset, we decided to use it ourselves.

With the dataset out of the way, the only thing left to do is build a machine learning model that uses and learns from it. However, a question arises: which machine learning algorithms should be used? The problem usually determines this. As a multi-classification problem, and by looking at some examples using the same dataset, we quickly realise that the norm appears to be the use of Deep Learning, via Neural Networks (NNs), specifically Convolutional Neural Networks (CNNs). NNs are inspired by human biology, specifically the nervous system, and are made up of interconnected neurons, allowing for multiple inputs and outputs. However, because this is not the focus of this work, we will not go through the entire theory of it, let alone explain all other algorithms that exist.

When training some machine learning models, there are several stages. It usually begins with the creation of a dataset, which in this case is already provided by Google, followed by preprocessing (*q.v.* Section 8.4.2.1), model definition and compilation (*q.v.* Section 8.4.2.2), training, and validation (*q.v.* Section 8.4.2.3).

¹³ <https://quickdraw.withgoogle.com/>

¹⁴ <https://quickdraw.withgoogle.com/data>

8.4.2.1 Preprocessing

We prepare the data for the yet-to-be-defined machine learning model at this stage. This is usually where data is normalised or transformed further so that it takes up less memory. In particular, we choose a number of classes ($N=100$), download 4000 entries from each, totalling 400.000 images, and resize them to 28x28 pixels (width, and height). We also convert all of them to grayscale and normalise the colour values to be between 0 and 1 (*q.v.* Figure 8.10). We then divide the data into two separate sets from the start: one for training and one for validation. The standard is to have 90% and 10%, 80% and 20%, or 70% and 30%, respectively. We chose the first option, a 90/10 separation (360.000 for training, and 40.000 for validation). We picked this option as NNs tend to perform better with more data. This separation is done to ensure that the model does not become biased towards the data used to test it.

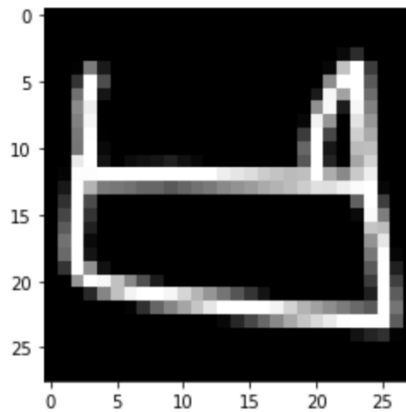


Figure 8.10 – Example of preprocessed image.

8.4.2.2 Architecture

A CNN architecture (*q.v.* Figure 8.11) was used in this work, with the first layer having an input size of 28x28x1 and the images being used in grayscale, with the actual hand drawing being white and the background being black. Three convolution layers, stride 1, and the Rectified Linear Unit (ReLU) activation function were

chosen for the feature extraction process in the CNN architecture to improve accuracy results. The second layer (Conv2D) is made up of 16 filters, each with a 3x3 size. The third layer is a Max Pooling layer with a 2x2 filter. The fourth layer (Conv2D) is made up of 32 filters of 3x3 size, all with the same activation function and padding, and is followed by another Max Pooling layer of 2x2 filter size. For the sixth layer (Conv2D), 64 filters of the same size (3x3) were used, followed by the same Max Pooling layer as before. A flatten layer was then used to normalise the output. Two fully connected layers were used for classification. The first dense layer employs 128 filters with ReLU activation, while the second dense layer, which is the final layer, employs 100 filters with Softmax activation. The Softmax function generates a probability distribution that sums to one of all classes at the end ($N=100$). The probability results show the percentage of an object's similarity to its class. Appendix D.1 (left) contains a more detailed view of the drawing machine learning model.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 16)	160
max_pooling2d (MaxPooling2D)	(None, 14, 14, 16)	0
conv2d_1 (Conv2D)	(None, 14, 14, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 32)	0
conv2d_2 (Conv2D)	(None, 7, 7, 64)	18496
max_pooling2d_2 (MaxPooling2D)	(None, 3, 3, 64)	0
flatten (Flatten)	(None, 576)	0
dense (Dense)	(None, 128)	73856
output (Dense)	(None, 100)	12900
Total params: 110,052		
Trainable params: 110,052		
Non-trainable params: 0		

Figure 8.11 – Architecture of Drawing CNN Model.

8.4.2.3 Training and Validation

The dataset was divided into training and validation sets with a 90-10 percent split. The Adam optimiser with a learning rate of 0.001 was used to train the model. The

categorical cross entropy was used as the loss function, and categorical accuracy was used to determine the model's accuracy. A batch size of 256 was used to train and validate the model. Finally, a validation accuracy of 93.12% was obtained.

8.4.2.4 Integration

With a working machine learning model that detects the drawings, an important step is still missing: integrating it with the experience. However, in order for that to happen, the model must be loaded into Unity3D. As previously stated (*q.v.* Section 8.4.1), one of the reasons we decided to build these systems ourselves was to use existing standards, in this case, **ONNX**. In Unity3D, these can be loaded using Barracuda¹⁵, a lightweight cross-platform (including mobile platforms) neural network package that allows them to run on both the CPU and GPU. After loading the model trained and saved in Google Collaboratory onto Unity3D, there is still the need to ensure that what is fed to the network is the same as what it expects. In other words, the preprocessing that occurs prior to training the model must still occur when the model is used to perform detections. First and foremost, we needed to figure out how to enable drawing in the experience. In the physical world, this is accomplished by grasping a pen or pencil with the hand, particularly the thumb and index fingers (sort of a pinching gesture). This is easily detectable with a device like leap motion, which tracks each finger individually. However, it is also possible to detect it using machine learning through InMERSE Decoders, which leads to another Gesture Detection system (*q.v.* Section 8.4.3). The Trail Renderers in Unity3D are a good way to show what is being drawn in real time. To ensure that the drawing fed into the machine learning model follows the preprocessing rules, we set up an additional camera that only captures the drawing's layer (*q.v.* Figure 8.12).

¹⁵ <https://docs.unity3d.com/Packages/com.unity.barracuda@1.0>



Figure 8.12 – Camera capturing drawing's layer only.

We invert the colours (black and white) and resize the render texture, provided by this extra camera when processing it, to 28x28 pixels. Then, using the runtime model, we create a GPU worker, feed it the textures in real time, and grab the top ten results, which, given the probability nature of the model's output, gives us the top ten drawing predictions (*q.v.* Figures 8.13, and 8.14).

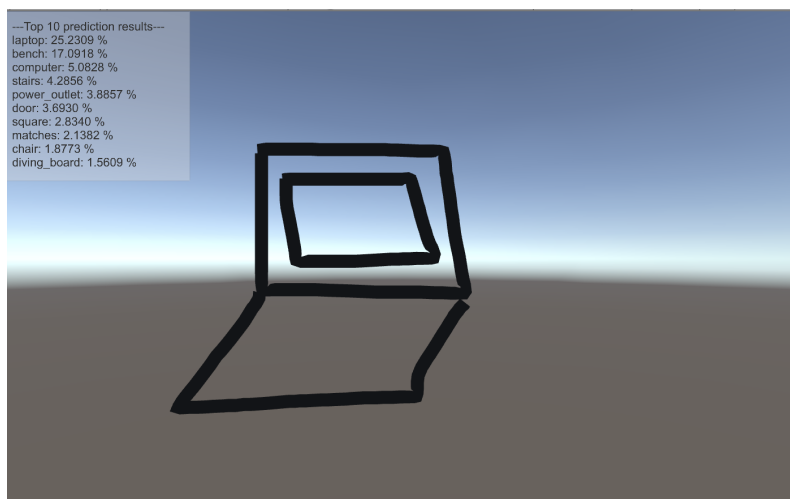


Figure 8.13 – Drawing system's predictions for a laptop drawing.

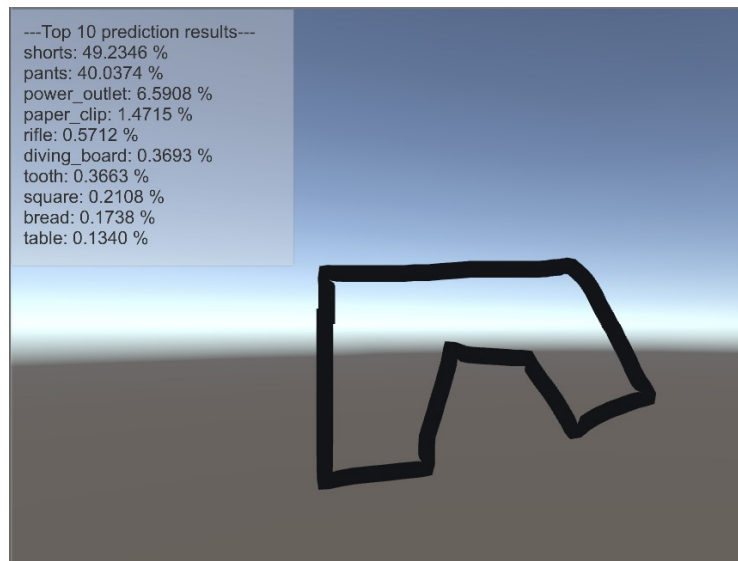


Figure 8.14 – Drawing system’s predictions for a shorts drawing.

This drawing system performs admirably, with no discernible lag in the predictions. For the time being, we’re only using it to draw a bird (*q.v.* Figure 8.15), but it could easily be used to spawn different objects (*q.v.* Figures 8.16) rather than selecting them from a list. This is especially true if there is a large selection of assets to choose from. Drawing gives you some flexibility, and it all boils down to the same type of detection. Even better, because we have access to the predictions, we can tell the user to keep attempting to draw the desired object, while providing additional feedback as needed.



Figure 8.15 – Drawing system in action within the VR Prototype.

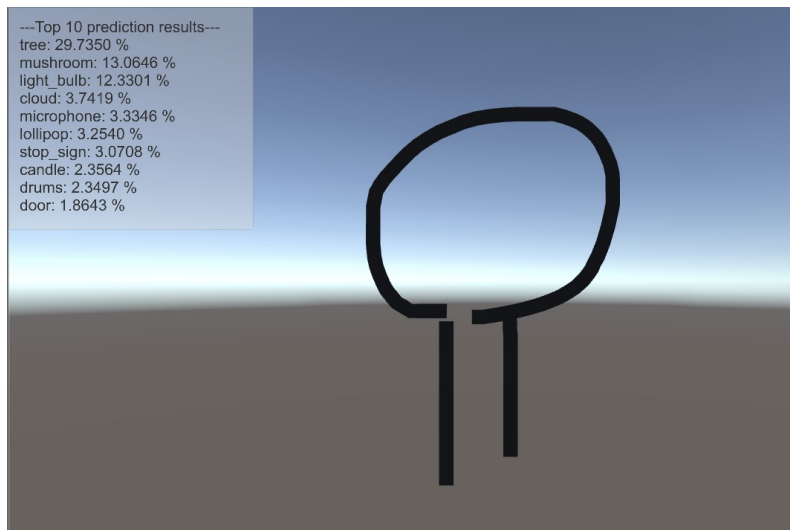


Figure 8.16 – Drawing system’s predictions for a tree drawing.

8.4.3 Gesture Detection

Another form of expression that is widely used is gesturing. As previously stated, the Portalble (Qian et al., 2019) solution is dependent on Accord.NET, which does not support ONNX. With their generous assistance, we based our solution on theirs by utilising the same features they did, namely rotation and scale invariance, which is critical when supporting different hand sizes and camera distances. Our implementation, like the Drawing detection (*q.v.* Section 8.4.2), was done using Google Colaboratory through its Jupyter Notebooks¹⁶, Tensorflow¹⁷, and Keras¹⁸.

8.4.3.1 Dataset Creation

This time around, however, the dataset still needed to be created, which was not as linear. The features we chose are essentially each fingertip and middle joint’s Fx , Fy , and Fz , for a total of 30 features per hand, as described in the Portalble (Qian et al., 2019) solution (*q.v.* Figure 8.17).

¹⁶ <https://colab.research.google.com/>

¹⁷ <https://www.tensorflow.org/>

¹⁸ <https://keras.io/>

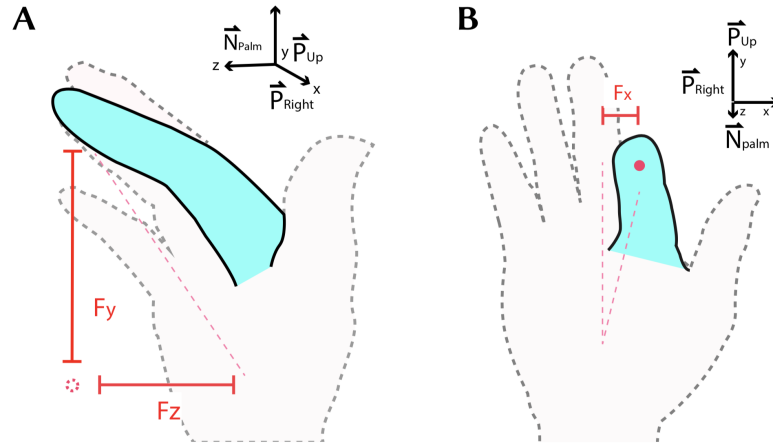


Figure 8.17 – Picture taken from Portalble (Qian et al., 2019). Each fingertip is represented by three features (F_x, F_y, F_z). F_z is a projection formed by a normal vector to the palm and the fingertip to the palm center ($NPalm$). The remaining two features are computed in the same manner, using a projection from the palm vector ($NPalm$) and each of the two local vectors PUp and $PRight$.

The dataset was actually created by running the sample WPF application from the InMERSE repository while wearing the Leap Motion in the forehead (as to imitate its position in the actual HMD). The recording was done in 6 sessions for each gesture ($N=5$, q.v. Figures 8.18, 8.19, 8.20, 8.21, and 8.22), with each session lasting 40 seconds. This resulted in approximately 27.000 entries per gesture. Despite the fact that each gesture had approximately 27.000 entries, only 2.000 were used after the entire set was shuffled. This is for memory reasons, as Google Colaboratory allows the use of 12GB of RAM, which allows up to around 36.000 entries to be used at the same time before it crashes. It is important to note, however, that the entire preprocessing is done at the time of recording, so it is not required later. The only thing needed is to remove three metadata columns, containing the gesture ID, subject ID, and attempt number. Similarly, when running the model, the same preprocessing is used before feeding the network. This preprocessing is the direct normalisation of all features in its InMERSE Decoder. The ontological basic data is sufficient for the Decoder to calculate and then normalise all of the features (distances).

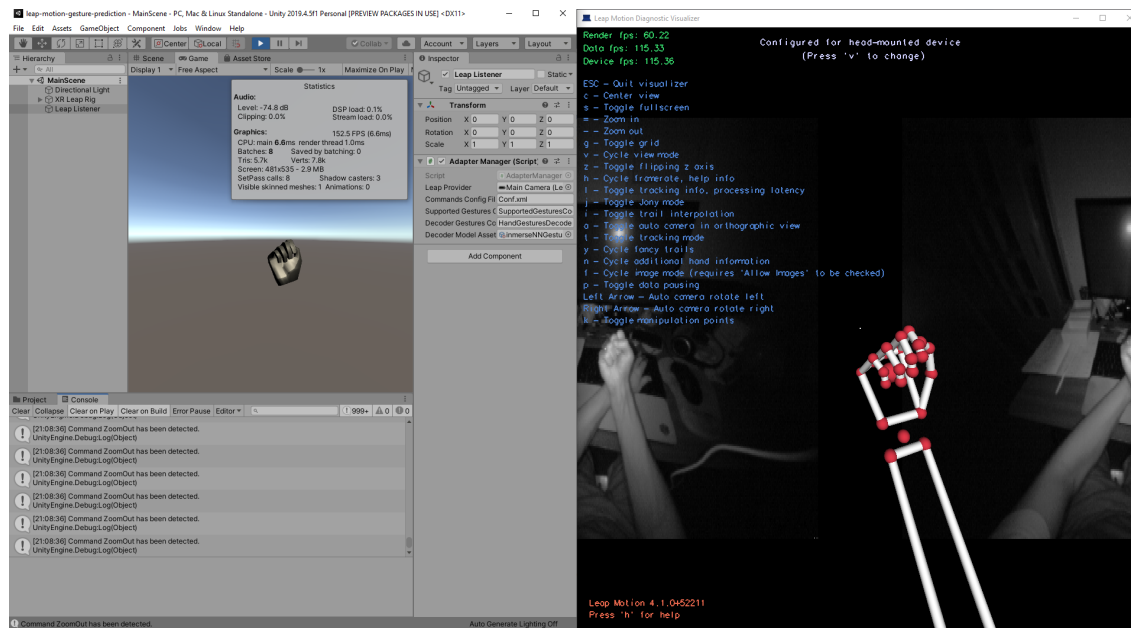


Figure 8.18 – Closed gesture.

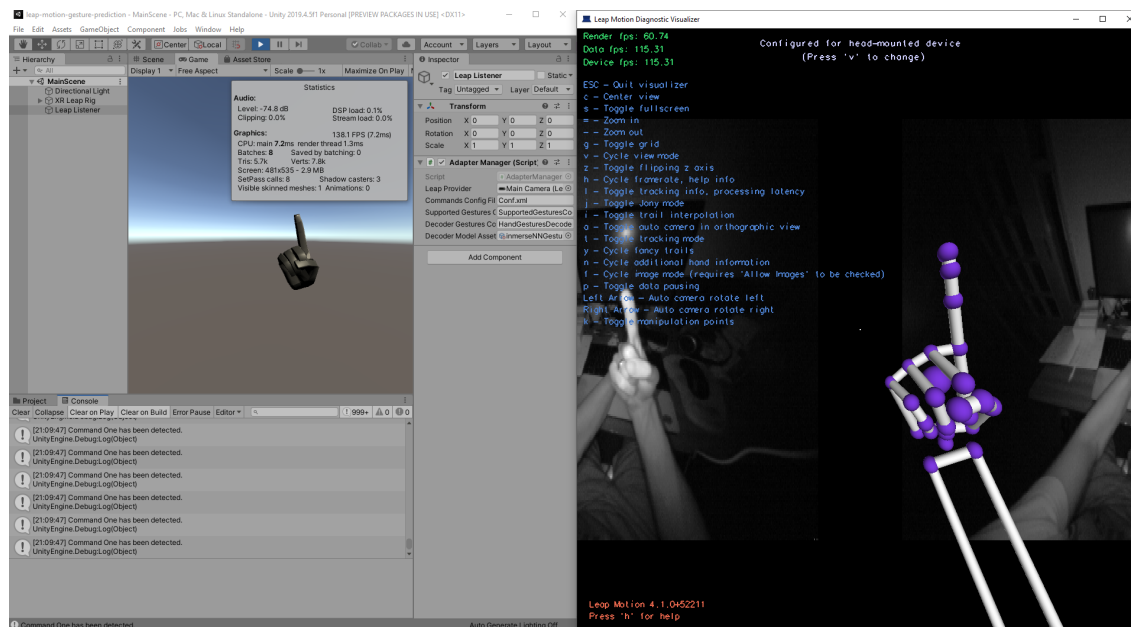


Figure 8.19 – One gesture.

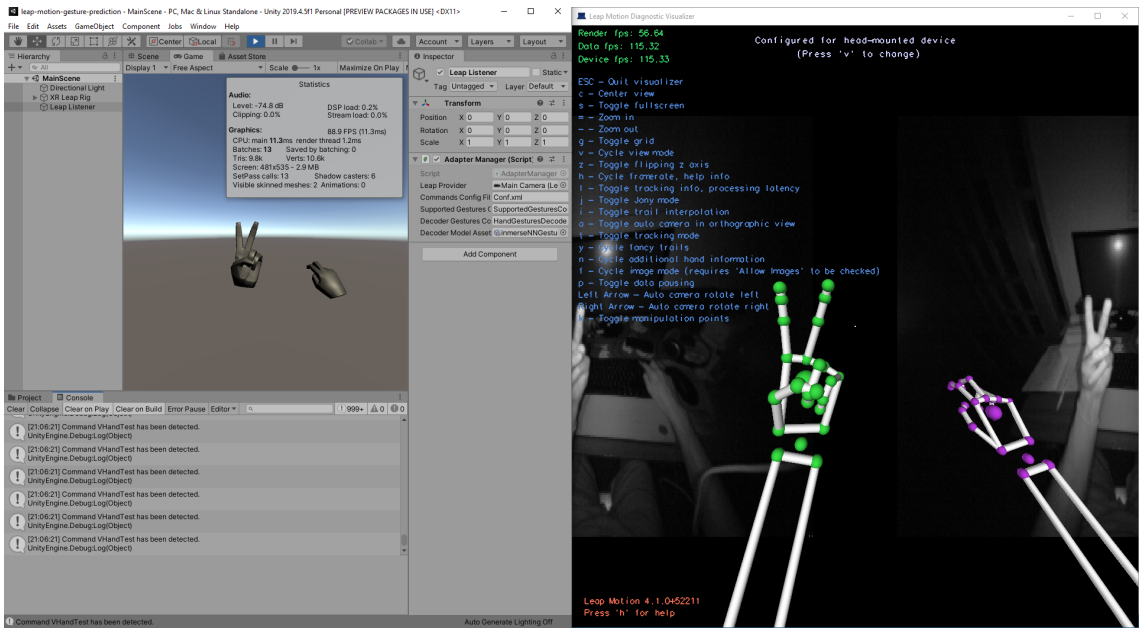


Figure 8.20 – V gesture.

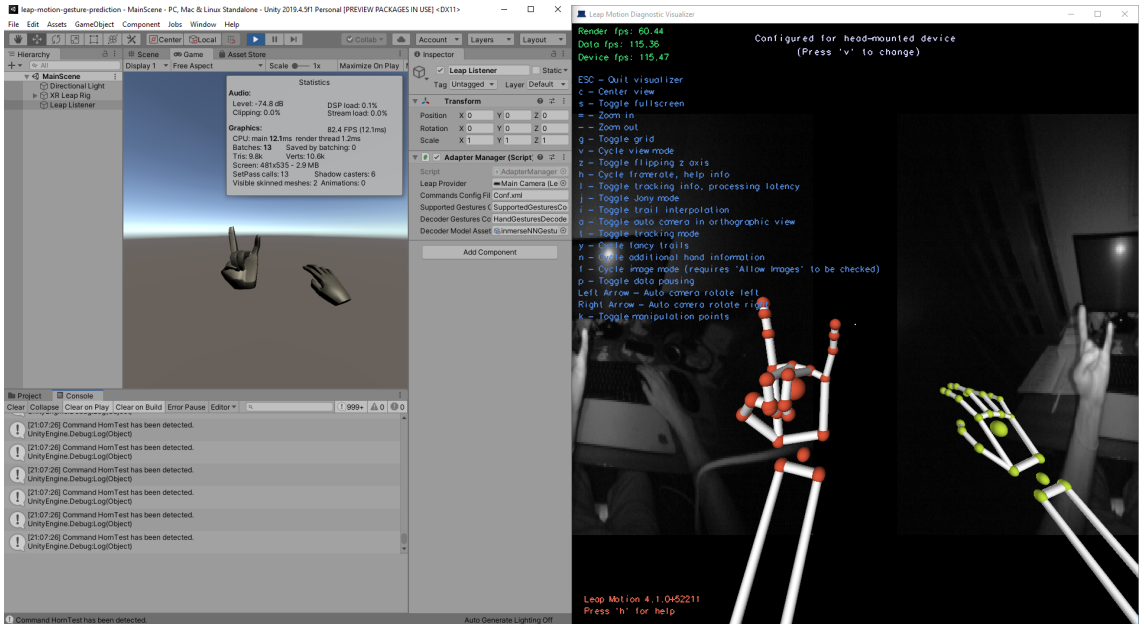


Figure 8.21 – Horn gesture.

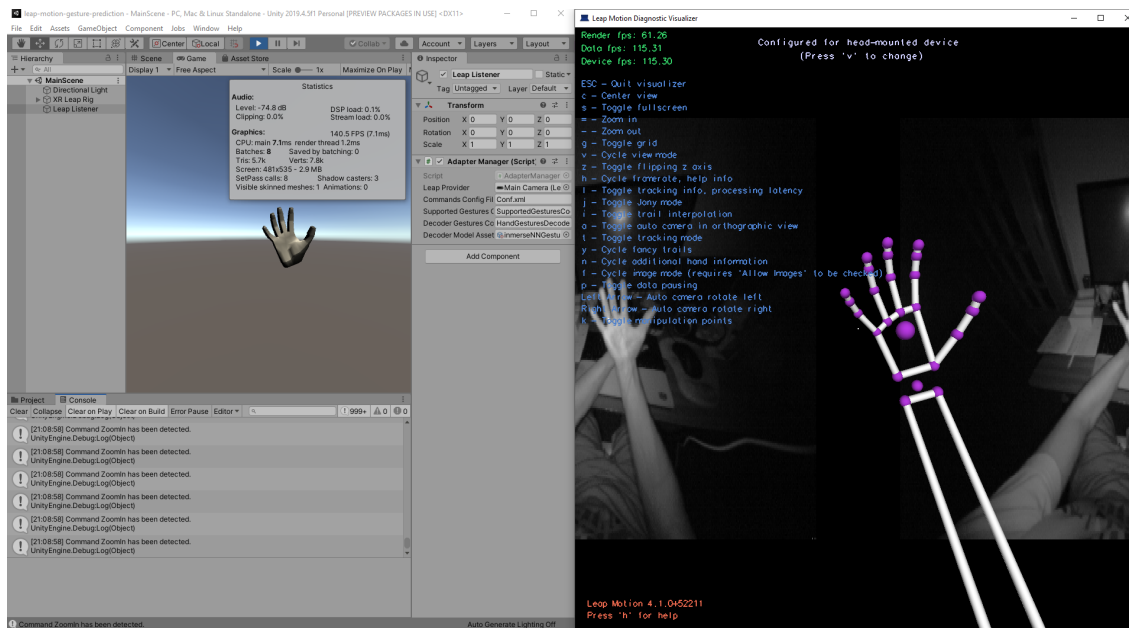


Figure 8.22 – Open gesture.

8.4.3.2 Architecture

One of the goals of this Gestural detection system was to continue using **NNs**, as this would allow us to continue using the **ONNX** format in the same way we do with the Drawing model (*q.v.* Section 8.4.2). We quickly realised that this wasn't going to be an easy task after testing several **NN** models with different architectures. The Portable authors used a **Support Vector Machine (SVM)**, which produces better results with less data. But because we couldn't convert a **SVM** model to **ONNX**, we came up with the idea of mimicking the operation of a **SVM** within a **NN**, which is indeed possible (*q.v.* Appendix D.1, right).

The input shape has one dimension and 30 elements, which correspond to the 30 features discussed above. Two fully connected (dense) layers were used to simulate a hinge model in this model's architecture, which is similar to the **SVM** procedure. The first layer serves as the input layer, with additional kernel initialises such as *he uniform*, which provides variance scaling initialisation by extracting samples from a uniform distribution within the weight tensors' negative and positive bounds. It also has 30 filters and uses **ReLU** as an activation function. The second layer is the

output layer, which contains five classes (*i.e.*, the 5 gestures we want to detect). To avoid overfitting, an L2 kernel regulariser with a value of 0.1 was chosen, along with a Softmax activation function corresponding to the percentage of the projected class, which adds up to 1. Using the hinge model in conjunction with a kernel regulariser causes it to behave similarly to a SVM, which is exactly what we intended.

8.4.3.3 Training and Validation

This model was trained using a categorical hinge loss function and an Adam optimiser with a learning rate of 0.001. The validation was carried out using K-fold validation with 10 folds and a batch size of 10. The sample dataset is divided into K sub-samples for K-fold cross validation, with one sub-sample serving as the testing set and the remaining $K-1$ samples serving as training. Each sub-sample is validated once against the rest of the sub-samples, which serve as the training set. The experiment is repeated K times to test each sub-sample individually. The average accuracy of all iterations is then computed. $K=10$ is a standard practice in K-fold cross-validation, which leads to it being referred to as 10-fold cross-validation. The general procedure in each validation method is to shuffle the dataset, divide it into K groups, and then train the model, evaluate the model, and finally evaluate the models based on the evaluation score for each group. Because the overlap between training sets is reduced with this method, the test error estimates are less correlated and have a higher value of error variance. The model's average resulting accuracy of validation is 99.86% (*q.v.* Figure 8.23), which is considerably high.

```

-----
Score per fold
-----
> Fold 1 - Loss: 0.18867716193199158 - Accuracy: 99.90010261535645%
-----
> Fold 2 - Loss: 0.10445418953895569 - Accuracy: 99.70030188560486%
-----
> Fold 3 - Loss: 0.0718192458152771 - Accuracy: 100.0%
-----
> Fold 4 - Loss: 0.0662582665681839 - Accuracy: 99.90010261535645%
-----
> Fold 5 - Loss: 0.053046803921461105 - Accuracy: 100.0%
-----
> Fold 6 - Loss: 0.04600997641682625 - Accuracy: 99.90000128746033%
-----
> Fold 7 - Loss: 0.04656017944216728 - Accuracy: 99.6999979019165%
-----
> Fold 8 - Loss: 0.03934652358293533 - Accuracy: 100.0%
-----
> Fold 9 - Loss: 0.054226212203502655 - Accuracy: 99.6999979019165%
-----
> Fold 10 - Loss: 0.03324434906244278 - Accuracy: 99.80000257492065%
-----
Average scores for all folds:
> Accuracy: 99.86005067825317 (+- 0.11996709481030252)
> Loss: 0.07036429084837437
-----

```

Figure 8.23 – Gesture detection system’s 10-fold cross-validation’s results.

8.4.3.4 Integration

The whole point of using **NNs** for this problem was to keep using the **ONNX** file format. This means that we can load the model and then get the class predictions using the same Barracuda¹⁹ package (*i.e.*, gesture detected). All of this to ensure that we have a cross-platform solution that supports mobile environments (for example, smartphones and **VR/AR/MR HMDs** that run Android). The real-time detection feels solid, and no noticeable delays were observed during our testing. When compared to the animated concept, the first impression is that it increases immersion and agency. However, in order to effectively validate this **VR** prototype, we conducted some user tests (*q.v.* Chapter 9). We were only able to test the Portuguese culture due to the current **COVID-19** pandemic and quarantine restrictions. We still need to replace the dialogue recordings with Salish before we can test with the tribe’s culture, which will have to wait until the pandemic has passed.

¹⁹ <https://docs.unity3d.com/Packages/com.unity.barracuda@1.0>

9

User Testing

9.1 Introduction

For COVID-19 health security reasons, the user trials were conducted at home. To be clear, this is not a random sample study. Rather, a small group of people (*q.v.* Section 9.2) has been selected for convenience. Each test was carried out in isolation. Only the researcher observing—the author of this document—and the user being tested were present, who was given a random ID to maintain anonymity. Each user completed a socio-demographic questionnaire prior to the actual test (*q.v.* Section 9.3). Additionally, following the method of Carvalho (2019), each user was asked to choose which gestures/somatic movements they would use to represent a set of predetermined general actions (*q.v.* Table 9.2). The answers are used to expand the set of gestures that is detected by the experience. To be clear, no hint was given to the actual use of gestures or somatic movements. With the user’s explicit consent, the researcher observed and recorded the user trials. Following the test, each user was interviewed using the open-ended questions listed below:

- **Q1** - What are your first impressions of the experience?

- **Q2** - Was there any situation mid-experience during which you didn't know/understand what you were supposed to do?
 - **Q2.1** - In that situation, what did you think you had to do? Or, what led you to think otherwise?
- **Q3** - What observations can you make about what you learnt from the experience?
- **Q4** - What can you tell me about the birds? What did they do? What happened?
- **Q5** - What can you tell me about the tribes?
- **Q6** - What did you think about the dialogues being in English? Did it affect your experience in any way?
- **Q7** - What did you think about the exploration phase?

The results of the observations (*q.v.* Section 9.4) and their analysis (*q.v.* Section 9.5) will be presented throughout this chapter. It concludes with a summary of the key findings (*q.v.* Section 9.6).

9.2 User Sample Distribution

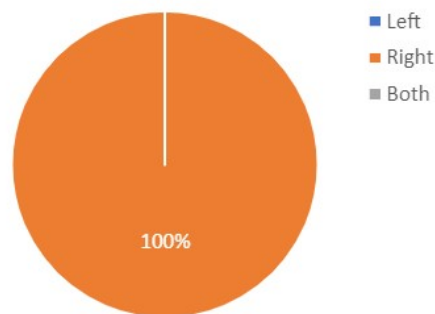
The trials were carried out with a non-random sample of ten people from the same Portuguese cultural group. Tests with the CSKT were not yet possible due to current COVID-19 restrictions. That being said, for the time being, we conducted user tests with a limited number of people, chosen purely for convenience, as no prior knowledge is required. They are all of Portuguese descent and range in age from 17 to 57 years old. Gender distribution was as follows: 60% (6) were female, 40% (4) were male (*q.v.* Table 9.1).

Table 9.1 – General sample's description.

User ID	Gender	Age	Cultural Group
1	Male	20	Portuguese
2	Female	23	Portuguese
3	Female	56	Portuguese
4	Male	57	Portuguese
5	Male	30	Portuguese
6	Female	50	Portuguese
7	Male	50	Portuguese
8	Female	21	Portuguese
9	Female	17	Portuguese
10	Female	28	Portuguese

9.3 Pretest

As previously stated, all users completed a socio-demographic questionnaire prior to the actual experience. Other, more experience-related questions were asked in addition to more general data (*q.v.* Table 9.1). We attempted to understand the users' conditions and habits prior to conducting the actual tests. For example, using the same socio-demographic form, we ask what their dominant hand is (*q.v.* Figure 9.1), as well as whether they have any vision (*q.v.* Figure 9.2, and 9.3) or auditory problems (and their specificities, if known), *q.v.* Figure 9.4. Because all of these have the potential to influence the end-user experience, we consider them to be more than minor details.

**Figure 9.1** – User Trials dominant hand.

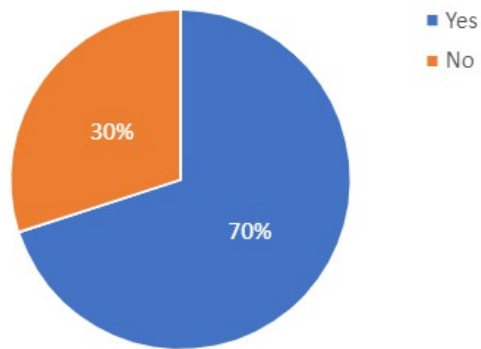


Figure 9.2 – User Trials - vision issues.

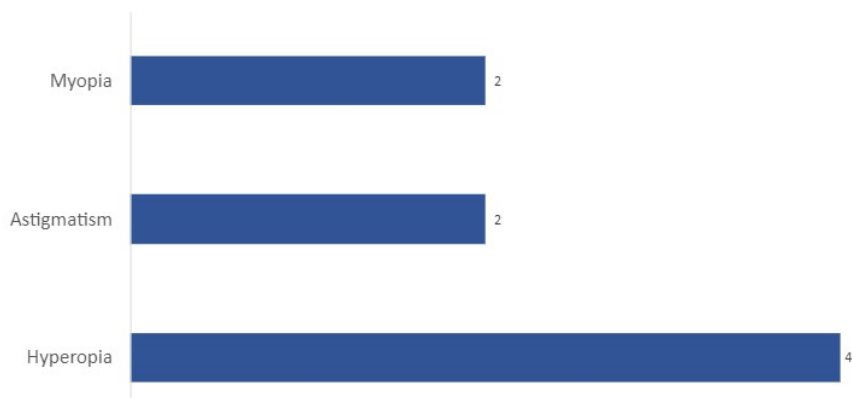


Figure 9.3 – User Trials - vision specific issues.

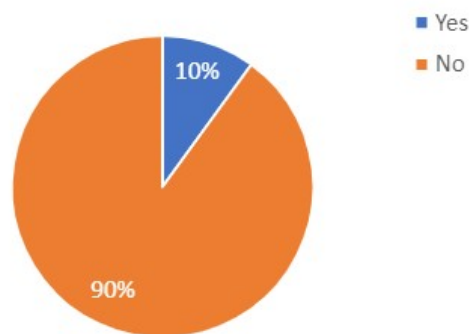


Figure 9.4 – User Trials - Auditory issues.

Despite the fact that we do not require any specific knowledge from the users, we do inquire about their education. With the exception of one user who is still in high

school, all of them have or are pursuing a higher education degree (*q.v.* Figure 9.5).

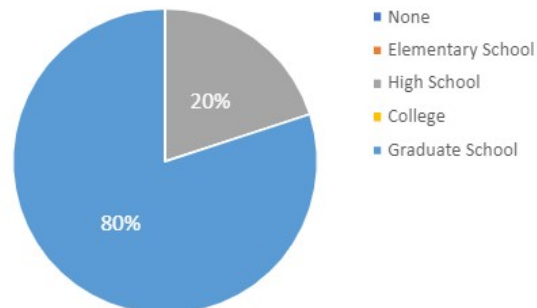


Figure 9.5 – User Trials - Education level.

They were also questioned about their prior experience with immersive environments and specific HMDs (*q.v.* Figures 9.6, 9.7, and 9.8).

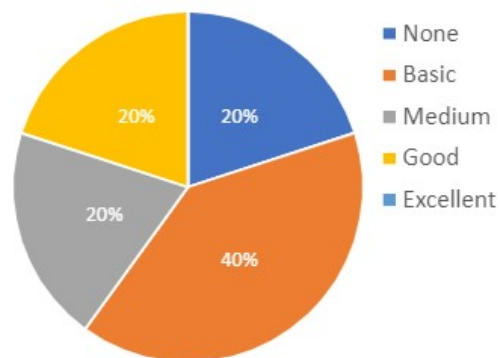


Figure 9.6 – User Trials - Prior experience with immersive technology.

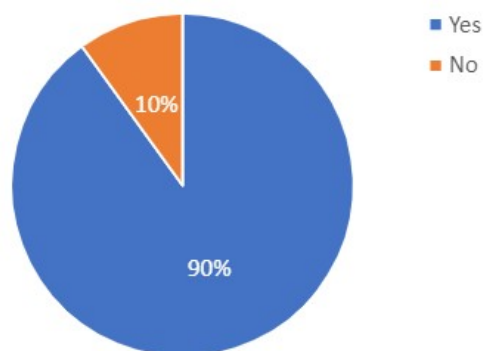


Figure 9.7 – User Trials - used Immersive technology before.

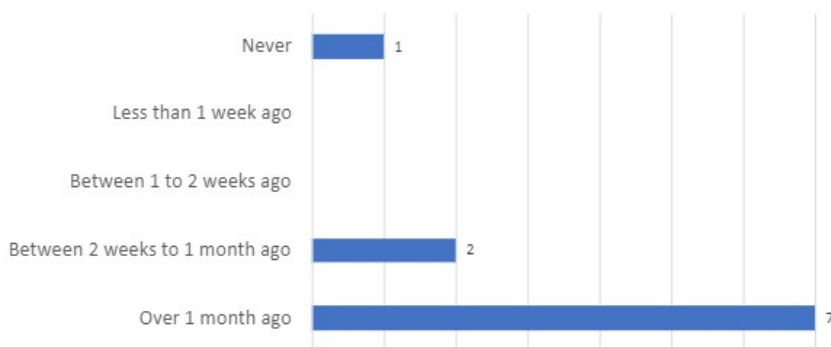


Figure 9.8 – User Trials - last time since the use of immersive HMDs.

Almost no one had ever used immersive technology before. Even if they did, the majority (7) did so more than a month ago (*q.v.* Figure 9.8).

After completing the form, all users were asked to represent actions such as grabbing, resizing, and flying with gestures, body or head movements, all of which corresponded to a specific Task or Action of the experience. This information, as well as hints, were not provided to the users. The researcher would observe and record how they did it. This was captured with a camera, while the hand movements were captured with the Leap Motion. The responses were compiled into a single form, with the number of occurrences assigned to each (*q.v.* Table 9.2).

Table 9.2 – Compiled pre-form with corresponding number of occurrences.

Pre-form User's Action Gestures and/or somatic movements		
Task or Action ID	Action meaning	Answers (observational)
T1	Grab	One hand closing (5); Both hands closing (3); Stretch the arm fully and close hand (1); Both hands facing with the palm upwards, and closing together (1).

Table 9.2 continued from previous page

Pre-form User's Action Gestures and/or somatic movements		
T4.1 and T5.1	Change (from one picture to another)	Look at one, and then the other (3); Walk to one and then the other (1); Full hand swipe (2); Index finger from right to left (sort of swipe) (1). Grab them, in turns, and switch their place (2); Grab them at the same time, and switch their place (1).
T6.1 and T9	Resize	Pinch (index and thumb fingers) (5); Both hands in and out (from the inside to the outside, and vice versa) (2); Hold it and stretch it to make it bigger, and smash it with the palm to make it smaller (1); Position the object further away to make it smaller, and vice versa (1); Walk away to make it smaller, and vice versa (1).
T7	Hold (a pen)	Hold it with the index, middle, and thumb fingers (8); Hold it with the index, and thumb fingers (1); Hold it with the index, middle, ring, and thumb fingers (1); 2 of the users removed the cap of an actual pen to show the movement. 1 of the users flipped the pen onto the right position.
T7.1	Draw	Same position as when holding a pen, but with motion (10); 1 user mentioned that if it was a bigger tool, such as a big brush, he would use the whole palm to hold it, instead.

Table 9.2 continued from previous page

Pre-form User's Action Gestures and/or somatic movements		
T8	Grow	Stretch back and forth the upper limbs (2); Two hands lifting at the same time (2); One hand lifting (1); Water it (1); Moving it farther away to make it smaller, and vice versa (1); Pinch (index and thumb fingers) with both hands (in and out) (1); Pinch with one hand moving hand upwards (1); Feed him (if alive), or, walk away to make it smaller, and vice versa (for inanimate objects) (1).
T10	Fly	Pinch gesture (index and thumb) and throw as it was a paper plane (1); Both hands interconnected to make a butterfly motion (1); Throw it frontwards (3); Throw it frontwards, or "carefully lift" it (like a bird) (1); Throw it upwards (with the palm) (1); Place some wings on its back, while holding the object (1); Draw some wings (1); Two hands moving as if they were wings (1).
T12, T13, and T14	Yes	Nodding up and down (5); Thumbs up (1); Aloud, always accompanied with nodding up and down (4);
T12, T13, and T14	No	Nodding left and right (4); Move index finger left and right (1); Nodding left and right with the eyes moving, too (1); Aloud accompanied with nodding left or right (3); Aloud with slight upwards movement of the shoulders (1);

Table 9.2 continued from previous page

Pre-form User's Action Gestures and/or somatic movements			
A19	Change eral)	(gen-	Fingers snap (1); Rubbing palms as if something doesn't exist anymore (1); Left hand move to the left side for past, and right hand move to the right side for future (1); Line in front with left hand side being past, middle being the present, and right hand side being the future (1); Flip of right hand, starting with palm up position (1); Hand to the back to refer to the past, and hand to the front to refer to the future (3); Today as both hands with palm facing upwards in front, Thumb finger to the back to refer to the past, and and standard 'change' to refer to the future (1); Thumb back to for past, palm facing face for present, and flipping hand forwards for future (1).

We would test the **HMD**, in this case an Oculus Quest, on each user after they completed the socio-demographic form and answered the follow-up actions questions compiled above (*q.v.* Table compiled pre-test). Before beginning the experience, we make sure they can read the subtitles, hear an example dialogue, and are comfortable. We adjusted the headset straps for each user, but, perhaps surprisingly, none of the users required manual adjustment of the **Interpupillary Distance (IPD)** (with 63mm being the default value). Because it is a stationary **VR** experience, we attempted to position the chair where they would sit so that they would not collide with any of the furniture around them. To ensure that, they were asked to fully extend their hands at an angle of 180 degrees (front and sides). Finally, users were told that if they felt any discomfort during the experience, they could simply say so and we would stop the experience. They were also asked if the experience could be fully recorded, to which all users verbally consented in addition to providing written consent.

9.4 Task Execution and Observation

A list of tasks and corresponding default gestures (for both the Portuguese and Salish cultures), *q.v.* Table 9.3, is provided below. Tasks are mandatory to progress in the experience and its story. Actions represent possible actions at specific times that can be done or not done depending on the progress of the story and have no effect on the actual progress. As the VR prototype was designed for default gestures only, with no guarantees of detection or use by users, 3-4 seconds were allotted per task for the user to perform some kind of action. If the action did not occur, the experience would continue normally.

Table 9.3 – Tasks and actions, and corresponding default gestures.

Task or Action ID	Task or Action	Portuguese Culture	Cul- ture	Salish Culture
A1	Look around			
A2	Hello	Wave		Wave
T1	Grab Micro-phone	Close hand in a fist		Open hand slightly
T1.1	Talk			
A3	Give Micro-phone	Drop/throw phone back	micro-	Drop/Throw micro-phone back
T2	Look at the V-shaped peak			
T3	Look back at narrator			
T4	Look at reservation pictures			

Table 9.3 continued from previous page

Task or Ac- tion ID	Task or Action	Portuguese Cul- ture	Salish Culture
T4.1	Interact with pictures	Swipe left or right	Swipe left or right
A4	Look at car and bear demonstration		
T5	Look at art pictures		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right
A5	Look at coyote		
T6	Look at Luli's book cover		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)
T7.1	Draw bird	Drawing motion	Drawing motion
A6	Look upwards when narrator points to the sun		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in/out with both hands

Table 9.3 continued from previous page

Task or Ac- tion ID	Task or Action	Portuguese ture	Cul- Salish Culture
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below
T10.1	Look at the Geese in the sky		
A7	Look at Luli at the beach		
T11	Look at light leading to the sky		
T11.1	Look at Geese flying back		
A8	Look at Geese and bear in lake		
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down
A9	Look at the sun and planets models		
A10	Look at the earth globe		

Table 9.3 continued from previous page

Task or Ac- tion ID	Task or Action	Portuguese culture	Cul- Salish Culture
A11	Look at the animals dis- appearing (hibernating)		
A12	Look at the mountains snow coming down like a blanket		
T13	Yes/No choice for further info on geese migra- tions	Ring with index and thumb; Nodding	Nodding, or thumbs up/down
A13	Look at Geese flying at differ- ent speeds		
A14	Look at goose going out of for- mation and re- covering		
A15	Look at the Geese forma- tion's leader rotating back when it gets tired		

Table 9.3 continued from previous page

Task or Ac- tion ID	Task or Action	Portuguese ture	Cul- Salish Culture
A16	Look at the Geese flying and motivating the ones in front to continue		
A17	Look at a goose falling and two others staying with him until he recovers		
A18	Look at the Geese joining back the same flock		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle
A20	Goodbye	Wave	Praying or Wave Ges- ture

9.5 Data Analysis

A smartphone camera was used to fully capture each user's experience. Following the completion of all trials, the videos were analysed and used to fill out a form for each user (*q.v.* Appendix E, Subsection E.1) that details which gesture, somatic movement, or voice (if any) was used on each predetermined Task or Action (*q.v.* Table 9.3), and whether they used what they said they would (*q.v.* Table 9.2).

All tasks and actions (*q.v.* Table 9.3) were assigned a three variable code, with each variable having the possibility of being Y (Yes) or N (No), and each having a different meaning:

- The leftmost column indicates whether the user's gesture was expected;
- The middle column indicates whether or not an action was carried out;
- The rightmost column indicates whether the gesture (if any) corresponds to the respective Pretest Answer.

When they are combined, they produce different interpretations:

- **YN** indicates that an expected action was not carried out. And there was no Pretest answer (such as looking or waving);
- **YNN** indicates that an expected action was performed, but it did not match the corresponding Pretest answer;
- **YNY** indicates that an expected action was not carried out, despite the presence of a Pretest answer;
- **YYY** indicates that an expected action matches the one performed by the user, and the corresponding Pretest answer. **NNN** is the inverse of this;
- **YYN** indicates that an expected action was performed by the user, but it did not match the corresponding Pretest answer;

- **NYN** indicates that there was an expected action that was not performed by the user, but it also did not correspond to the corresponding Pretest answer. This is not to say that the user interacted when they were not supposed to;
- **NYN** indicates that an expected action was not performed by the user, but, rather, that it matched the corresponding Pretest answer.

NNY does not exist because no Pretest for non-existent and/or expected actions was created. Cases of **YNY** and **NYN** were assigned the colour yellow because they both contradict what was in the Pretest and the predetermined defaults. Green was given when the interaction occurred and the outcome was either expected or matched their response (or both). When an interaction was expected but the user did not respond, the colour red was assigned. The observation compiled tables for each of the users can be seen in Appendix E.

All of the observation data was then compiled into a single table to determine how all of the users handled each of the predetermined tasks and/or actions collectively (*q.v.* Figures 9.9 to 9.30). Appendix E.2 depicts some of the, not as important, non-mandatory actions (mostly looking somewhere).

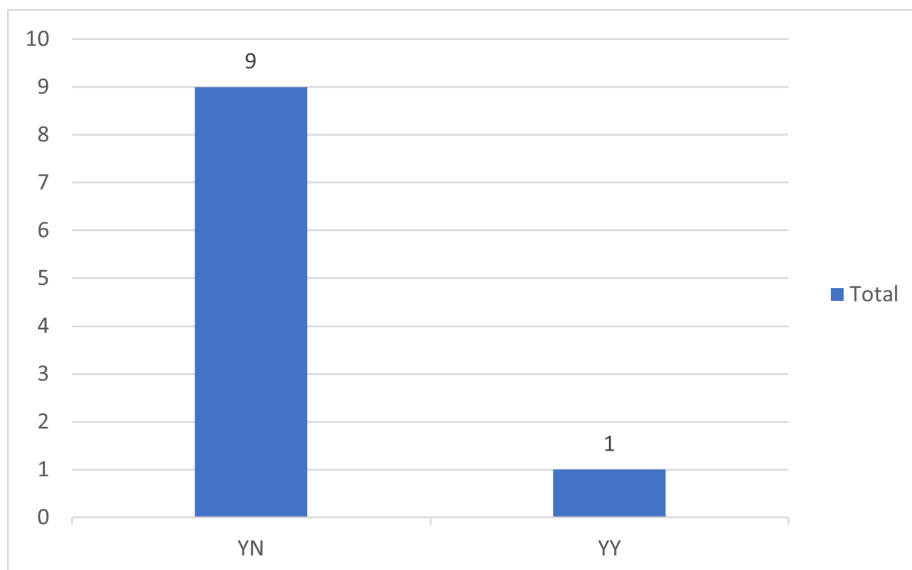


Figure 9.9 – Action A1 - Look around.

The main result for **A1** (*q.v.* Figure 9.9), which was about entering the experience and the possibility of looking around to explore, was that only one of the ten users looked around. The majority of people did not.

Following a similar pattern, **A2** (*q.v.* Figure 9.10), Waving for hello, was not performed by any of the users.

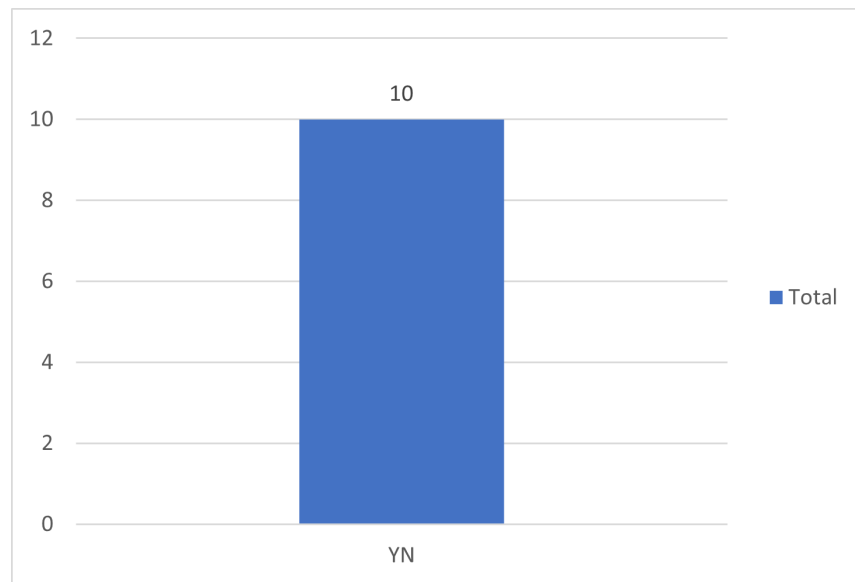


Figure 9.10 – Action A2 - Wave for Hello.

T1 (*q.v.* Figure 9.11) performed a little differently, causing the trend to shift. Six of the ten users grabbed the microphone, which not only corresponded to the predetermined gesture, but also to what they answered (**YYY**). Two of the users grabbed because it was predetermined, but it did not correspond to their answer. And, despite the Pretest corresponding answer, only two of the users did not take any action (**YNN**).

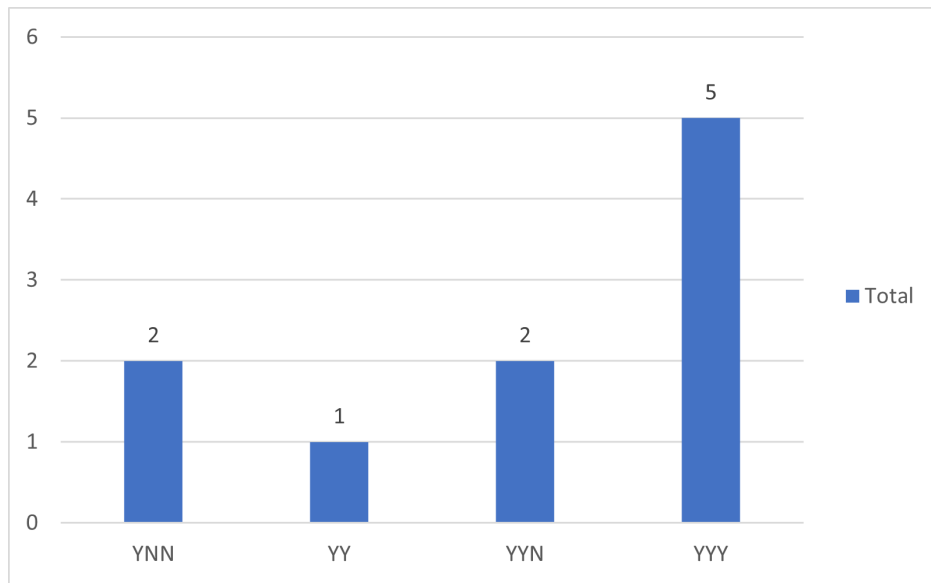


Figure 9.11 – Task T1 - Grab Microphone.

The results for the sub task **T1.1** (*q.v.* Figure 9.12) were completely split. The first five interacted as expected (saying their name), while the remaining five did not.

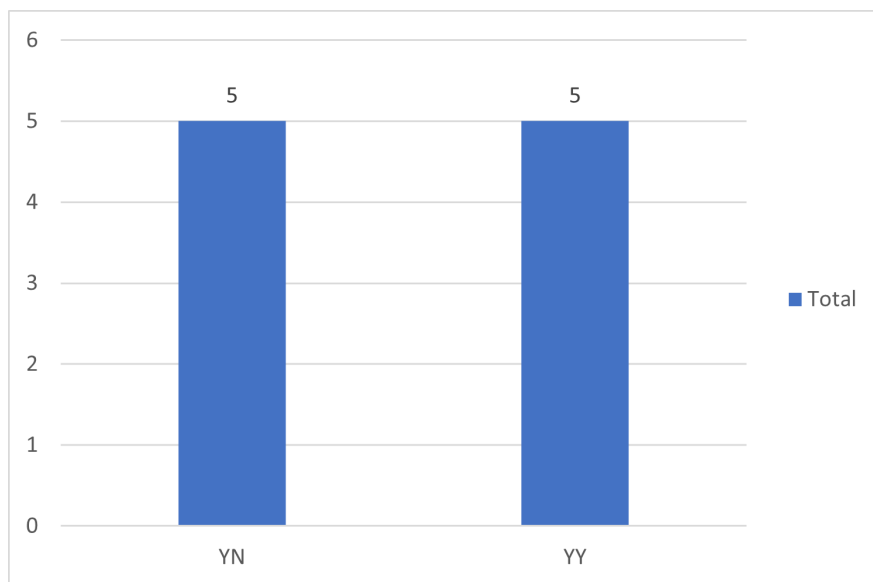


Figure 9.12 – Task T1.1 - Talk.

For **A3** (*q.v.* Figure 9.13), what happened with the microphone, 8 out of 10 users dropped it, while 2 users did nothing.

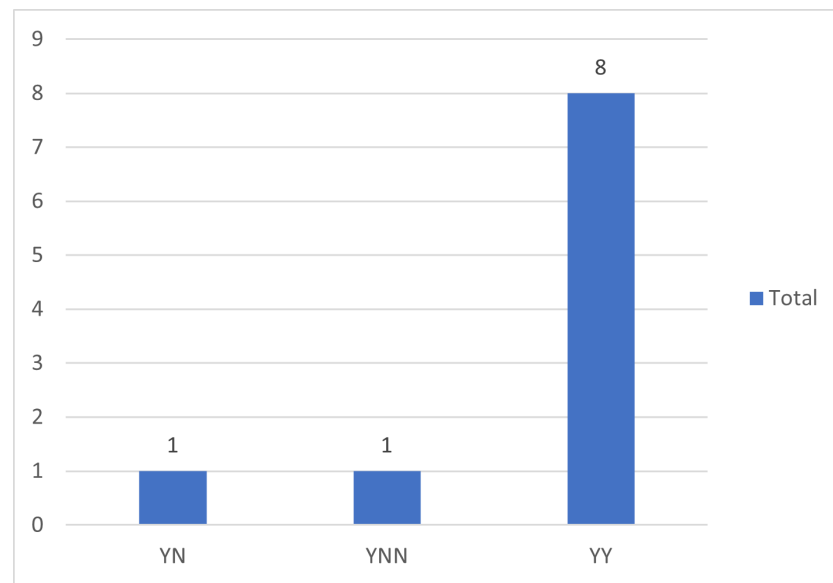


Figure 9.13 – Action 3 - Give Microphone.

The results of **T2** (*q.v.* Figure 9.14) show that everyone looked at the V-shaped peak as intended (**YY**). The narrator's pointing gesture appears to have been effective.

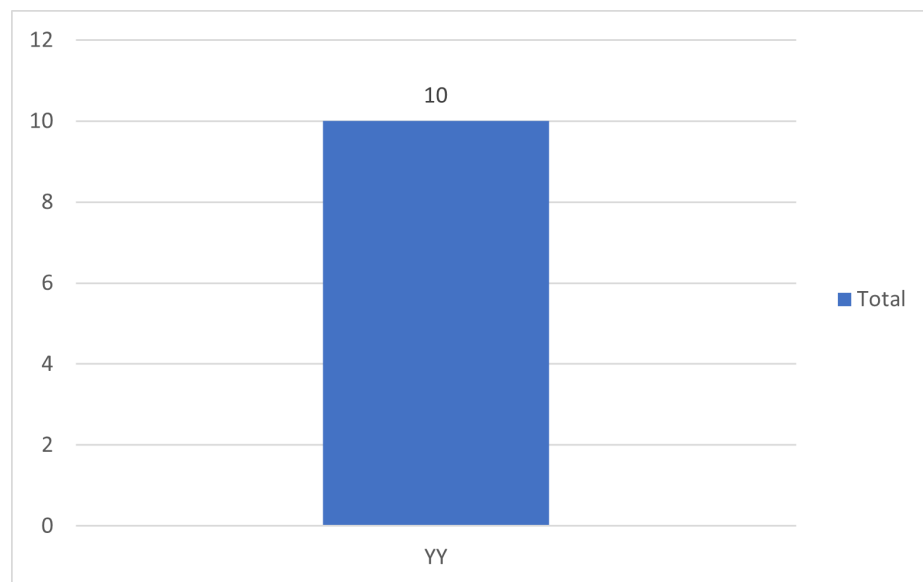


Figure 9.14 – Task T2 - Look at V-shaped peak.

T3 (*q.v.* Figure 9.15) shows that after looking at the mountain's peak, the majority of the users (9) looked back at the narrator, rather than waiting for the next dialogue

to begin.

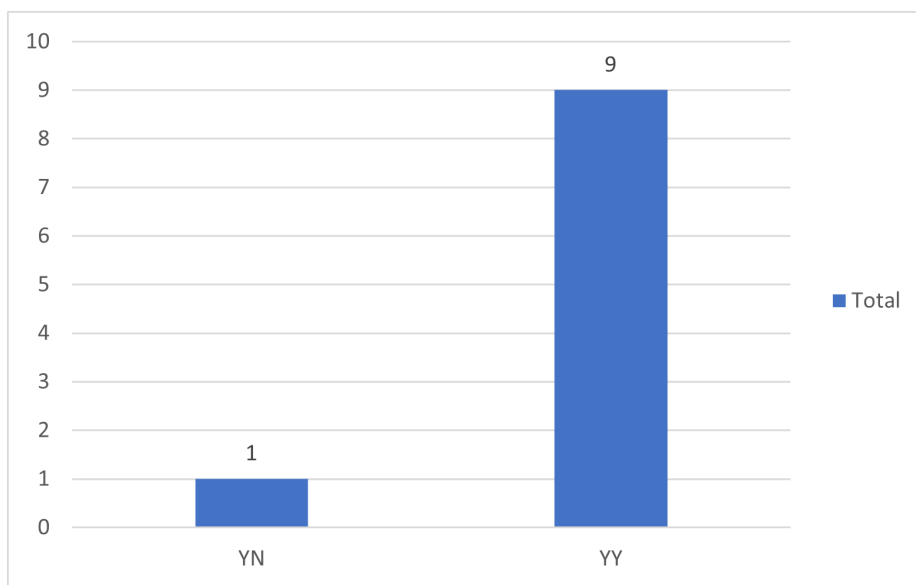


Figure 9.15 – Task T3 - Look back at narrator.

T4.1 (*q.v.* Figure 9.16) demonstrates that the majority of people interacted with the pictures, either agreeing with the standard gestures (**YYN**) or completely agreeing with them (**YYY**). One user interacted in a way that was neither standard nor corresponding to the answer (**NYN**), whereas two users did not interact at all (**YNN**).

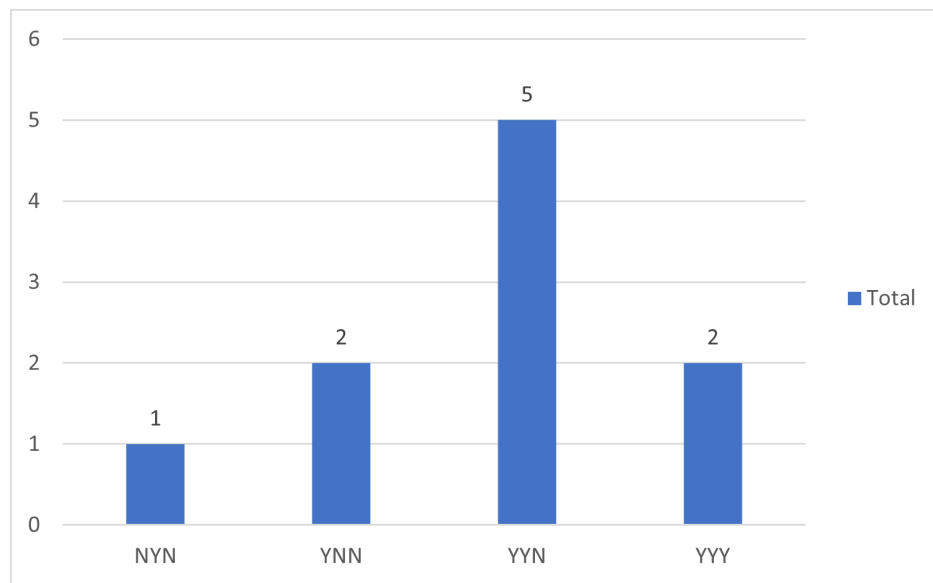


Figure 9.16 – Task T4.1 - Interact with pictures.

T5.1 (*q.v.* Figure 9.17) exhibits the same pattern as **T4.1**, above. Four users agreed with the standard gesture but not with their corresponding answer (**YYN**), while two users completely matched expectations (**YYY**). One user unexpectedly interacted (**NYN**), and another agreed with his corresponding response (**NYY**). Two users did not interact with the images in any way (**YNN**).

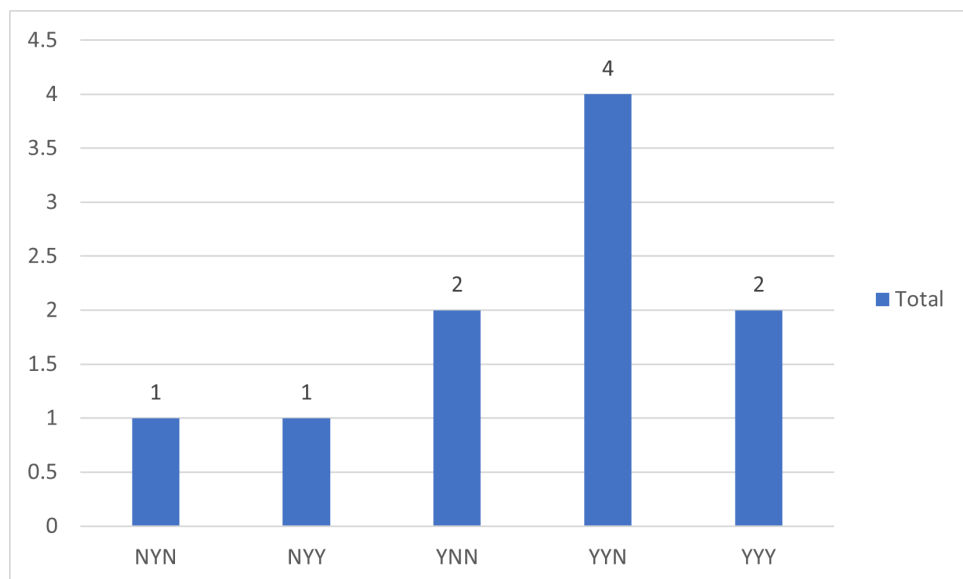


Figure 9.17 – Task T5.1 - Interact with pictures.

The interaction with Luli's book (*q.v.* Figure 9.18), on the other hand, was minimal. Only one matched the standard gesture (**YYN**), and the remaining two matched their corresponding answers (**NYY**). Three users interacted in an unexpected way, while four others did not (**YNN** and **YNY**)

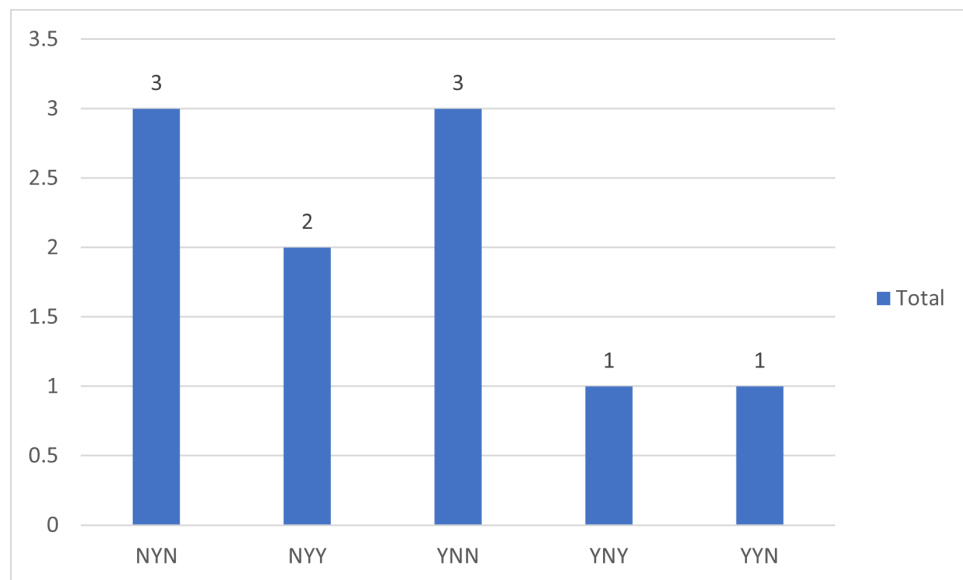


Figure 9.18 – Task T6.1 - Resize Luli's book.

T7 (*q.v.* Figure 9.19) depicts almost everyone (9 out of 10) holding the brush while answering the questions. One user did not interact with it.

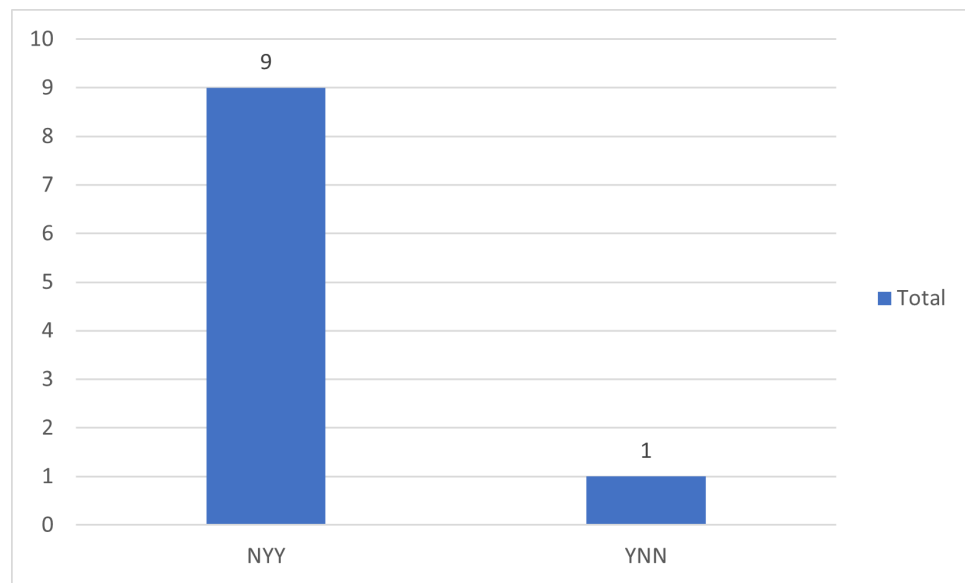


Figure 9.19 – Task T7 - Hold brush.

Almost everyone (8) drew a bird (*q.v.* Figure 9.20) as intended and matched their corresponding answers (**YYY**), with the exception of one who did not interact at all (**YNN**), and one who interacted unexpectedly (**NYN**).

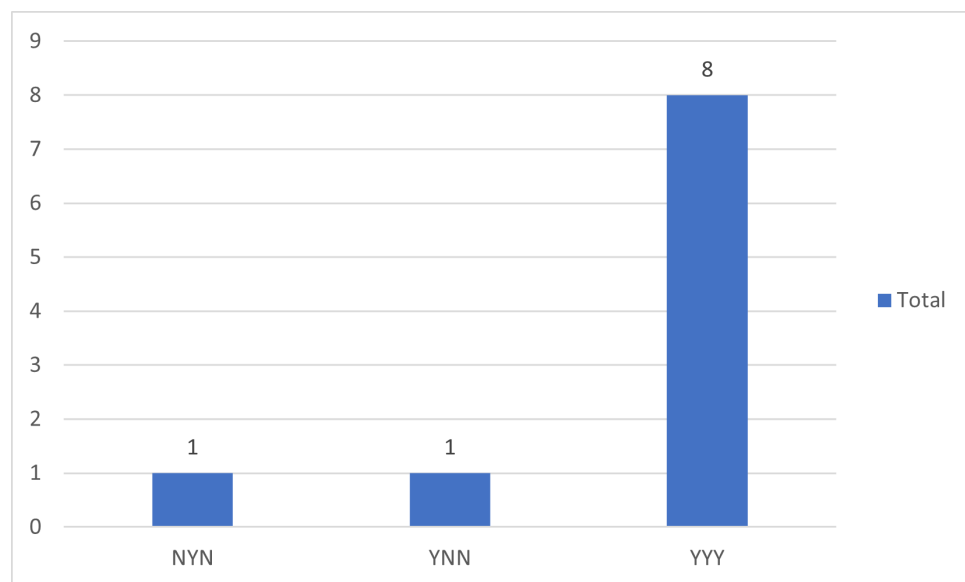


Figure 9.20 – Task T7.1 - Draw bird.

T8 (*q.v.* Figure 9.21) depicts four users interacting in an unexpected manner

(**NYN**), as well as four others interacting with either the standard gesture or the corresponding pretest answer (**NYN** and **YYN**). Two users did not interact (**YNN**).

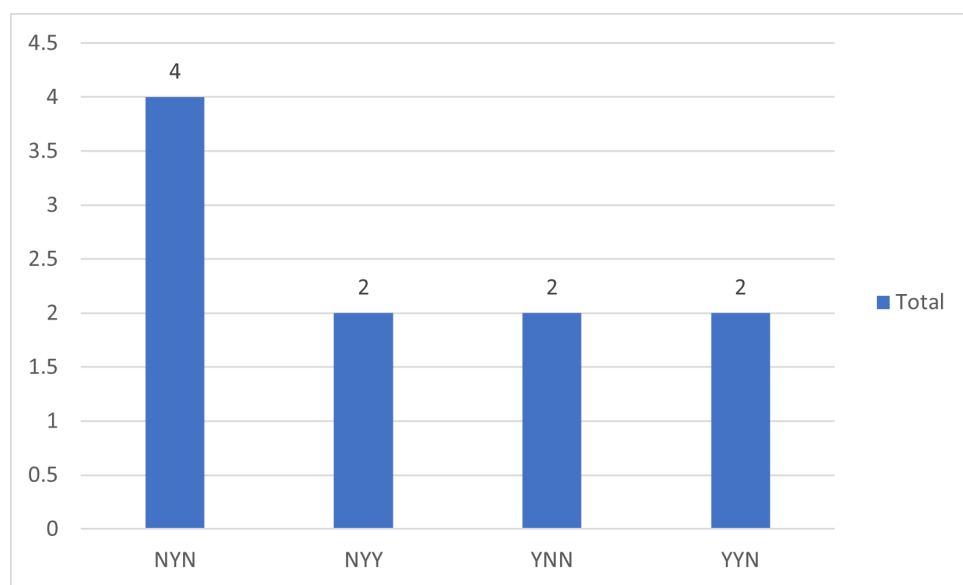


Figure 9.21 – Task T8 - Make Luli grow.

T9 (*q.v.* Figure 9.22) shows that half of the users (5) do not interact at all (**YNN**), three match the standard action or answer (**YYN** and **YYY**), and two users behave unexpectedly (**NYN**).

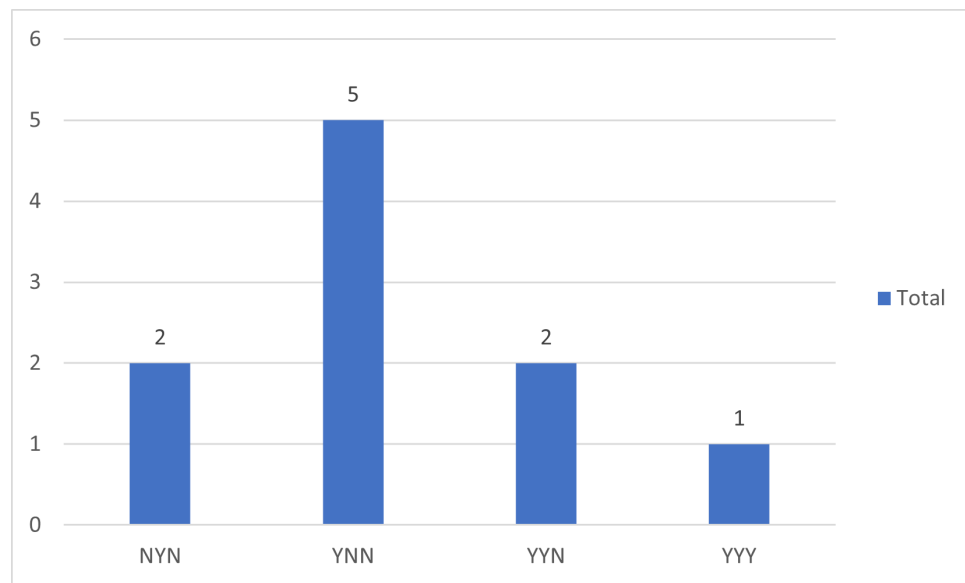


Figure 9.22 – Task T9 - Adjust Map.

T10 (*q.v.* Figure 9.23) depicts four users interacting based on their corresponding answer (**NYY**), and two others interacting based on the default action (**YYN**). Three users did not interact (**YNN**), and one acted unusually (**NYN**).

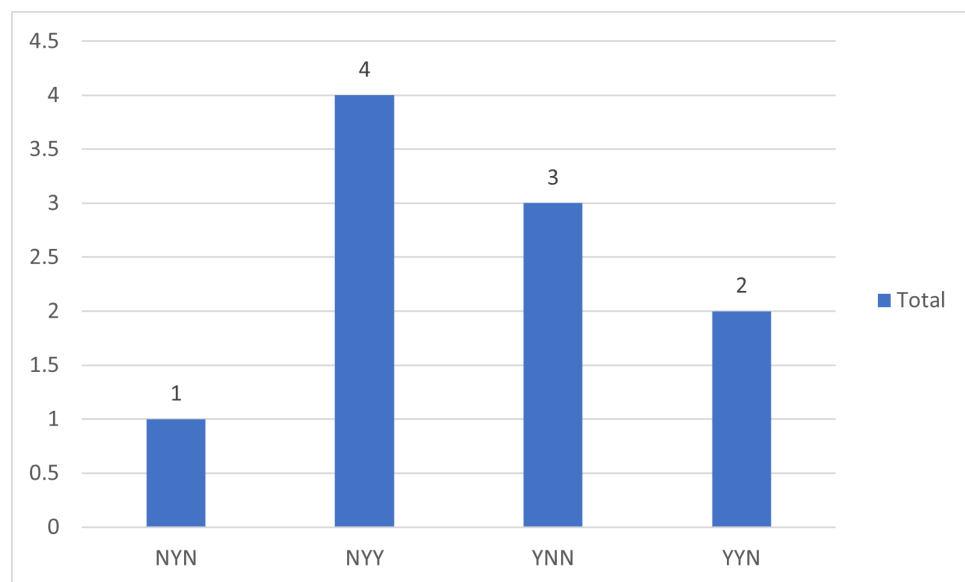


Figure 9.23 – Task T10 - Make Luli fly.

T10.1 (*q.v.* Figure 9.24) depicts eight people looking up at the sky as intended

(YY), with only two users not looking up (YN).

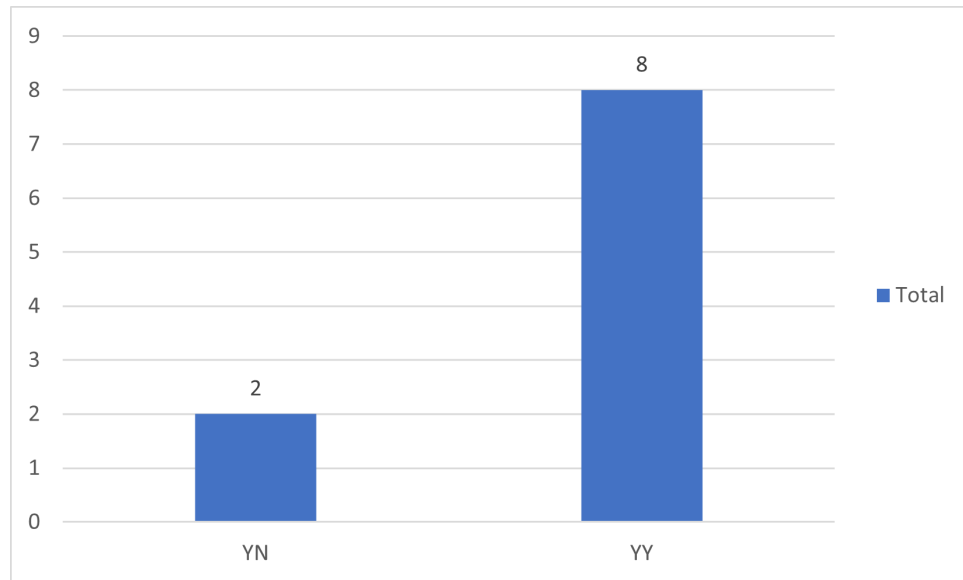


Figure 9.24 – Task T10.1 - Look at the Geese in the sky.

T11 (*q.v.* Figure 9.25) depicts seven users following the light trail leading to the sky (YY), with the remaining three ignoring it (YN).

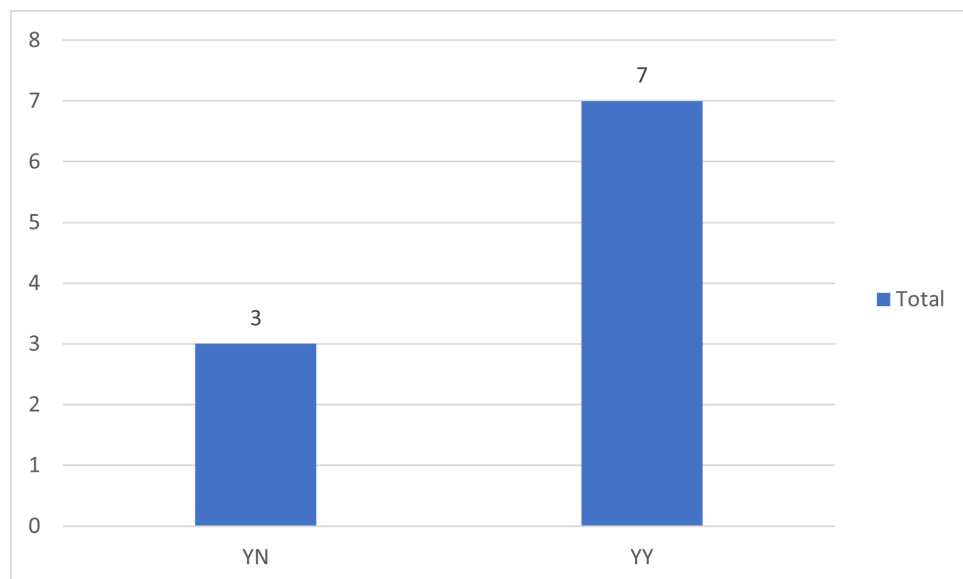


Figure 9.25 – Task T11 - Look at light leading to the sky.

T11.1 (*q.v.* Figure 9.26) depicts one more user than **T10.1** (*q.v.* Figure 9.24) looking up at the sky (**YY**), with one still missing (**YN**).

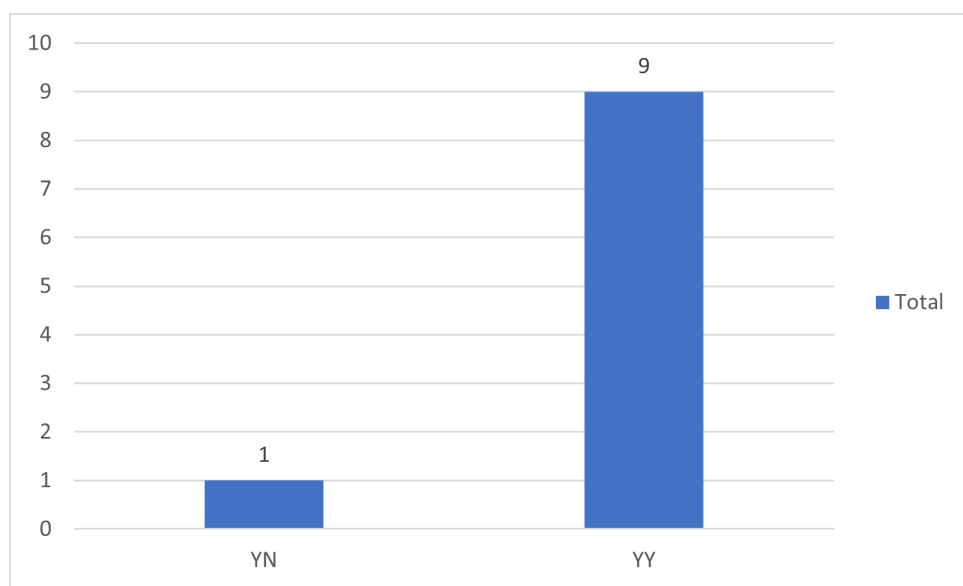


Figure 9.26 – Task T11.1 - Look at Geese flying back.

T12 (*q.v.* Figure 9.27) depicts seven people who are completely in agreement when answering a yes or no question (**YYY**), with one matching the standard action but not his corresponding answer (**YYN**). Unexpected interactions occurred with two users (**NYN**).

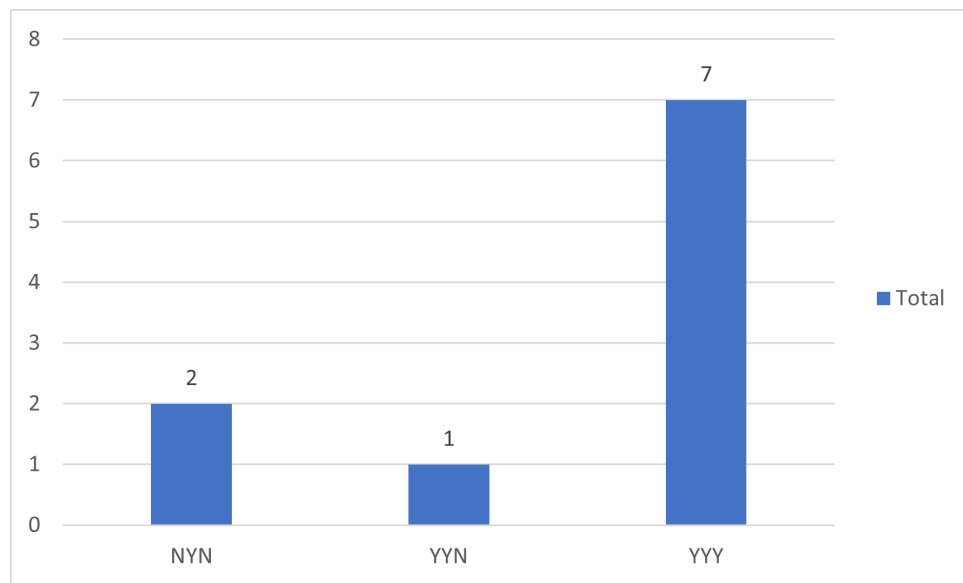


Figure 9.27 – Tasks T12 - Yes/No Answer for more information on tribes.

T13 and **T14** produce identical results (*q.v.* Figure 9.28), with eight users completely agreeing (YYY)—one more than **T12**. One user still matches the standard gesture/action, whereas one less user behaved unexpectedly when compared to **T12** (NYN).

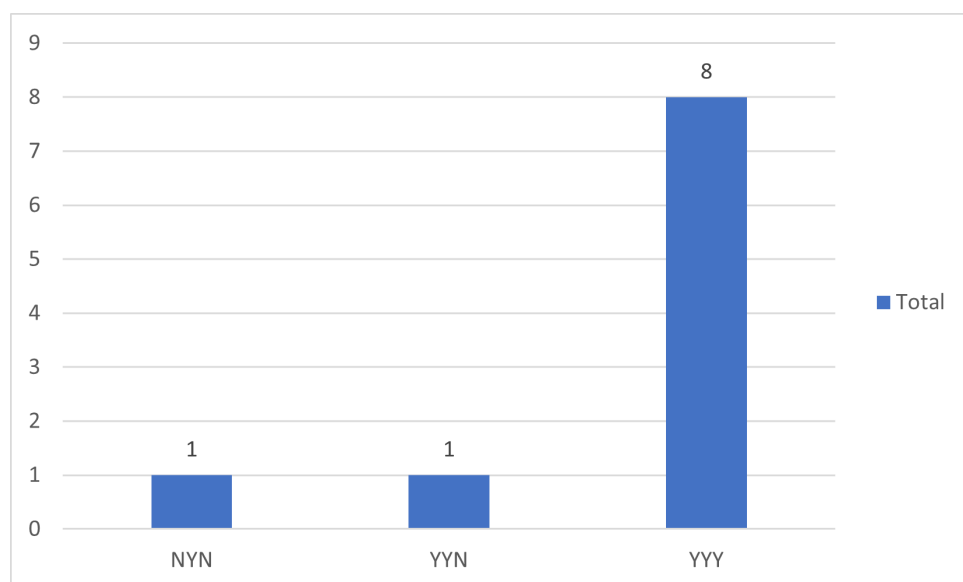


Figure 9.28 – Tasks T13 and T14 - Yes/No Answer.

A19 (*q.v.* Figure 9.29) shows two users fully agreeing with the predetermined change gestures (**YYY**), four agreeing with their corresponding answers (**NYY**), and one interacting unexpectedly (**NYN**). Three of the users did not perform a 'change' in the environment (**YNN**).

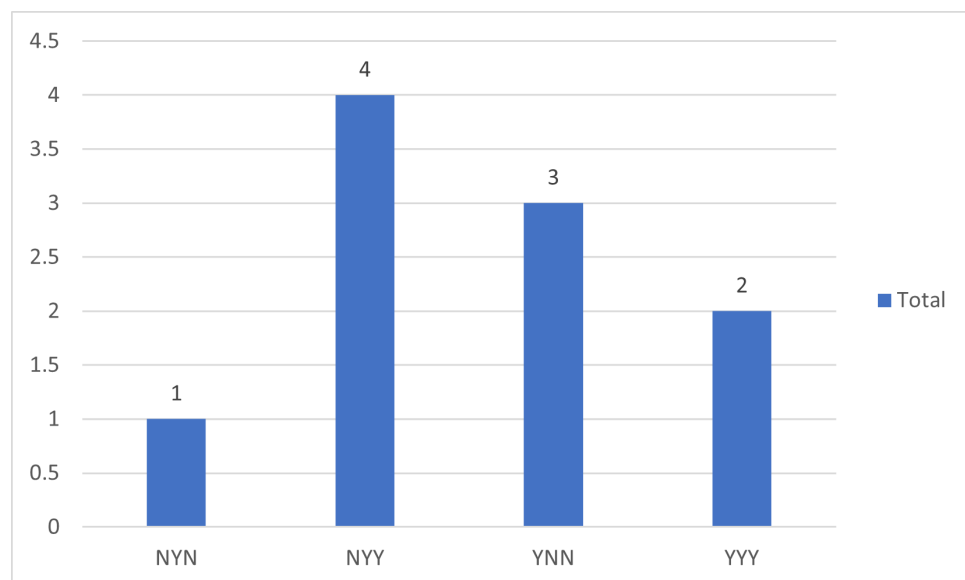


Figure 9.29 – Action 19 - Exploration phase with 'change'.

A20 demonstrates that, despite the fact that no users waved hello at the start of the experience (**A2**, *q.v.* Figure 9.10), six of them did wave goodbye at the end (**YY**), with four users still not interacting (**YN**).

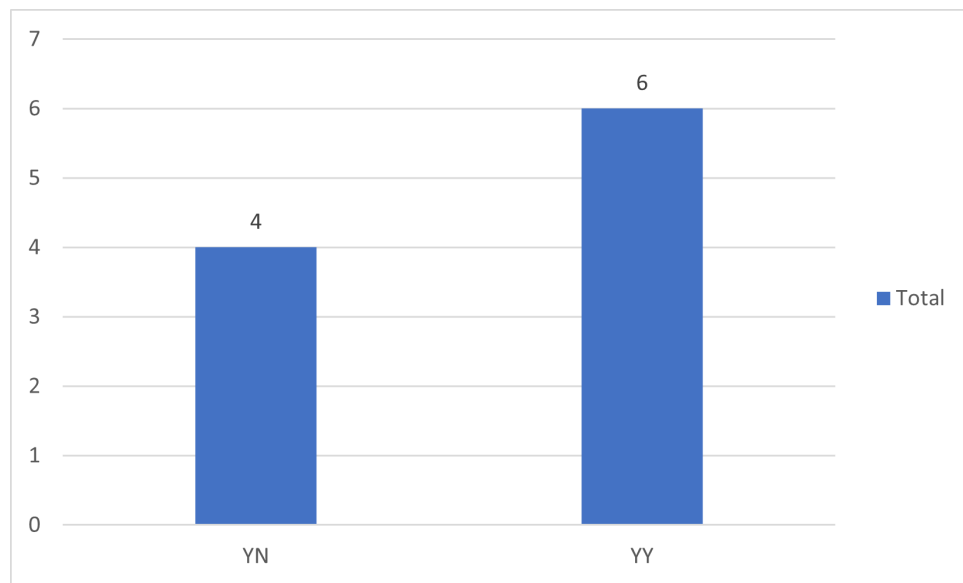


Figure 9.30 – Action 20 - Goodbye Wave.

9.6 Main Findings

Overall, the majority of users interacted a lot during the experience and found it enjoyable and interesting. User 4 in particular did not do much, though he did draw and perform some other actions, presumably because he did not realise from the start that he was being interacted with by the narrator. Similarly, User 9 lacked some interactions, presumably due to a lack of knowledge about the possible interactions. It is important to remember that the users had no prior experience with the setup, so that the experience was not skewed in any way. Our goal was to figure out how they would interact intuitively if they had no prior experience. After all, we're hoping to capitalise on the experiences of those who aren't as computer-savvy. However, one can return to the socio-demographic results to try to understand how they self-evaluated their computer expertise (*q.v.* Figure 9.31).

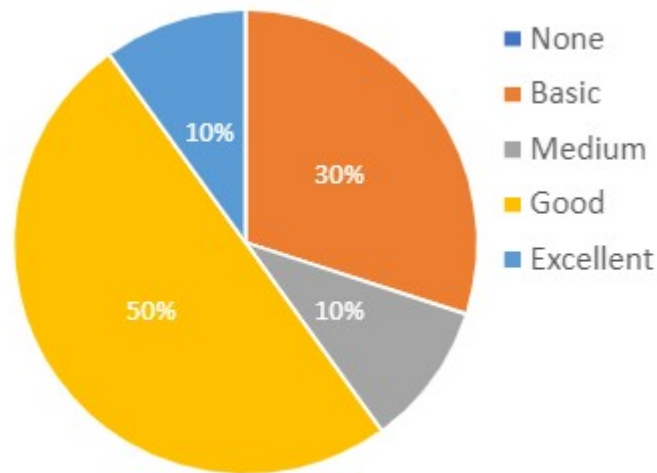


Figure 9.31 – Users' computer experience.

Only one user claims to have excellent computer skills, while five others claim to have a good level. Three are basic, and one is medium. One could argue that this demonstrates all levels of computer expertise interacting with the immersive experience in some way. Despite the fact that nine out of ten users claim to have tried immersive technology before, their actual experience *q.v.* Figure 9.6) shows only two users claiming a good AR experience (via Instagram and Pokemon Go), with the rest rating it as basic or less. Again, this appears to indicate that the VR Prototype (*q.v.* Section 8.4) is suitable for all types of users, with varying degrees of expertise in the relevant fields.

Some users, on the other hand, interacted in advance, for example, by saying their names as soon as the question was posed. They still grabbed the microphone and even repeated their names after that. This could be interpreted as a clear indication of immersion. A curious side note is that no user waved hello at the start (*q.v.* Figure 9.10), but six did wave goodbye at the end (*q.v.* Figure 9.30). This, in addition to being a good immersion arguing point, may indicate that the actual immersion increased as a result of the experience. That is, the experience and its narrative appear to evolve the users more and more over time.

Except for a few cases where users did not follow the light trail at all (a small part of YN, *q.v.* Figure 9.32), all users appear to look in the direction we expect them

to. Some claimed it was because they didn't recognise it as a cue, whereas hearing the geese or seeing the narrator point somewhere has been shown to be effective.

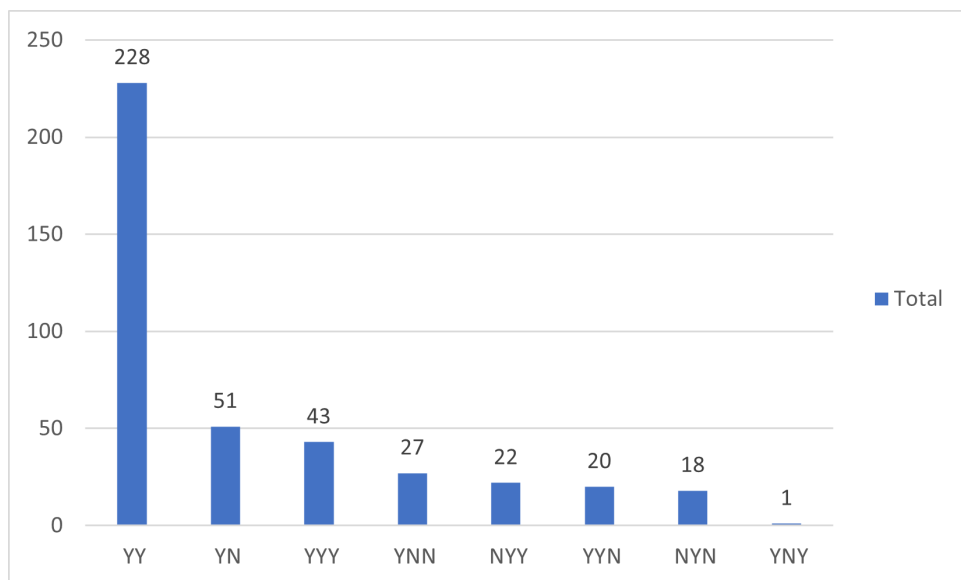


Figure 9.32 – Users trials observation codes' total count.

Despite the short time interval between the questionnaire and the experience, users appear to agree on either the predetermined gestures (**YYN**, accounting for 10% of total observed interactions, *q.v.* Figure 9.33, below) or the corresponding pretest answers, **NYY** (*q.v.* Figure 9.32, above). When it comes to actions like looking and talking, users demonstrated that they behaved pretty much as expected (**YY**), accounting for 56% of total observed interactions (*q.v.* Figure 9.33, below). The agreement with both the standard and the corresponding pretest answers (**YYY**) also indicates a high level of affluence and indicates that the experience performed well in terms of defaults. Nonetheless, the percentages of no interactions when they are expected and even asked about beforehand (**YN** and **YNN**, with 12% and 7%, respectively) indicate that the experience needs to be improved in that regard. Unusual behaviour was still observed in some cases (**NYN**). All of this suggests that users are acting instinctively rather than thinking about and remembering what they previously said. Even more intriguing was hearing the users speak after all trials had been completed, claiming they had answered some parts of the questionnaire to determine how they would behave like other users when they had not. Some gestural

and/or somatic movements appear to make more sense in their minds after the experience, whereas the actions were not as contextualised when they first answered the questionnaire. Again, it appears that context is important, and that context increases immersion throughout the experience.

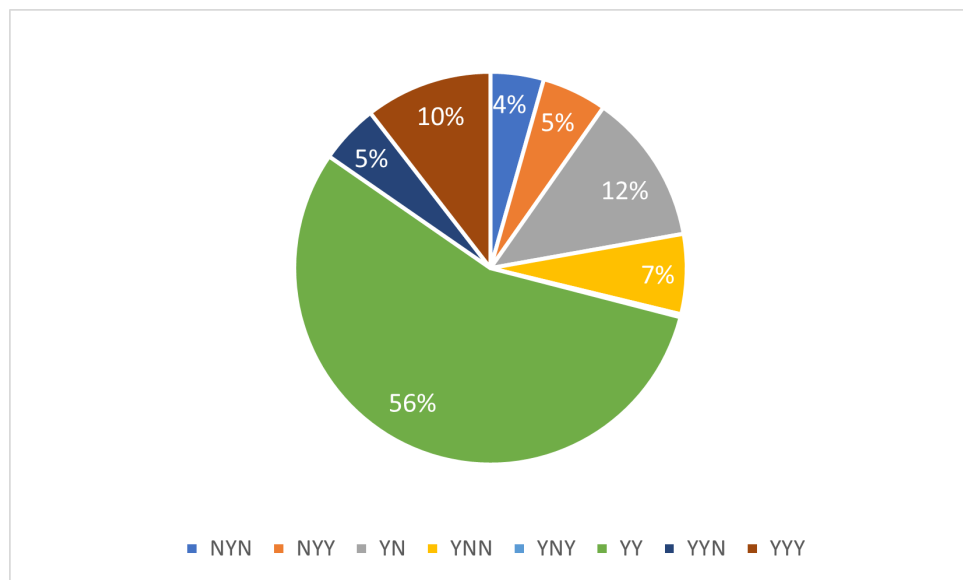


Figure 9.33 – User trials observation code's percentages.

Finally, during the post-interviews, some of the users who did not understand much English (if any) argued that the subtitles hampered or limited their interactions with their surroundings. Apparently, because the subtitles are fixed and do not move with the head movements or the **HMD**, they could either read or interact with what was happening around them. Others, mostly English speakers, contended that this was not the case and that they could interact. This appears to indicate that this technicality may become an issue going further. One possible solution is to move the subtitles whenever the user moves his or her head, allowing the user to move freely while still reading. Still, as pointed out, listening, reading, and interacting at the same time is a bit too quick, and more time to interact freely is suggested.

In terms of the exploration phase, each user appeared to have a different idea of what to do. Despite the fact that locomotion was not envisioned in this stational experience, some wanted to go to the mountains. Others wanted to go inside the

cabins, which was also not planned. Finally, some investigated the dynamics of the birds by rewatching the cycles through the ‘change’ action, emphasising, in reference to the geese, the importance of working together as a group for the best interests of their flocks. Some users even dug deeper into the landmarks they used, including the magnetic field. Most, on the other hand, enjoyed exploring the V-shaped formation while grabbing one of the birds backwards just to see them reunite. The experience dynamics at the end, with four or five sequenced geese animations, appear to have hyped the majority, if not all, of the users.

In terms of actual comprehension of the complex concepts, only one user mentioned the importance of home bonding and how the geese always return home during the post-interview. Even though it wasn’t one of the most explored concepts, it was still understandable. The majority of users appear to have understood the importance of nature and wildlife to the native, and even mentioned that, despite rarely being used nowadays, similar things are true in Portugal: the appearance of swallows as a harbinger of Spring.

10

Final Considerations

10.1 Contributions and Limitations

There are philosophical arguments arguing that, when compared to more traditional interfaces, people will be able to use a wider range and depth of richer interactions with cultural gestures. An earlier paper ([Pinto et al., 2015](#)) proposed a method for assessing this by analysing how people act in a videogame or a similar immersive system. [Carvalho \(2019\)](#) recently demonstrated that cultural gestures have a higher rate of remembering after 15 days and fewer errors than non-cultural gestures. It also demonstrates that they are equally satisfied with either set.

Starting from these three points, this thesis explored the conjecture that combining cultural gestures, which can lead to richer interactions, immersive environments, and more context, can enable better learning of complex concepts and systems. There was, however, no pre-existing method of approaching or accomplishing this. The contribution of this thesis is thus the devised start-to-finish process of going from cultural exemplar to complex concepts, to designing of an immersive environment for their exploration (with narrative, content, and interactions), and of identifying cultural gestures to interact with it, enabling that exploration.

A systematic literature review on how to teach and learn about complex concepts is presented, first emphasising the importance of context's multidimensionality. This significance demonstrates that context and gestures are not being used effectively for this problem, at least not yet, so there wasn't a way of telling if it was even a relevant problem.

After conducting ethnographic research on the Flathead Indian Reservation (CSKT), we present a set of complex concepts based on the children's book Luli's Journey. After conducting a multimodal analysis to define a more grounded, less subjective narrative, we iteratively design a VR concept and prototype based on this analysis, leveraging interaction through an architecture (InMERSE) that we also improve in an attempt to free end-user applications from specific input devices and modalities. With this final combination, we first determine whether the prototype complies with the complex learning literature and its guidelines. We observe their interactions, which leads to the following remarks:

- If they don't know how to interact, the prototype or its concepts aren't being presented in the best light. In those situations, we explain what to do but not how to do it;
- They attempt to use gestures, somatic movements, or other modes of interaction, and we register them all as part of Portuguese culture:
 - Follow along if that gesture is expected;
 - Otherwise, the researcher interrupts, apologises, and makes the appropriate gesture to continue the experience, demonstrating the inadequacy of the standard expected interaction.

We record and register all of the interactions that users have, later analysing it to understand what they explore and what they do not explore. What commands/gestures they use, why they use them, whether or not they repeat them throughout the experience, and so on. Then, we have a feasible, tested way of reiterating the artefact (the prototype) with a new, improved, default set of gestures that should make more sense in Portuguese culture.

To summarise, this thesis sets forth from a cultural exemplar (a child's tale) and provides (1) a process for identification and extraction of complex concepts, (2) an improved version of the InMERSE software architecture to enable device-independent gesture interaction, (3) a method for designing immersive interactive environments for exploring complex concepts associated with a cultural exemplar, and (4) a method for creating cultural gestures for interaction with the environment. This is the entire process and outcome of how we put the conjectured idea into action. It is neither better nor worse than other approaches; it is simply the first step toward something novel.

10.2 Future Research Directions

The conceptualised and designed solution is not meant to be a dogma, but rather a means of contributing to the solution and ongoing discussion of best technological solutions that, in an increasingly decisive way, help with end-user interaction by leveraging people's cultural diversity and prior knowledge. It is about adapting technology to people rather than people to technology. After all, what is science if not a naive, selfless attempt to improve society?

As a result, and because this manuscript only describes a proof of concept tested with the Portuguese culture, the artefact presented herein can proceed to implement some changes based on feedback and analysis of user trials, observations, and interviews. Moving the subtitles slightly in tandem with the head movement is one example of this, allowing for more interactions with the surroundings while avoiding losing track of the story being told. Finally, after validating the entire narrative with them, changes to the narration voice should be made to Salish, allowing for user trials on the Flathead Indian Reservation (CSKT) in Montana. The prototype can then be expanded and tested with other cultures, such as those in Central Europe and the Middle East.



SLR's Studies and Quality evaluation

A.1 Included Studies

Table A.1 – Studies included in final review.

ID #	Reference	Type
S1	Falck, H. W., & Halskov, K. (2013). Towards a framework for projection installations. In PerDis 2013 - Proceedings: 2nd ACM International Symposium on Pervasive Displays 2013 (pp. 67–72).	Conference
S2	Jiménez Fernández-Palacios, B., Morabito, D., & Remondino, F. (2017). Access to complex reality-based 3D models using virtual reality solutions. Journal of Cultural Heritage, 23, 40–48.	Journal
S3	Morgado, L., Rodrigues, R., Coelho, A., Magano, O., Calçada, T., Cunha, P. T., . . . Ramos, L. (2015). Cities in Citizens' Hands. In Procedia Computer Science (Vol. 67, pp. 430–438).	Conference

-
- | | | |
|-----------|---|------------|
| S4 | Bazlamit, R. (2018). Communicating experiential and spatial qualities of the occupation of Palestine using mixed video-virtual environments. <i>Virtual Creativity</i> , 8(1), 7–21. | Journal |
| S5 | Artopoulos, G., Charalambous, P., & Wehmeier, C. E. (2019). Immersive computing and crowd simulation techniques in modelling urban commons: The case of Nicosia-Cyprus. <i>International Journal of E-Planning Research</i> , 8(1), 35–49. | Journal |
| S6 | Cornelisse, R., Blundell, D., 2016. A Taiwan virtual reality memory project: Rituals in the circle, in: Addison A.C. Thwaites H., L.L.S. (Ed.), <i>Proceedings of the 2016 International Conference on Virtual Systems and Multimedia, VSMM 2016</i> . Institute of Electrical and Electronics Engineers Inc. | Conference |
| S7 | Partarakis, N., Antona, M., Zidianakis, E., Stephanidis, C., 2016. Adaptation and content personalization in the context of multi user museum exhibits, in: <i>CEUR Workshop Proceedings</i> . pp. 5–10. | Conference |
| S8 | Vogel, D., Lubos, P., Steinicke, F., 2018. AnimationVR - Interactive Controller-Based Animating in Virtual Reality, in: <i>2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)</i> . p. 1. | Conference |
| S9 | Fonseca, D., Navarro, I., de Renteria, I., Moreira, F., Ferrer, Á., de Reina, O., 2018. Assessment of Wearable Virtual Reality Technology for Visiting World Heritage Buildings: An Educational Approach. <i>Journal of Educational Computing Research</i> . 56, 940–973. | Journal |
-

- S10** Agarwal, S., Dagli, C.H., 2013. Augmented cognition in human system interaction through coupled action of body sensor network and agent based modeling, in: *Procedia Computer Science*. pp. 20–28. Conference
- S11** Zünd, F., Ryffel, M., Magnenat, S., Marra, A., Nitti, M., Kapadia, M., ... Sumner, R. W. (2015). Augmented creativity: Bridging the real and virtual worlds to enhance creative play. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*, SA 2015. Journal
- S12** Wiehr, F., Daiber, F., Kosmalla, F., & Krüger, A. (2016). BetaCube - Enhancing training for climbing by a self-calibrating camera-projection unit. In *Conference on Human Factors in Computing Systems - Proceedings* (Vol. 07-12-May-, pp. 1998–2004). Conference
- S13** Zhang, Y., Liu, S., Tao, L., Yu, C., Shi, Y., & Xu, Y. (2015). ChinAR: Facilitating Chinese Guqin learning through interactive projected augmentation. In *ACM International Conference Proceeding Series* (Vol. 18-19-April, pp. 23–31). Conference
- S14** Searle, K. A., Casort, T., Litts, B. K., Jones Brayboy, B. M., Dance, S. L., & Kafai, Y. (2018). Cultural repertoires: Indigenous youth creating with place and story. In *Proceedings of International Conference of the Learning Sciences, ICLS* (Vol. 2, pp. 697–704). Conference
- S15** Puyuelo Cazorla, M., Val Fiel, M., Higón Calvet, J. L., & Merino Sanjuán, L. (2015). FROM THE REPRESENTATION TO THE EXPERIENCE. Augmented Reality for the interpretation of the monument heritage in “La Lonja” of Valencia. *EGA. Revista de Expresión Gráfica Arquitectónica*, 20(26), 180. Journal
-

-
- | | | |
|------------|---|------------|
| S16 | Palma, V., Spallone, R., & Vitali, M. (2019). Digital interactive baroque atria in turin: A project aimed to sharing and enhancing cultural heritage. <i>Advances in Intelligent Systems and Computing</i> , 919, 314–325. | Journal |
| S17 | Marti, P., Peeters, J., Trotto, A., Tittarelli, M., True, N., Papworth, N., & Hummels, C. (2015). Embodying culture: Interactive installation on women's rights. <i>First Monday</i> , 20(4). | Journal |
| S18 | Muender, T., Fröhlich, T., & Malaka, R. (2018). Empowering creative people: Virtual reality for previsualization. In <i>Conference on Human Factors in Computing Systems - Proceedings (Vol. 2018-April)</i> . | Conference |
| S19 | Jowsey, S., & Williams, M. (2017). Engaging the Invisible. In <i>SIGGRAPH ASIA 2017 SYMPOSIUM ON EDUCATION (SA'17)</i> . | Conference |
| S20 | Delgado, F. J., Martinez, R., Finat, J., Martinez, J., Puche, J. C., & Finat, F. J. (2013). Enhancing the reuse of digital resources for integrated systems to represent, understand and dynamize complex interactions in architectural cultural heritage environments. In <i>ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. 2, pp. 109–114)</i> . | Conference |
| S21 | Bottino, A., Martina, A., & Toosi, A. (2015). GAINÉ - tanGible augmented interaction for edutainment. In <i>2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)</i> . Turin: IEEE. | Conference |
-

-
- S22** Kosek, M., Koniaris, B., Sinclair, D., Markova, D., Rothnie, F., Smoot, L., & Mitchell, K. (2017). IRIDIUM+: Deep media storytelling with non-linear light field video. In ACM SIGGRAPH 2017 VR Village, SIGGRAPH 2017. Association for Computing Machinery, Inc. Conference
- S23** Bardakjian, S.-S., Mokhov, S. A., Song, M., & Mudur, S. P. (2016). ISSv3: From human motion in the real to the interactive documentary film in AR/VR. In SA 2016 - SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments. Association for Computing Machinery, Inc. Conference
- S24** Krishnaswamy, M., Lee, B., Murthy, C., Rosenfeld, H., & Lee, A. S. (2017). Iyagi: An immersive storytelling tool for healthy bedtime routine. In TEI 2017 - Proceedings of the 11th International Conference on Tangible, Embedded, and Embodied Interaction (pp. 603–608). Conference
- S25** Bugalia, N., Kumar, S., Kalra, P., & Choudhary, S. (2016). Mixed reality based interaction system for digital heritage. In Proceedings - VRCAI 2016: 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry (Vol. 1, pp. 31–37). Conference
- S26** Nurminen, A. (2015). Mobile multisensory augmentations with the CultAR platform. In SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications, SA 2015. Conference
- S27** Chu, J., Bryan, C., Shih, M., Ferrer, L., & Ma, K.-L. (2017). Navigable videos for presenting scientific data on affordable head-mounted displays. In Proceedings of the 8th ACM Multimedia Systems Conference, MMSys 2017 (pp. 250–260). Conference
-

-
- S28** Bock, A., Hansen, C., & Ynnerman, A. (2018). Journal
OpenSpace: Bringing NASA Missions to the Public. *IEEE*
Computer Graphics and Applications, 38(5), 112–118.
- S29** Panou, C., Ragia, L., Dimelli, D., & Mania, K. (2018). Conference
Outdoors mobile augmented reality application visualiz-
ing 3D reconstructed historical monuments. In R. L. Lau-
rini R. Grueau C. (Ed.), *GISTAM 2018 - Proceedings of*
the 4th International Conference on Geographical Informa-
tion Systems Theory, Applications and Management (Vol.
2018-March, pp. 59–67). SciTePress.
- S30** Murru, G., Fratarcangeli, M., & Empler, T. (2013). Prac- Conference
tical augmented visualization on handheld devices for cul-
tural heritage. In *21st International Conference in Central*
Europe on Computer Graphics, Visualization and Com-
puter Vision, WSCG 2013 - Communication Papers Pro-
ceedings (pp. 97–103).
- S31** Mancini, A., Clini, P., Bozzi, C. A., Malinverni, E. S., Journal
Pierdicca, R., & Nespeca, R. (2017). Remote touch inter-
action with high quality models using an autostereoscopic
3D display. *Lecture Notes in Computer Science (Includ-*
ing Subseries Lecture Notes in Artificial Intelligence and
Lecture Notes in Bioinformatics), 10325 LNCS, 478–489.
- S32** Kampa, A., Stöbener, K., & Spierling, U. (2016). User in- Journal
terface prototyping for handheld mobile augmented reality
applications. *Lecture Notes in Computer Science (Includ-*
ing Subseries Lecture Notes in Artificial Intelligence and
Lecture Notes in Bioinformatics), 9926 LNCS, 229–234.
-

- S33** Flynn, B. (2013). V-Embodiment for cultural heritage. Conference
In Proceedings of the DigitalHeritage 2013 - Federating
the 19th Int'l VSMM, 10th Eurographics GCH, and 2nd
UNESCO Memory of the World Conferences, Plus Special
Sessions fromCAA, Arqueologica 2.0 et al. (Vol. 1, pp.
347–354).
- S34** Brumana, R., Oreni, D., Caspani, S., & Previtali, M. Journal
(2018). Virtual museums and built environment: Narra-
tives and immersive experience via multi-temporal geodata
hub. *Virtual Archaeology Review*, 9(19), 34–49.
- S35** Lo, P., Chan, H. H. Y., Tang, A. W. M., Chiu, D. K. W., Journal
Cho, A., See-To, E. W. K., . . . Shaw, J. (2019). Visualising
and revitalising traditional Chinese martial arts: Visitors'
engagement and learning experience at the 300 years of
Hakka Kungfu. *Library Hi Tech*.
- S36** Kidd, J. (2019). With New Eyes I See: embodiment, em- Journal
pathy and silence in digital heritage interpretation. *Inter-
national Journal of Heritage Studies*, 25(1), 54–66.
- S37** Marks, S., Estevez, J. E., & Connor, A. M. (2014). To- Conference
wards the holodeck: Fully immersive virtual reality visual-
isation of scientific and engineering data. In *ACM Interna-
tional Conference Proceeding Series* (Vol. 19-21-Nove, pp.
42–47). Association for Computing Machinery.
- S38** He, Y., Zhang, Z., Nan, X., Zhang, N., Guo, F., Rosales, E., Journal
& Guan, L. (2017). vConnect: perceive and interact with
real world from CAVE. *Multimedia Tools and Applications*,
76(1), 1479–1508.
-

A.2 Quality evaluation

Table A.2 – Quality evaluation of included studies.

ID #	QAQ1	QAQ2	QAQ3	QAQ4	QAQ5	QAQ6	QAQ7	QA Score
S1	1	0.5	1	1	0.5	0.5	0	4.5
S2	1	1	0.5	1	0.5	0.5	0.5	5
S3	1	1	1	1	1	0	0	5
S4	1	1	1	1	1	0.5	0	5.5
S5	1	0.5	1	1	1	1	0	5.5
S6	1	0.5	1	1	0.5	0	0.5	4.5
S7	1	1	0.5	1	1	1	0	5.5
S8	1	1	1	1	0.5	0.5	0.5	5.5
S9	1	1	1	1	1	1	1	7
S10	1	1	0.5	1	0.5	0	0	4
S11	1	1	0.5	1	0	0	0	3.5
S12	1	1	0.5	1	1	0	0	4.5
S13	1	1	0.5	1	1	1	1	6.5
S14	1	1	0.5	1	1	1	0	5.5
S15	1	1	0.5	1	1	0	0	4.5
S16	1	1	1	1	0.5	0	0	4.5
S17	1	1	0.5	1	1	0	0	4.5
S18	1	1	0.5	1	1	0.5	0.5	5.5
S19	1	0.5	0.5	1	0.5	0	0	3.5
S20	1	1	0.5	1	1	0	0	4.5
S21	1	1	0.5	1	1	0	0.5	5
S22	1	1	0.5	1	0.5	0	0	4
S23	1	0.5	0.5	1	0	0	1	4
S24	1	1	1	1	1	0	0.5	5.5
S25	1	1	0.5	1	1	1	0.5	6



CSKT Collected Evidence

B.1 Luli's Journey collected evidence

Table B.1 – Collected evidence for each unit of analysis, for pages 4 and 5 of the book Luli's Journey.

Unit of Analysis	Evidence (page 4 and 5)
Intramodal considerations	
Linguistics	Salish
Characters	Female Canadian Goose (Luli); Mother or Father adult geese (black and grey)
Pose	(Luli) Hatching from egg; Mother (or father) laying down next to Luli
Objects	Nest
Actions	Nurturing Luli
Setting	Shore of McDonald Lake; Green background
Emotions	Love and care
Gaze	Luli and mother (or father) looking at the reader
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.2 – Collected evidence for each unit of analysis, for pages 6 and 7 of the book Luli's Journey.

Unit of Analysis	Evidence (page 6 and 7)
Intramodal considerations	
Linguistics	Salish; "Summer sun brings lots of fun"
Characters	Luli (gosling); Father adult goose; Mother adult goose;
Pose	Father and Luli swimming; Mother standing while collecting berries
Objects	Red Berries bush
Actions	Wade in the water; following her father; eating berries or plants brought by her mother (who is collecting them); Luli learns how to fly and run;
Setting	Blue Water (McDonald Lake?); Summer sun; Blue sky; Mountains background (greenish)
Emotions	"Lots of fun"
Gaze	N/D
Interpersonal distance	Close personal (Luli and father); Far personal (in the picture, as mother collects berries to bring Luli)
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.3 – Collected evidence for each unit of analysis, for pages 8 and 9 of the book Luli’s Journey.

Unit of Analysis	Evidence (page 8 and 9)
Intramodal considerations	
Linguistics	Salish;
Characters	Luli (adult–big and strong; black and grey); Father and Mother (Geese); Brown Bear; Brown Squirrel
Pose	All standing
Objects	Flowers (lavender, Arrowleaf Balsamroot’s)
Actions	“Won’t be long until their journey to the land of the sun”
Setting	Blue sky; Green Mountains; Green field; Mention of “land of the sun”
Emotions	Confidence/Admiration (from Mother), as she considers Luli big and strong enough for the journey
Gaze	All seem to be focusing on Luli, in the center
Interpersonal distance	Around Luli (not too close)
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.4 – Collected evidence for each unit of analysis, for pages 10 and 11 of the book Luli's Journey.

Unit of Analysis	Evidence (page 10 and 11)
Intramodal considerations	
Linguistics	Salish;
Characters	Luli (adult) and Mother or Father (maybe father given recall from father on next page)
Pose	Swimming; Luli with head below water
Objects	yellowish/reddish leaves on tree, and on the ground; red barn (within farm's boundaries)
Actions	Leaves dropping to the ground; Luli finds food all around (e.g., grain from farmer's crops and plants); Luli finds plants beneath murky ponds
Setting	Autumn; Farm; Water; Green Mountains; Blue sky;
Emotions	N/D
Gaze	N/D
Interpersonal distance	Close Personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.5 – Collected evidence for each unit of analysis, for pages 12 and 13 of the book Luli's Journey.

Unit of Analysis	Evidence (page 12 and 13)
Intramodal considerations	
Linguistics	Salish; “before Kerr Damn was built”
Characters	Luli's father; Friend Adult Goose
Pose	Standing with open and closed wings
Objects	Dark blue water; rocky background; muddy ground
Actions	Luli's father and friends (another Goose) playing
Setting	Autumn days; Kerr Dam; Muddy flats; Flat-head Lake
Emotions	Longing; Playfulness
Gaze	Looking at each other
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.6 – Collected evidence for each unit of analysis, for pages 14 and 15 of the book Luli's Journey.

Unit of Analysis	Evidence (page 14 and 15)
Intramodal considerations	
Linguistics	Salish; “Long before winter’s frost blows in” =>Before winter arrives
Characters	Luli; Luli’s friends (maybe from first picture with many other birds)
Pose	Standing to ruffle feathers; Flying pose
Objects	North America Map, showing the flathead reservation and possible destination in mexico (Matamoros)
Actions	Luli ruffles her feathers; Luli flies
Setting	From linguistics =>Before Winter; Flies to Mexico (Matamoros)
Emotions	Goodbye (leaving friends); Preparation to fly
Gaze	N/D
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.7 – Collected evidence for each unit of analysis, for pages 16 and 17 of the book Luli's Journey.

Unit of Analysis	Evidence (page 16 and 17)
Intramodal considerations	
Linguistics	Salish; “Canadian Geese let them know: if the summer will be hot and really dry or if the winter will be harsher as the days go by” => Giving seasonal information to people
Characters	Luli and Luli's family; People (mother and daughter-child) from one of the tribes
Pose	V formation => self-organisation; Mother and daughter standing; Mother Pointing to birds in the sky, while touching daughter's back/clavicle
Objects	People's Clothes have drawings of stars and some sort of diamonds. Both wearing necklaces
Actions	Geese Flying; Mother pointing to birds; Daughter looking at the same direction
Setting	Together in flight (sky); day and night; Blueish mountains with snow on peaks; Green field; Sun
Emotions	Union (Geese staying together); Mild happiness from mother and daughter (while watching the geese); Caring (mother teaching daughter)
Gaze	N/D for birds; Mother and daughter looking at birds
Interpersonal distance	Close personal (both for birds and people)
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.8 – Collected evidence for each unit of analysis, for pages 18 and 19 of the book Luli's Journey.

Unit of Analysis	Evidence (page 18 and 19)
Intramodal considerations	
Linguistics	Salish; “weeks of hard flying” =>time to relax; “sun and bliss” =>happiness
Characters	Luli
Pose	Looking at Palm Trees
Objects	Palm trees; blue ocean; blue sky
Actions	Arrives to Mexico; Looking at palm trees; Feeling the warm sand and ocean breeze
Setting	Mexico; sun and bliss; Warm sand; Ocean breeze;
Emotions	From linguistics =>Relaxation/Filled duty; Excitement (to see palm trees =>discovery); From linguistics =>Happiness
Gaze	Looking at ocean
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.9 – Collected evidence for each unit of analysis, for pages 20 and 21 of the book Luli's Journey.

Unit of Analysis	Evidence (page 20 and 21)
Intramodal considerations	
Linguistics	Salish; “paradise shore”; “feels like she needs so much more” =>insatisfaction
Characters	Luli
Pose	Swimming
Objects	Palm tree; reddish sky =>sunset-like
Actions	Remembering times at home
Setting	Paradise, but not home (McDonald Lake, Mountains, and trees);
Emotions	Longing; from linguistics =>insatisfaction
Gaze	Looking down at water
Interpersonal distance	N/D
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.10 – Collected evidence for each unit of analysis, for pages 22 and 23 of the book Luli's Journey.

Unit of Analysis	Evidence (page 22 and 23)
Intramodal considerations	
Linguistics	Salish; “winter has passed”; “at last” =>excitement to go back; “each taking their turn to be in the lead” =>feedback loops and self-organisation
Characters	Luli and Luli's parents;
Pose	Ruffle feathers =>standing; Flying in V formation =>self-organisation
Objects	N/D
Actions	Ruffling their feathers; Flying north
Setting	From linguistics =>no longer winter (spring); Above water; Sunset-like sky (colorful, too)
Emotions	Preparation (ruffling feathers); from linguistics =>excitement
Gaze	Looking in front of them
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.11 – Collected evidence for each unit of analysis, for pages 24 and 25 of the book Luli's Journey.

Unit of Analysis	Evidence (page 24 and 25)
Intramodal considerations	
Linguistics	Salish; “the time when birds arrive after a late winter snowstorm is called Bird’s sock”
Characters	Luli and Luli’s family
Pose	V formation =>self-organisation
Objects	Lavender flowers;
Actions	Spring melts the snow; Flower buds grow; Luli’s family arrives home flying;
Setting	Spring; Green Mountains (no more snow at peaks); Some now being melt on the ground; After winter’s last blow; White clouds; Sun
Emotions	N/D
Gaze	N/D
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.12 – Collected evidence for each unit of analysis, for pages 26 and 27 of the book Luli's Journey.

Unit of Analysis	Evidence (page 26 and 27)
Intramodal considerations	
Linguistics	Salish; "Someday, Luli will lay eggs of her own and she will take them on the same journey that she has flown" =>circle of life; "... but her lake-side nest will always be home" =>sense of belonging; "... make the earth come alive" =>joy
Characters	Luli and her friends (Bear)
Pose	Luli laying on her nest; Bear sitting; Luli's parents swimming
Objects	Rocks; grass; wild flowers
Actions	Wild flowers bloom; Luli's parents swim;
Setting	Green Mountains; Blue sky;
Emotions	Comfort (Luli on her nest); Love (between the parents); From linguistics = >Joy
Gaze	Luli looking at bear friend; Parents looking at each other
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.13 – Collected evidence for each unit of analysis, for pages 28 and 29 of the book Luli's Journey.

Unit of Analysis	Evidence (page 28 and 29)
Intramodal considerations	
Linguistics	Salish; “birds’ knowledge of seasons is so precise Native people depended on them to survive. The Canadian Geese’ annual migrations gave the people an idea of weather patterns as well as timelines for season changes”
Characters	Geese (sky); Mother and daughters
Pose	Daughter standing, pointing at the sky; Mother with knees bent
Objects	Teepee (sort of a tent); Baby in sort of a backpack, with flowers drawings on it, carefully wrapped in clothes (warmed up); Older daughter wearing long sleeves
Actions	Daughter pointing at the geese flying (learnt from mother =>circle of life); From linguistics =>Mother preparing some clothes for winter; Birds migrating; Season changing
Setting	Blueish mountains; Snow at the peaks; Green field; Blue sky; Winter is coming (birds are flying away; mother is preparing clothes)
Emotions	Nurturing (mother taking care of daughters); Comfort (baby); Hard-work (mother)
Gaze	Daughter looking at sky (birds); Baby seems to be looking up (learning!); Mother is looking at the clothes she is putting together
Interpersonal distance	Close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

Table B.14 – Collected evidence for each unit of analysis, for pages 30 and 31 of the book Luli's Journey.

Unit of Analysis	Evidence (page 30 and 31)
Intramodal considerations	
Linguistics	Salish; “Like Native people long ago, you can continue to observe the birds’ migration and make it a part of your life”; “Canadian Geese will occupy the same nest ... throughout their lifespan” =>sense of belonging; “Resident Geese ... will live near their nesting sites” =>not all birds migrate; “migrate in order to have plentiful food supply during colder months” =>birds migrate to survive/be in better conditions
Characters	From linguistics =>Reader; Birds (migrating)
Pose	October-January and February-April =>flying;
Objects	Migration cycle
Actions	March-May =>choose a mate and lay eggs; October-January =>after 6-9 weeks after hatching, goslings are ready for flight; February-April =>flying north (back home)
Setting	Birds migration cycle; October-January =>ground and water freezes; December and January =>warmer place (southern part of the USA or Mexico)
Emotions	N/D
Gaze	N/D
Interpersonal distance	Birds migrating =>close personal
Intermodal considerations	
Ideational concurrence	Exemplification
Ideational complementarity	Joint/synergistic
Ideational counterpoint	Non contradictory

B.2 SKC IRB Protocol



SKC IRB protocol -1 -

IRB NEW PROTOCOL

Project Title:	Salish, Pend d'Oreille, and Kootenai (immersive) Storytelling
-----------------------	---

Investigator Information:

Principal Investigator:	Professor Jonathon Richter	Secondary Investigator or Project Supervisor*:	Luís Fernandes
Department:		Department:	
Department Phone:		Department Phone:	
Contact Phone:		Contact Phone:	
Contact Address:		Contact Address:	
City/State/Zip:		City/State/Zip:	
E-Mail Address:	jonathon_richter@skc.edu	E-Mail Address:	lfernandes@utad.pt

* Student projects must be submitted with a faculty member listed as Secondary Investigator or Project Supervisor.

Principal Investigator is:

☐	Faculty	☐	Staff	☐	Student
X	Outside Researcher	☐	Other (Please specify:)		

Type of Project:

X	Research	☐	Grant Activity	☐	Class Project
☐	Other (please specify):				

Does the research involve an outside institution/agency other than SKC*?	Yes ☐	No ☒
--	------------------------------	--

If yes, please list the institutions/agencies.

* Has Written Permission been obtained from the cooperating institutions/agencies? **If so, please attach.**

Project Information:

Present/Proposed Source of Funding:	
Anticipated Project Start Date:	Anticipated Project End Date:

*Please attach a copy of the funding application.

* You may not start this project until IRB approval is received.

SKC IRB protocol -2 -

Type of Review Requested:*Please check your response to each question.*

	Yes	No	1. Does the research involve prisoners?
	Yes	No	2. Does the research involve using survey or interview procedures with children (under 18 years of age) that is not conducted in an educational setting utilizing normal educational practices?
	Yes	No	3. Does the research involve the observation of children in settings where the investigator will participate in the activities being observed?
	Yes	No	4. Will videotaping or audio tape recording be used?
	Yes	No	5. Will the participants be asked to perform physical tasks?
	Yes	No	6. Does the research attempt to influence or change participants' behavior, perception, or cognition?
	Yes	No	7. Will data collection include collecting sensitive data (illegal activities, sensitive topics such as sexual orientation or behavior, psychological characteristics, or other data that may be painful or embarrassing to reveal)?
	Yes	No	8. For research using existing or archived data, documents, records or specimens, will any data, documents, records, or specimens be collected from subjects after the submission of this application?
	Yes	No	9. Can subjects be identified, either directly or indirectly, from the data, documents, or records?
	Yes	No	10. Does the research involve potentially culturally sensitive topics pertinent to the Confederated Salish and Kootenai Tribes?

Please check the category of review. Note that the SKC IRB will make a final determination of the review category.

	Exempt
	Expedited
	Full Board
	Cultural

Description of Subjects:

Total number of participants who are minors (less than 19):	
Total number of participants who are adults:	

What are the participants' characteristics? If study participants are restricted to one gender, one race/ethnicity, or other single demographic characteristics, include the rationale.

We are looking for tribal members, particularly elders who can help us design the storytelling experience. In a testing phase, students from SKC may also participate.

SKC IRB protocol -3 -

Special Considerations:

Do any of the following considerations apply to the proposed research?

If yes, please check all appropriate blanks below.

X	Audio taping	X	Videotaping	X	Archival/Secondary Data Analysis	X	Cultural materials
X	Photography		Web-based research (Research involving online surveys, email, or other electronic communication.)		Biological Samples		Protected Health Information

Project Personnel List:

Please list the names of all personnel working on this project, starting with the principal investigator and the secondary investigator/project advisor. Research assistants, students, data entry staff and other research project staff should also be included. Note that all personnel must complete human subjects and cultural training. See the SKC IRB site for more information.

Name of Individual:	Project Role:	Status at SKC (faculty, staff, student, co-investigator from other institution, etc)	Human Subjects Protection and Cultural Training Certification*
Jonathon Richter	Principal Investigator	Faculty	Yes
Luís Fernandes	Project Supervisor	Visiting Researcher	No

* Attach certificate of Human Subjects Protection and Cultural Training, or note certificate is on file in the SKC IRB office.

Required Signatures:

Your signature indicates agreement that the study will be performed in accordance with the Salish Kootenai College Institutional Review Board policies.			
Principal Investigator:		Date:	
Secondary Investigator/Project Advisor:		Date:	

SKC IRB protocol -4 -

Part B:
PROJECT DESCRIPTION

FOR OFFICE USE ONLY

INITIALS OF IRB CHAIR:

DATE APPROVED:

1. Describe the significance of the project.

What is the significance/purpose of the study? (Please provide a brief 1-2 paragraph explanation in lay terms.)

We are trying to assess the feasibility of employing cultural gestures and symbols (i.e., semiotics somatic emblems) towards a "more natural" interaction in immersive environments (e.g., virtual or mixed reality), by leveraging users' cultural semantics towards improving learning/teaching of complex systems.

The idea is to develop an immersive storytelling prototype for the Confederated Salish and Kootenai Tribes (CSKT) in Montana, United States. Specifically, we are planning on basing it on a published child story told by a Pend d'Oreille elder, Luli's Journey, about Canadian Geese migrations. The goal within this storytelling prototype would be to learn about their V-shaped flight formations (i.e., complex flight dynamics) as well as the insight people obtained in the past from observing the birds (i.e., weather patterns), while still giving the users a sense of agency (i.e., control). It can be extended to the topic of climate change and its effects on the food chain (e.g., whitebark pine and Clark's Nutcracker). It should take place, through the use of immersive technology, somewhere in the Flathead Indian reservation (e.g., Kicking Horse Reservoir, facing the Eagle Pass Mountain), and enable meaningful interaction through cultural gestures and emblems. Examples of that could be switching to bird's eye view (e.g., using a gesture that represents 'sky') or changing the current season (e.g., winter, summer) and reflecting that through landscape changes within the immersive environment. This idea can be further developed. Acting upon the complexity and systemic concepts to modify them, users could also be able to demonstrate to others their perspectives on the subjects, in similarly location-aware, immersed fashion. Examples of areas where systemic concepts are critical for end-user understanding and participation include urban planning, ecology, entrepreneurship, psychology, among others.

These tribes could use these technologies to perpetuate the language and knowledge through storytelling. They could have full control of it and use this powerful medium and use it as they see fit.

In the process, we plan to collect strikingly different sets of somatic emblems directly from tribal members and prototype their use for interaction in immersive environments, i.e., involve in cooperative prototyping users who retain a deeper tie with cultural semantics and overall significance of those emblems.

The end result would then be a set of emblems and interaction prototypes which could serve as a markedly distinct case from Western European cultural environments, and thus enable this research to later establish parallel teams between the US and Portugal to field test new prototypes and evolve them.

SKC IRB protocol -5 -

2. Describe the methods and procedures. <i>Describe the data collection procedures and specifically what participants will have to do. In an initial phase, through questioning/interview. In a prototype phase, through questionnaires/surveys.</i> <i>How long will this take participants to complete?</i> The interviews (to elders or others that know details regarding the child story) shouldn't take longer than 20min each. If small prototypes are developed throughout the process, they should take no longer than a few minutes to test (10 to 15 minutes). <i>Exactly where (physical location) will the study take place?</i> Research will take place in an established or commonly accepted educational setting.
3. Describe recruiting procedures. <i>How will the names and contact information for participants be obtained?</i> Through a basic form just asking for the name, contact, and their approval to take photos and/or record video during the participation. <i>How will participants be approached about participating in the study?</i> At first, we will contact people directly (elders of tribes if possible). In a second phase, we will try to get both elders and students to test the immersive prototypes. <i>Will follow-ups or reminders be sent? If so, explain.</i> *Please submit copies of recruitment flyers, ads, phone scripts, emails, etc. These require IRB approval.
4. Describe Benefits and Risks. <i>Explain the benefits to participants or to others.</i> There is no clear benefit. <i>Explain the risks to participants. What will be done to minimize the risks? If there are no known risks, this should be stated.</i> There are no risks involved.
5. Describe Compensation. Will compensation be provided to participants? Yes No

SKC IRB protocol -6 -

If 'Yes', please describe amount and type of compensation, including money, gift certificates, extra course credit, etc.

6. Informed Consent

How will informed consent/assent be obtained?

Their consent to participate in the research, or not, will be through direct communication with the principal investigators. It will also be done through a consent form along with their name and signature. None of its information will be disclosed.

****Please attach copies of informed consent/assent forms, emails, and/or letters. Please refer to the SKC IRB website for information which must be included in informed consent materials.**

*** Please note SKC IRB policies concerning assent and parental permission for participants under the age of 18.**

7. Describe how confidentiality will be maintained.

How will confidentiality of records be maintained?

In the second phase (prototype testing), surveys will be completely voluntary. Participants can choose to not allow the use of their reflections/responses for the research purpose. Reflection responses will be redacted so that no personal identifying information, such as personal name, will be released. There is no scoring or grading involved.

Will individuals be identified?

Responses and reflection prompts will be redacted so that no personal identifying information (personal name, etc.) will be released. The surveys and questionnaires will be identified using a random ID.

How long will records be kept?

Where will records be stored?

Who has access to the records/data?

Only the principal investigator, project supervisor, and PhD supervisors.

How will confidential information be destroyed after the study is finished?

For web-based studies, how will the data be handled? Will the data be sent to a secure server? Will the data be encrypted while in transit? Will you be collecting IP addresses?

If transcriptions are required, how will transcriptions be handled? Who is doing the transcriptions? Please attach a copy of the confidentiality agreement that transcriptionists will sign.

The transcriptions will be handled in 3 steps:

SKC IRB protocol -7 -

1. Ask the participant to avoid saying something that can identify him/her
2. Go through the tape and remove sensitive/personal parts, if any
3. Ask the transcriber to not transcribe sensitive/personal topics, names, etc. (just in case step 2 didn't catch those)

** For studies utilizing Protected Health Information (PHI; e.g., information obtained from a hospital, clinic, or treatment facility), how will this PHI data be obtained and safeguarded? Please provide a copy of the release of authorization that will be used to obtain permission from the participant for the agency/institution to release protected health information for project purposes or a letter from the agency/institution documenting agreement to provide protected health information for project purposes.*

8. Plans for Publication.

How will the results of this project be reported? For example, will they be published, presented at conferences, sent to other agencies or individuals, or distributed in other ways?

This research is part of an ongoing PhD. As such, the results will be present both on the thesis and papers (e.g., journals, conferences).

9. Copies of questionnaires, survey, or testing instruments.

Please list all questionnaires, surveys, and/or assessment instruments/measures used in the project.

- Socio Demographic Questionnaire (pre-questionnaire)
- Consent Form
- Interview Questions

Please submit copies of all research instruments, tools, surveys, etc.

B.3 Socio Demographic Pre-Questionnaire

Salish, Pend d'Oreille, and Kootenai (immersive) Storytelling

Pre-Questionnaire

To be completed by the researcher:

ID: _____

Age: _____

Gender: ☐ Male ☐ Female ☐ Other

Do you wear glasses/contact lenses? ☐ Yes ☐ No Do you use a hearing device? ☐ Yes ☐ No

Job: _____

Education:

- ☐ None
- ☐ Elementary school
- ☐ High school
- ☐ College
- ☐ Graduate school

In case you are still a student, which year and field?

Classify your experience with computers:

- ☐ None
- ☐ Basic
- ☐ Medium
- ☐ Good
- ☐ Excellent

How do you classify your knowledge regarding immersive technology (e.g., virtual reality)?

- ☐ None
- ☐ Basic
- ☐ Medium
- ☐ Good
- ☐ Excellent

Salish, Pend d'Oreille, and Kootenai (immersive) Storytelling

Pre-Questionnaire (continuation)

How often do you play on the computer?

- ☐ Never
- ☐ Occasionally (1 to 2 times per week)
- ☐ Frequently (3 to 4 times per week)
- ☐ Very Frequently (5 to 6 times per week)
- ☐ Always (everyday)

Have you used immersive technology, such as virtual/mixed reality headsets, before?

- ☐ Yes
- ☐ No

Have you participated in this experience before?

- ☐ Yes
- ☐ No

When was the last time you used virtual/mixed reality headsets?

- ☐ Never
- ☐ Less than 1 week ago
- ☐ Between 1 to 2 weeks ago
- ☐ Between 2 weeks to 1 month ago
- ☐ Over 1 month ago

B.4 Interview's Form of Consent

Salish, Pend d'Oreille, and Kootenai (immersive) Storytelling

Consent to take part in research

- I,, voluntarily agree to participate in this research study.
- I understand that even if I agree to participate now, I can withdraw at any time or refuse to answer any question without any consequences of any kind.
- I understand that I can withdraw permission to use data from my interview within two weeks after the interview, in which case the material will be deleted.
- I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study.
- I understand that I will not benefit directly from participating in this research.
- I agree with my interview/participation being audio/video-recorded.
- I understand that all the information I provide for this study will be treated confidentially.
- I understand that in any report on the results of this research my identity will remain anonymous. This will be done by changing or removing my name and disguising any details of my interview which may reveal my identity or the identity of people I speak about.
- I understand that disguised extracts from my interview may be quoted in the principal investigator's PhD thesis, as well as in published conference and journal papers.
- I understand that if I inform the researcher that myself or someone else is at risk of harm they may have to report this to the relevant authorities—they will discuss this with me first but may be required to report with or without my permission.
- I understand that signed consent forms and original recordings will be retained by the principal investigator and that they can also be accessed by the project supervisor and PhD supervisors.
- I understand that under freedom of information legalisation I am entitled to access the information I have provided at any time while it is in storage as specified above.
- I understand that I am free to contact any of the people involved in the research to seek further clarification and information.

Salish, Pend d'Oreille, and Kootenai
(immersive) Storytelling

Consent to take part in research (continuation)

Signature of research participant

Signature Email (optional) Date (dd-mm-yy)

Signature of researcher

I believe the participant is giving informed consent to participate in this study

Signature Date (dd-mm-yy)

B.5 Interview's Questionnaire

Salish, Pend d'Oreille, and Kootenai (immersive) Storytelling

Interview Questions (1 of 2)

Note: Please read carefully the following questions, and mark those you don't want to answer or you want to be discarded from the interview. Ideally, the interview should be recorded (audio/video), but you can tell the researcher if you don't want it to, or have any special request. You can stop the interview at any point. The results/information you provide will only be considered with your approval. Your identity will be protected, as we identify each interview with a random ID.

1. What can you tell us about Storytelling?
 - a. How important is it in your culture?
 - b. Why do you do it?
 - c. What do you try to accomplish when doing so?
 - d. Do you have some special days or months for that matter?
 - e. Who would be eligible to listen? The whole village or tribe?
2. Do you know the child book story, **Luli's Journey**? (Canadian Geese migrations)
 - a. How old were you when you first heard this story?
 - b. Who told you that story?
 - i. Was it an elder?
 - ii. What was his/her name?
 - iii. Was it from the same tribe?
 - iv. Is he/she still with us?
 1. Can we contact him/her?
 - v. What was the context? Was it a special day or month?
 - c. Where were you when you were told that story? Inside the Flathead reservation?
 - i. Are there any special remarks in that place? At **Kicking Horse** with a view to **Eagle Pass** Mountain or a specific lake for instance. **Medicine Tree**?
 - d. How were you influenced by that story?
 - i. What insight/lessons did it give to you? Weather patterns?
 - e. Can you tell that story? Go into as much detail as possible.
 - i. Which cultural elements (e.g., Thunderbird, clothes, animals) were present in the story?
 - ii. Can you go through the whole process of the **Canadian Geese** migration?
 1. In which month do they fly south?
 - a. To where exactly? Mexico?
 2. And back north?
 3. Is there any specific event happening before doing so?
 - a. Do you know a possible reason why they return after a winter snowstorm?

Salish, Pend d'Oreille, and Kootenai (immersive) Storytelling

Interview Questions (2 of 2)

3. Do you know anything related to **Bird's Sock**, or the **Thunderbird**? What about the **coyote, howl**, or other mischievous beings? Howl?
 - a. Why does it have that name?
 - b. What does it represent?
4. Do you remember similar stories?
 - a. Any of them related to **weather patterns**, or other methods you used to **forecast**? Any related to **bears** hibernating as a sign for people to gather stuff because winter is coming?
5. What other **cultural elements, symbols, rituals** and **gestures** can you share with us?
 - a. How would you best represent the **sky**? And a **bird**?
 - b. How would you best represent the **changing of seasons** (e.g., from winter to summer by using a gesture)? Sort of changing the chapter with your hands.

B.5.1 Interview with Salish Student

Location: CSKT, Montana, USA

Date: 6th March, 2019

Hour: 3:14pm

Timezone: Mountain Time (UTC-6)

Duration: 26 minutes and 52 seconds

Transcript

Interviewer: So, at any point, if you wanna pause or cancel, just let me know. We will identify the interview with a random ID, so that this is totally anonymous.

Salish Student: Alright.

Interviewer: What can you tell us about Storytelling? In general. How important is it in your culture?

Salish Student: It's important in our culture because stories have always taught about mischief, about strength, and growing. A big thing in our culture is 'coyote' stories. They are about life lessons, and the coyote is always... he's not a bad character. He's not evil. He's just mischievous. In some cases he can be bad. But, it's kinda like the stories of him.... You know, trying to... the bird will catch the snake. When the bird tries to catch the snake, the coyote gets the bird. The snake bites the bird. You know, it's a cycle. It's the circle of life through one story. Coyote stories are taught to the youth, to really... It solidifies the culture. It's our main teaching point. Cause, you know, when the culture was being 'produced', it wasn't about going to a school. You were taught everything by experiencing it. The coyote stories can only be taught, they can

only be told, after the first snowfall. They can't be told after the trees pop in the Spring. So our cultural belief is that you can only tell them during heavy winters.

Interviewer: So basically, when there is snow?

Salish Student: Yes. There is a time period to tell them and those are really, kinda, the down-months. They are mainly a teaching point, just to get on the right path.

Interviewer: Is there any real reason why you don't tell the stories in any other month?

Salish Student: It's mainly just cultural beliefs. We got many bad omens. Not a lot, but. An omen in native culture is owls. Owls are a very bad sign. So, owls are messengers. If you see an owl in your path, at night, it's delivering a message. If you see an owl during the day it means someone is passing. They are messenger birds. When you have these beliefs, the stories can only be told in a certain time because of the cultural way they hold, as far as the bad luck around gather-months.

Interviewer: That is similar to Portugal. I think owls symbolise death.

Salish Student: We have the same belief. I have a hat with an owl I wear a lot, and it's not appreciated by some people. For me it's a mascot, sometimes. But my cultural belief is that it is still bad. It's keeping your culture connected, but also being associated with the real world.

Interviewer: As to why you do storytelling, in general, I guess you always answered.

Salish Student: Yeah, it's mainly a teaching point.

Interviewer: It's the way people transmit knowledge from generation to generation, right?

Salish Student: Yeah, and it puts it in a way that it will stick with you. Because I still remember growing up and going to cultural camps, and such, and coyotes' stories always stuck with you. So, you know, it's that aspect of... it's the respect that comes with them, and just, you know, in our culture, it's holding your beliefs so strong, so that you don't tell them in summer, or spring.

Interviewer: Are they usually real stories?

Salish Student: They are associated with the Creator. They are real settings, sometimes real situations, but it's added in to get the aspects of the mythicity and be a teaching point.

Interviewer: Is there a special day, a holiday, to tell certain stories?

Salish Student: No. It's just in that time period. And they are usually told at gatherings between the people of the tribe.

Interviewer: And just to make sure, you are from Salish, right? Bitterroot...?

Salish Student: Yes. Salish Bitterroot. They were originally residing in the Bitterroot Valley, and then the government decided to... Salish were forced to go all through Missoula, and this is where we reside now. So, this wasn't the original land. And if you go down there, you will find the 'Madison Tree', a very cultural spot. It's kinda on the side of the highway, which is sad to see.

Interviewer: Who would be eligible to listen to those stories? The whole village, or the whole tribe?

Salish Student: They are mainly told towards all the youth. There is no script for coyote stories. Every person has different outlooks. I don't know about other people coming in, as we learn them in schools within the reservation. Other tribes have their own thing. In Dakota they have their own mischief maker. A lot of tribes use the coyote, but others use different animals.

Interviewer: Do you remember the game from yesterday? I think they were using the coyote there, no? To help, though. Or was it a fox?

Salish Student: Oh. That was a fox. Having them play a role in the game is also a teaching point. The mischief maker, if you do veer off the path, the coyote would be drawing you, so.

Interviewer: Do you know the child book story, Luli's Journey?

Salish Student: Yes.

Interviewer: How old were you when you first heard it?

Salish Student: I was a grad schooler in an american studies class.

Interviewer: And who told you that story?

Salish Student: The teacher.

Interviewer: So it wasn't an elder.

Salish Student: No.

Interviewer: Was the teacher from the same tribe?

Salish Student: Yes.

Interviewer: Can we contact him or her?

Salish Student: No... If you go to any local school, you will find books, and you will find multiple teaching styles.

Interviewer: What was the context? Was it a special day, or month? Probably not.

Salish Student: No.

Interviewer: Where were you when you were told that story?

Salish Student: At the school, within the flathead indian reservation.

Interviewer: How were you influenced with that story?

Salish Student: When you are that young... what it did was strengthen our salish. You go through the basics of colours, numbers, and then stories in salish. It helps strengthen our language.

Interviewer: So you don't really recall the idea that back then native people used the birds' migration to predict weather patterns?

Salish Student: Oh, yeah. That definitely popped through, but when you are that young you don't really grasp the concept.

Interviewer: Can you tell that story?

Salish Student: Nahh...

Interviewer: Is there any particular meaning you give to the Geese? Canadian Geese.

Salish Student: Not that I was raised with.

Interviewer: Do you know about the Geese constellation?

Salish Student: Yeah, but I was just not raised with that strong belief of it.

Interviewer: Can you walk me through the migration process, however?

Salish Student: Yeah, so. When the native people picked on the migration of the Geese, that is when they would usually start kicking in for the final of preparing blankets, food, shelter. Sometimes tribes moved, but not a long distance. But that's really when they would settle down for the winter. They would prepare.

Interviewer: Is there any specific event happening before they migrate?

Salish Student: No.

Interviewer: Do you know of a possible reason they will return after winter?

Salish Student: No.

Interviewer: Do you know about the totems?

Salish Student: No.

Interviewer: The thunderbirds?

Salish Student: We don't do totems in this reservation. That's like Pacific tribes.

Interviewer: Aren't there any totems in the reservation?

Salish Student: We got sculptures, but we are not a totem tribe. That is how everyone perceives tribes. Riding horses through the planes. Living in teepees. Everyone huntings, everyone warriors. When in reality you had the pacific natives who were totem, everyone gathering, fishing, boats. They are not in the planes riding horses. Everyone assumes every tribe is just jumping on a horse and... our tribe wasn't... It wasn't like a rodeo.

Interviewer: Yeah, depending on what you have around.

Salish Student: Exactly, the resources that nature provides.

Interviewer: Do you remember any similar stories?

Salish Student: I believe there is one about the bears hibernating. I can't remember the exact title.

Interviewer: So they used that information, somehow?

Salish Student: Because when the bears go away, it's because they are preparing for winter. When you see the bears preparing, you prepare. There are a lot of stories about the snow coming down the mountains, because you can tell... in the summer... were you here during last summer?

Interviewer: No.

Salish Student: so, during the summer you can tell the snow hits the peaks. You can tell it gets colder. As the winter gets closer, the snow moves slowly down the mountain, and then it's like a blanket. It's just a blanket of snow across the mountain. There are multiple ways to tell. That's like, in the spring you can tell by the lake, because of the runoff.

Interviewer: What other symbols, rituals, gestures, can you share?

Salish Student: Intertwined with the story?

Interviewer: Yes. For instance, do you know the 'growing' gesture? Apparently growing is like this (shows).

Salish Student: Ah, yeah. I know, but I can't, like, name them. How I learn is tactile. What I really think is cool about you doing this for our culture, is because kids, growing up and stuff, it fades out after a while. Right now we are talking about this, and it's coming back, but until then I'm sitting, like, I'm disconnected of a sort.

Interviewer: Because you are not really aware you are doing that?

Salish Student: Yeah.

Interviewer: How would you represent a 'Yes'?

Salish Student: Fist up.

Interviewer: Fist up?

Salish Student: Yeah, as in the stick game. Or thumbs up, in general.

Interviewer: How about 'No'?

Salish Student: Just nodding.

Interviewer: And 'sky'?

Salish Student: It's just up. With both hands.

Interviewer: I think the american sign language one would be with both hands up, going opposite ways. Would that make sense to you?

Salish Student: Oh, yeah. It could be 'open', too. But if you look up and do that, yes.

Interviewer: How about flying?

Salish Student: You can do something like with your thumbs, moving over each other. A simple flick.

Interviewer: How about getting back down?

Salish Student: Maybe by pulling it to you, almost. If you are in the sky, you can pull it up to you.

Interviewer: And bird?

Salish Student: I thought the fly was kinda the same. I would do the same with the thumbs over each other, moving. Because you always connect birds to flying.

Interviewer: Yesterday I remember you talking about some songs that are appropriate for the summer or the winter.

Salish Student: Yes.

Interviewer: Is that a Salish thing?

Salish Student: It's a multiple tribe thing. There are certain songs for certain times. The grass songs wouldn't be played in the winter because there is... no grass... It's not a big thing, more of a general thing.

Interviewer: But do you think that would help give more context to the setting?

Salish Student: Yes, it might.

Interviewer: How do you best represent the change of seasons?

Salish Student: A gesture or animation?

Interviewer: Could be a pose or an 'animated' gesture. It doesn't have to be static or anything.

Salish Student: O.K. Like you do turning a page. Like the next chapter to you. Because you would like, select a chapter looking at the mountain, almost. It's probably hard, but if you could like, as you are flipping, change the landscape.

Interviewer: Half half?

Salish Student: Not as your hand goes, but, like, progressed. Little progress. Like trees popping, or something. Not like a sudden change. Like the whole cycle. Fall leaves drop, etc. The snow, not like, showing up from nowhere, but progressing.

Interviewer: I have also heard, on Monday, that the Salish use the term 'praying' as a way of wishing luck to someone. There is a praying word, even though there is no ritual. (shows)

Salish Student: Yeah, that's 'on your way'.

Interviewer: And the other person kind of waves?

Salish Student: Yeah, sort of a gesture to say 'thank you'.

Interviewer: But you use the term 'praying', not giving or wishing luck, correct?

Salish Student: Yeah. Because you are praying to the Creator. Like, before our meals, we pray for thanks, we pray for safe travel, you pray for guidance, safekeeping.

Interviewer: Can you go to someone and ask to pray for you? Is that how the approach goes?

Salish Student: Usually someone notices and they will pray for you. Put you back on your way.

Interviewer: Alright. Cool. Can I stop the interview?

Salish Student: Yup.

B.5.2 Chat with Professor Frank Finley

Location: Facebook Messenger

Date: June 17, 2021

Hour: 6:40 pm

Timezone: (UTC+1)

Duration: 7 minutes

Transcript

Interviewer: Hello! I was rereading one of the children's books I found there, and there is a particular illustration showing some birds (*q.v.* Figure B.1), but I was wondering about their species. Do you happen to know the one on the bottom left corner?

Prof. Frank: It's a mountain bluebird. They are tiny little birds. The male is iridescent blue, very pretty.

Interviewer: Ah, great! What about on the right? It seems to be an owl and a magpie?

Prof. Frank: Looks like it to me too.

Interviewer: Thanks!

Prof. Frank: I'm not sure where the peacock came from...

Interviewer: I found it weird too. Not so common around there is it?

Prof. Frank: Not at all.

Interviewer: The dialogue mentions big and small birds.

Prof. Frank: Probably just to make a point.

Interviewer: Thanks for also confirming that.

Interview related picture



Figure B.1 – Picture from Luli's Journey book that lead to chatting with Professor Frank Finley.



C.1 Narrative Details

Place:

- South Crow Creek Area, south of Ronan, Montana, near Kicking Horse Reservoir, located on the Flathead Indian Reservation
 - Environment can have text annotations for specific places, like eagle's pass; maybe the house names (as per the drawing; the lake; etc.

Characters:

- User (part of the tribe or not; depends on the culture)
 - Full body avatar with inverse kinematics (IK)
 - Maybe make the user choose a set of hands he/she wants to use
- Elder or someone from the tribe eligible to tell the story (even teachers of the Salish language tell them, so it's not something only elders can do)

- The story seems to be Salish, so it makes sense to have the narration in Salish (at least for this first version)

Other characteristics:

- **Interactive storytelling** (mixed with **animated storytelling**);
- Scenario starts in broad daylight. And changes accordingly, so that the story can be told. The environment can also change as the story progresses;
- **Stational experience/tracking**, so that elderly can be targeted, too. If we have the time, we can try to implement a good mechanism for locomotion, etc.

C.2 Narrative Script

Story:

1. Elder introduces himself and the tribes
 - (a) “Hello! My name is Michel, and I am going to tell you about the Confederated Salish and Kootenai Tribes. This includes the Salish, Kootenai, and Pend d’Oreille tribes. Pend d’Oreille are also known as Kalispel”
 - (b) Asks the user’s name: “What is your name?”
 - i. The user needs to answer with voice.
 - ii. The name can then be used (at least in subtitles, or so)
 - (c) “Nice name!”
2. “Before going any further, let me just tell you a bit about complex systems.”
(remove from the script after first iteration);
3. Go over complex systems. (removed from the script after first iteration)

- (a) “The world is an extremely complex place and has been rapidly evolving for some decades, consisting of complex and interconnected systems, creating novel challenges that call for a new way of thinking. Ecosystems, societal issues, politics, weather, and the human body are examples of such complex systems. They are complex – not to be confused with chaotic – in the sense that they act in interdependent and unpredictable (non-deterministic) ways. A key concept that makes complexity diverge from chaos is emergence. Simply stated, some features that emerge in large systems cannot be traced down to lower components, causing uncertainty to be inherent. Examples of such properties are the behaviours of water or of a bird flock. Even if initial conditions of individual birds or water molecules are known, the behaviours seen afterwards – of the whole – cannot be predicted directly such as flight formation, and surface tension; rather, they have emerged from interactions of individual system components. Complex systems, or the study thereof, may hold the key to our understanding in this regard. Complexity science replaces the reductionist paradigm with a more holistic approach.”
- (b) “I hope that wasn’t too boring for you. What is important to understand is that the traditional divide and conquer approach is not always the answer. There are many things that cannot be explained in such a way.”
- (c) “Now a bit more about the tribes.”

4. Elder tells the location of the experience.

- (a) “Salish were originally located in the Bitterroot valley, but were forced to move here for political reasons. Anyway, you and I are now at South Crow Creek Area, south of Ronan, Montana, near Kicking Horse Reservoir, located on the Flathead Indian Reservation, where Elders from the tribes used to tell stories, through what is known as storytelling.”
- (b) Tell the user to look into the mountains (Eagle’s pass)
 - “Look at the mountains over there. That little V on top is known as Eagle’s pass.”;

- The User needs to look into a specific part; only then the experience progresses.
5. Elder shows some pictures of the reservation (sort of an interactive ring menu)
 - (a) “Great. Now take a look into these pictures to see what our land actually looks like!”
 6. Elder mentions the importance of the land, wildlife, and water. They take care of it.
 - (a) Talk
 - i. “We take from the land, wildlife, and water, so it’s only right to give it back in some shape or form”;
 - ii. Show the example of structures to help bears cross a road (bridge or tunnel);
 - A. “An example of that is trying to not disrupt the wildlife. We build bridges and tunnels so that bears and others can cross roads safely”;
 - B. A hologram showing up with a small road and a bridge, with a car and bear passing by.
 7. Elder mentions the importance of art, such as paintings/drawings and colours, music, etc.
 - (a) “We also give utmost importance to drawings, paintings, colours, and music. They are a big part of our culture. They show our vision of the world and our surroundings”;
 - (b) Show in sort of an interactive ring menu with some photos of paintings, etc., taken there.
 8. Elder explains the importance of storytelling in Salish (and other tribes) culture
 - (a) “Stories always tell about mischief, strength, and growing (e.g., a tree)”;

- (b) “Firstly, they were (and are) a way of conveying an idea or teaching, often in a nonlinear way.”
 - i. Can perform certain gestures here to represent these. Closing hand with palm facing upwards.
 - (c) “A single story often tells about the circle of life. Today is no different!”
 - i. Maybe animation with VFX - particle going around in a circle.
 - (d) “They solidify the culture.”;
 - (e) “They are a way of transmitting knowledge from generation to generation”;
 - (f) “There were no schools back then; you would learn by experiencing it.”;
 - (g) “An important remark to make is that these stories are not necessarily scripted. Each tells them their own way, with their own words.”;
 - (h) “Another is that there might be specific scenarios that define when and where these stories can/should be told.”;
 - (i) “A good example of this are the Coyote stories.”
 - i. ”They are about life lessons.”;
 - ii. (The coyote can be represented as a real animal in the scene, or through a spirit-like animation floating around.);
 - iii. “The coyote, for the Salish (other tribes might have different animals), is not necessarily a bad character (he can be). He’s more of a mischievous character.”;
 - iv. “They are a teaching point to get in the right path.”;
 - v. “They can only be told in a certain time period, particularly the “down months”. That is, during winter (with snowfalls). They cannot be told after the first bloom of Spring.”
- A. Reason:
- A. “This is because of Cultural beliefs”;
 - B. “Salish, and other tribes, have signs of bad omens. One of which is the Owl-sign of misfortune, death. They are messaging birds.”

C. “Other stories might be told in some specific days, depending on the memories they bring. Context matters!”;

D. “They are more than stories. They are full of meaning.”

9. Elder Talks

(a) “Now that we went through a bit of the Salish culture and the importance of storytelling, it’s time to set an appropriate environment to go through a story ourselves!”

i. Gesture to change from broad day light to night time—the fire starts here, too.

A. (Environment changes);

B. Certain background song starts.

(b) “This is based on the children’s book story ‘Luli’s Journey’ about a little Canadian Goose called Luli, her family, and their journeys.”

i. Show picture of book while talking:

A. “It was written by Alyssa Kelly and Jess Nenemay, and translated by Pat Pierre. It was partly funded by the Administration for Native Americans.”.

10. Elder Talks

(a) “There are many birds on the reservation. Some are big and some are small, but a Canadian goose named Luli is different from them all.”

i. Ask the user to draw a bird, by giving her/him a frame to do so: “Can you draw a bird?”;

ii. The drawing in mid-air can be done with pinch gestures;

iii. Show a hologram of a Gosling, when the drawing is detected as a bird.“.

(b) “Luli hatched from an egg. Her nest is on the shore of McDonald Lake. Her mother and father were also born there. They raised Luli with love and care.”

i. Show a hologram with father, mother, and Luli as gosling.

- (c) “Summer brings a lot of fun as Luli learns to fly and run. She wades in the water as she follows her father and eats berries or plants her mother brought her”
 - i. show a hologram with father, mother, and Luli a bit older.
- (d) “Luli has grown so big and so strong, her mother says it won’t be long until their journey to the land of the sun.”
 - i. Ask the user to make the goose bigger: “Can you make the goose bigger?”
 - A. General gesture would be zoom in/zoom out or using both hands;
 - B. The cultural one for Salish would be the ‘growing’ gestures shown before by the elder;
 - C. Show an hologram of a bigger goose
- (e) “Luli’s father recalls the autumn days, before Kerr Dam was built when he and his friends played on the muddy flats of Flathead Lake.”
 - i. Leaves falling (autumn like);
 - ii. Sound of leaves and wind.
- (f) “Long before winter’s frost blows in, Luli prepares to leave her friends (some birds are resident. They do not migrate). She ruffles her feathers and away she goes. Maybe she will travel to Mexico.”
 - i. Show map of North America with POIs being Montana and Mexico;
 - ii. Another option here is to show a timeline where the user can then drag the goose to.
- (g) “Can you make Luli fly?”
 - i. Hologram of Luli appears, with some indication pointing upwards;
 - ii. User performs cultural gesture of lifting the bird onto the sky.
 - A. General gesture could be grabbing the bird in direction to the sky.
 - iii. Elder points/looks to the sky as animation goes from the fire to the sky (cueing for some event happening in the sky instead);
 - iv. User needs to look to a specific location, too, as the Geese start showing up in a V formation.

- A. This happens by showing some sort of arrow to show where to look at, *etc.*
- v. User needs to look back to fire/elder
- (h) “Luli’s family stays together in flight, they fly in a ‘V’ through the day and the night. People watch from the earth below. Canadian Geese let them know: if the summer will be hot and really dry or if the winter will be harsher as the days go by.”
 - i. Snow starts falling (next to the user and fire); environment starts changing.
- (i) “Weeks of hard flying have come to this: Luli arrives to sun and bliss. Mexico has warm sand and an ocean breeze, but Luli was excited most to see the palm trees.”
 - i. Show hologram of Palm tree / beach diorama;
 - ii. Play some sound of waves.
- (j) “Months have passed on the paradise shore, but Luli feels like she needs so much more. She remembers her nest near McDonald Lake; the mountains and trees were her favorite place.”
 - i. Environment starts changing to Spring (not as much snow; trees will start popping);
 - ii. Change to a more appropriate song for spring; birds chirping, etc.
- (k) “Luli’s parents tell her winter has passed and it’s time to return to their home at last. So they ruffle their feathers and flyaway north in a ‘V’, each taking their turn to be in the lead.”
 - i. Show globe/map with an animation from Mexico to the reservation (a simple line suffices).
- A. Or again timeline, where the user needs to drag goose from one location to another.
 - ii. Elder looks to the sky again;
 - iii. User needs to look as well;
 - iv. Birds show up flying in V again.

- (l) “Spring melts the snow and flower buds grow. Luli’s family arrives after winter’s last blow. The time when birds arrive after a late winter snowstorm is called Bird’s Sock.”
 - i. No more snow in the environment.
- (m) “Wild flowers bloom on the mountainside and blue skies make the earth come alive. Luli and her friends have all grown, but her lakeside nest will always be home. Someday, Luli will lay eggs of her own and she will take them on the same journey that she has flown.”;
- (n) “The birds’ knowledge of seasons is so precise that the Native people depended on them to survive. The Canadian Geese’ annual migrations gave the people an idea of weather patterns as well as timelines for season changes.”;
- (o) “Tribes look forward to birds as harbingers of Spring.”

11. Elder asks

- (a) “Would you like to know more about the information the Native people gathered from observing the birds and the environment?”
 - i. User needs to answer with a gesture for yes or no
 - A. Salish would be thumbs up (or fist up);
 - B. General would be maybe a ‘ring’ (done with thumb and index fingers).
 - ii. If the answer was no; Move to next question.
 - iii. If the answer is yes:
 - A. “Cold weather triggers migrations.”
 - A. Wind blowing (with sound effect, too).
 - B. When the birds fly to the South, that would mean the tribes needs to gather food, shelter, and settle down for Winter.”;
 - C. “They can use the Sun, and the stars as guidance.”
 - A. Show hologram of Sun and stars.
 - B. “They can also use landmarks, such as coastlines, rivers, or mountain ranges.”;

- C. "They can even use Earth's magnetic field!";
- D. "They can go as far as Brazil or Argentina. But most get to Texas or Mexico, which will already be considerably warmer.";
- E. "Tribes also used other information available in the environment.";
- F. "When bears start disappearing; when they start hibernating, it's because they are also preparing for winter.";
- G. "You can also tell it is getting colder by watching the mountains. The peaks will start to get white first, and it will slowly go down the mountain, like a blanket.";
- H. "Hopefully this will give you some insight into using the land, and the wildlife, as part of your life."

12. Elder asks

- (a) "Would you like to know more about the emergent nature of V-shaped flights?" (Yes/No)
 - i. Show the birds in a V-formation as a hologram next to the fire.
- (b) "As each bird flaps its wings, it creates an uplift for others behind him. There is 71 percent more flying range in V-formation than flying alone.";
- (c) "Whenever a goose flies out of formation, it quickly feels the drag and tries to get back into position";
- (d) "When the lead goose gets tired, it rotates back into the formation and another goose flies at the head";
- (e) "The geese in formation honk from behind to encourage those up front to keep up their speed";
- (f) "When a goose gets sick or wounded and falls, two geese fall out and stay with it until it revives or dies. Then they catch up or join another flock."

13. Elder asks

- (a) "That is it for now! Would you like to explore this environment freely? (Yes/No)"
 - i. * Allow exploring

14. Elder says

- (a) “Hope we can meet again soon!”;
- (b) “Goodbye”;



Machine Learning Detailed Models

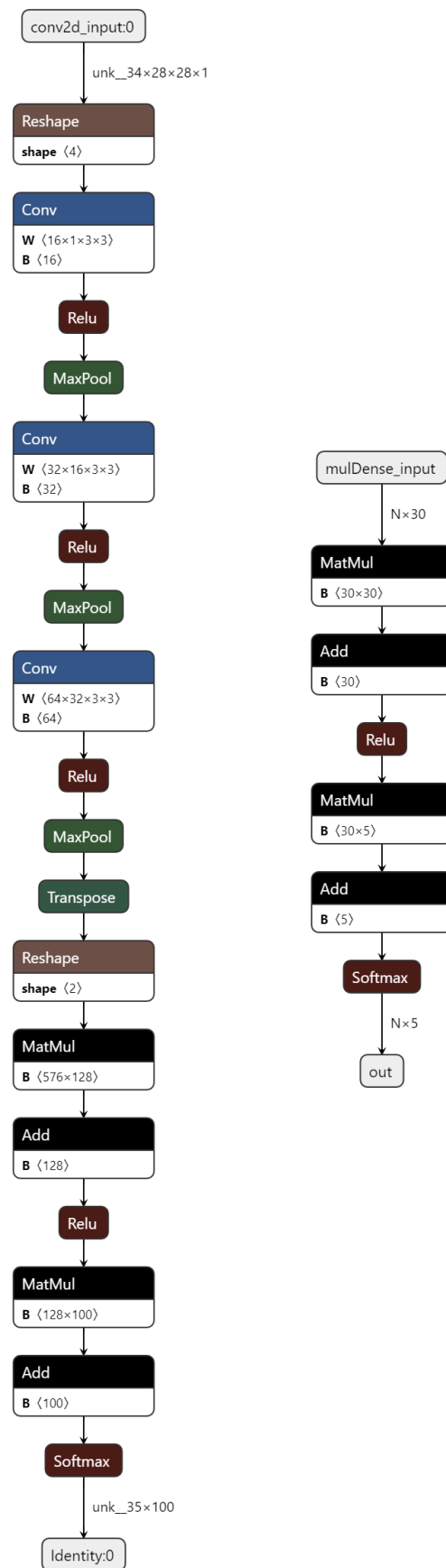


Figure D.1 – Detailed Drawing (left), and Gesture (right), Machine Learning Model Views using Netron¹.

¹ <https://netron.app/>



User Trials Data Compilation

E.1 Observational Compiled Data

Table E.1 – Observation compilation of User's one experience.

Task ID	Task Action	or	PT	Salish	User 1	Pre-Test Answer	Observation
A1	Look around				YN		
A2	Hello		Wave	Wave	YN		
T1	Grab microphone	Mi-	Close hand in a fist	Open hand slightly	YYY	One hand closing	
T1.1	Talk				YN		
A3	Give microphone	Mi-	Drop/throw micro-phone back	Drop/Throw micro-phone back	YY		Drop

Table E.1 continued from previous page

Task ID	Task Action	or	PT	Salish	User 1	Pre-Test Answer	Observation
T2	Look at the V-shaped peak				YY		
T3	Look back at narrator				YY		
T4	Look at reservation pictures				YY		
T4.1	Interact with pictures	Swipe left or right	Swipe left or right		YYN	Grab them, in turns, and switch their place	Now used one hand to hold another to swipe
A4	Look at car and bear demonstration				YY		
T5	Look at art pictures				YY		

Table E.1 continued from previous page

Task ID	Task or Action	PT	Salish	User 1	Pre-Test Answer	Observation
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YYN	Grab them, turns, and switch their place	Now used one hand to hold another to swipe
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	NYN	Both hands in and out (from the inside to the outside, and vice versa)	Used hand slightly facing upwards with fingers almost touching the palm
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NYN	Hold it with the index, and thumb fingers	

Table E.1 continued from previous page

Task ID	Task or Action	PT	Salish	User 1	Pre-Test Answer	Observation
T7.1	Draw bird	Drawing motion	Drawing motion	YYY	Hold it the same way and move the hand	
A6	Look upwards when narrator points to the sun			YN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	NYN	Stretch back and forth the upper limbs	Used hand slightly facing upwards with fingers almost touching the palm
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in/out with both hands	NYN	Both hands in and out (from the inside the outside, vice versa)	Used hand slightly facing upwards with fingers almost touching the palm

Table E.1 continued from previous page

Task ID	Task Action	or	PT	Salish	User 1	Pre-Test Answer	Observation
T10	Make fly	Luli	Grab it with one hand and move it upwards	Lift with one or two hands, from below	NY Y	Pinch gesture (index and thumb) and throw as it was a paper plane	
T10.1	Look at the Geese in the sky				YY		
A7	Look at Luli at the beach				YY		
T11	Look at light leading to the sky				YN		
T11.1	Look at Geese flying back				YY		
A8	Look at Geese and bear in lake				YY		

Table E.1 continued from previous page

Task ID	Task or Action	PT	Salish	User 1	Pre-Test Answer	Observation
T12	Yes/No choice for further info on in-sights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Thumbs up; or Move index finger left and right	
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		
A12	Look at the mountains snow coming down like a blanket			YN		

Table E.1 continued from previous page

Task ID	Task or Action	PT	Salish	User 1	Pre-Test Answer	Observation
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Thumbs up; or Move index finger left and right	
A13	Look at Geese flying at different speeds			YY		
A14	Look at goose going out of formation and recovering			YY		
A15	Look at the Geese formation's leader rotating back when it gets tired			YY		

Table E.1 continued from previous page

Task ID	Task or Action	PT	Salish	User 1	Pre-Test Answer	Observation
A16	Look at the Geese flying and motivating the ones in front to continue			YY		
A17	Look at a goose falling and two others staying with him until he recovers			YN		
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Thumbs up; or Move index finger left and right	

Table E.1 continued from previous page

Task ID	Task or Action	PT	Salish	User 1	Pre-Test Answer	Observation
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	YYY	Fingers snap	
A20	Goodbye	Wave	Praying or Wave Gesture	YY		

Table E.2 – Observation compilation of User two's experience.

Task ID	Task Action	or PT	Salish	User 2	Pre-Test Answer	Observation
A1	Look around			YY		This user explored the surrounding environment more. Always looking around, etc. The user understood english, and didn't really need the subtitles.
A2	Hello	Wave	Wave	YN		
T1	Grab microphone	Close hand in a fist	Open hand slightly	YYN	Both hands closing	

Table E.2 continued from previous page

Task ID	Task Action	or PT	Salish	User 2	Pre-Test Answer	Observation
T1.1	Talk			YY		But said her name before having the microphone. Showed some confusion after grabbing, almost talking again.
A3	Give microphone	Mi-Drop/throw micro-phone back	Drop/Throw micro-phone back	YY		Drop
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YN		
T4	Look at reservation pictures			YY		

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
T4.1	Interact with pictures	Swipe left or right	Swipe left or right	YYY	Full hand swipe	
A4	Look at car and bear demonstration			YY		Nods as if agreeing
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YYY	Full hand swipe	
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	YNY	Pinch (index and thumb fingers)	

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	YYY	Hold it with the index, middle, and thumb fingers	
T7.1	Draw bird	Drawing motion	Drawing motion	YYY	Hold it the same way and move the hand	
A6	Look upwards when narrator points to the sun			YN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	NYN	One hand lifting	Used one hand closing
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in/out with both hands	YNN	Pinch (index and thumb fingers)	

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
T10	Make fly	Luli	Grab it with one hand and move it upwards	Lift one or two hands, from below	NYN	Both hands interconnected to make a butterfly motion
T10.1	Look at the Geese in the sky			YY		
A7	Look at Luli at the beach			YY		
T11	Look at light leading to the sky			YY		
T11.1	Look at Geese flying back			YY		
A8	Look at Geese and bear in lake			YY		

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right	
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
A12	Look at the mountains snow coming down like a blanket			YN		The user actually looked to the back, where there are no mountains
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right	
A13	Look at Geese flying at different speeds			YY		
A14	Look at goose going out of formation and recovering			YY		

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
A15	Look at the Geese formation's leader rotating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		
A17	Look at a goose falling and two others staying with him until he recovers			YY		

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right	

Table E.2 continued from previous page

Task ID	Task or Action	PT	Salish	User 2	Pre-Test Answer	Observation
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	NYN	Rubbing palms as if something doesn’t exist any-more	She nods as if understanding when the narrator performs the changing of seasons gesture. For the exploration part, the user performed the semi-circle with one hand (not totally like the salish)
A20	Goodbye	Wave	Praying or Wave Gesture	YY		

Table E.3 – Observation compilation of User three's experience.

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		
T1	Grab microphone	Close hand in a fist	Open hand slightly	YYY	One hand closing	
T1.1	Talk			YY		
A3	Give microphone	Drop/throw micro-phone back	Drop/Throw micro-phone back	YY		Drop
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YY		
T4	Look at reservation pictures			YY		
T4.1	Interact with pictures	Swipe left or right	Swipe left or right	YYY	Full hand swipe	

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
A4	Look at car and bear demonstration			YY		
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YYY	Full hand swipe	
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	NYN	Pinch (index and thumb fingers)	
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NYY	Hold it with the index, middle, and thumb fingers	

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
T7.1	Draw bird	Drawing motion	Drawing motion	NYN	Hold it the same way and move the hand	She kept the holding position but didn't move
A6	Look upwards when narrator points to the sun			YN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	NYY	Two hands lifting at the same time	
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in/out with both hands	NYN	Pinch (index and thumb fingers)	Didn't pinch however pointed index finger
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	NYY	Throw it frontwards, or "carefully lift" it (like a bird)	Careful lift but not a full grab

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
T10.1	Look at the Geese in the sky			YY		
A7	Look at Luli at the beach			YY		
T11	Look at light leading to the sky			YY		Not until end
T11.1	Look at Geese flying back			YY		Looks around because of audio
A8	Look at Geese and bear in lake			YY		
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Aloud, always accompanied with nodding up and down; or Nodding left and right	She does both thumb action and nodding without voice

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		
A12	Look at the mountains snow coming down like a blanket			YY		

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Aloud, always accompanied with nodding up and down; or Nodding left and right	Just nodding without voice
A13	Look at Geese flying at different speeds			YY		
A14	Look at goose going out of formation and recovering			YY		

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
A15	Look at the Geese formation's leader rotating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		
A17	Look at a goose falling and two others staying with him until he recovers			YY		

Table E.3 continued from previous page

Task ID	Task or Action	PT	Salish	User 3	Pre-Test Answer	Observation
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YNN	Aloud, always accompanied with nodding up and down; or Nodding left and right	
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle		Left hand move to the left side for past, and right hand move to the right side for future	
A20	Goodbye	Wave	Praying or Wave Gesture	YY		With both hands

Table E.4 – Observation compilation of User four's experience.

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		
T1	Grab microphone	Close hand in a fist	Open hand slightly	YNN	Both hands closing	
T1.1	Talk			YN		
A3	Give microphone	Drop/throw micro-phone back	Drop/Throw micro-phone back	YNN		
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YY		
T4	Look at reservation pictures			YY		
T4.1	Interact with pictures	Swipe left or right	Swipe left or right	YNN	Index finger from right to left (sort of swipe)	

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A4	Look at car and bear demonstration			YY		
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YNN	Index finger from right to left (sort of swipe)	
A5	Look at coyote			YN		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	YNN	Both hands in and out (from the inside to the outside, and vice versa)	

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	YNN	Hold it with the index, middle, and thumb fingers	
T7.1	Draw bird	Drawing motion	Drawing motion	YNN	Hold it the same way and move the hand	
A6	Look upwards when narrator points to the sun			YNN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	NY Y	Two hands lifting at the same time	

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in-/out with both hands	YNN	Both hands in and out (from the inside to the outside, and vice versa)	
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	YNN	Two hands moving as if they were wings	
T10.1	Look at the Geese in the sky			YY		
A7	Look at Luli at the beach			YY		
T11	Look at light leading to the sky			YY		Does not see the end and looks after
T11.1	Look at Geese flying back			YY		

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A8	Look at Geese and bear in lake			YY		
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right with the eyes moving, too	Both nods and thumbs up
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YN		

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A12	Look at the mountains snow coming down like a blanket			YY		
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right with the eyes moving, too	Both nods and thumbs up
A13	Look at Geese flying at different speeds			YY		
A14	Look at goose going out of formation and recovering			YY		

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A15	Look at the Geese formation's leader rotating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		
A17	Look at a goose falling and two others staying with him until he recovers			YY		

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right with the eyes moving, too	
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	YYY	Line in front with left hand side being past, middle being the present, and right hand side being the future	

Table E.4 continued from previous page

Task ID	Task or Action	PT	Salish	User 4	Pre-Test Answer	Observation
A20	Goodbye	Wave	Praying or Wave Gesture	YY		

Table E.5 – Observation compilation of User five's experience.

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
A1	Look around			YN		Arms closed position for listening
A2	Hello	Wave	Wave	YN		
T1	Grab Microphone	Close hand in a fist	Open hand slightly	YYY	Stretch the arm fully and close hand	
T1.1	Talk			YN		
A3	Give Microphone	Drop/throw microphone back	Drop/Throw microphone back	YY		Drop
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YY		

Table E.5 continued from previous page

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
T4	Look at reservation pictures			YY		
T4.1	Interact with pictures	Swipe left or right	Swipe left or right	YYN	Look at one, then the other	Swiped
A4	Look at car and bear demonstration			YY		
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	NYY	Look at one, then the other	
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	YYN	Pinch (index and thumb fingers)	

Table E.5 continued from previous page

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NY Y	Hold it with the index, middle, and thumb fingers	
T7.1	Draw bird	Drawing motion	Drawing motion	Y Y Y	Hold it the same way and move the hand	
A6	Look upwards when narrator points to the sun			Y Y		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	Y Y N	Water it	
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in/out with both hands	Y Y N	Pinch (index and thumb fingers)	

Table E.5 continued from previous page

Task ID	Task Action	or	PT	Salish	User 5	Pre-Test Answer	Observation
T10	Make fly	Luli	Grab with one hand and move it upwards	Lift one or two hands, from below	NYN	Draw some wings	Both hands upwards
T10.1	Look at the Geese in the sky				YY		
A7	Look at Luli at the beach				YY		
T11	Look at light leading to the sky				YY		
T11.1	Look at Geese flying back				YY		Keeps looking around
A8	Look at Geese and bear in lake				YY		

Table E.5 continued from previous page

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	NYN	Nodding up and down; or Nodding left and right.	Grabs microphone
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		
A12	Look at the mountains snow coming down like a blanket			YN		

Table E.5 continued from previous page

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right.	
A13	Look at Geese flying at different speeds			YY		
A14	Look at goose going out of formation and recovering			YY		
A15	Look at the Geese formation's leader rotating back when it gets tired			YY		

Table E.5 continued from previous page

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
A16	Look at the Geese flying and motivating the ones in front to continue			YY		
A17	Look at a goose falling and two others staying with him until he recovers			YY		
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right.	

Table E.5 continued from previous page

Task ID	Task or Action	PT	Salish	User 5	Pre-Test Answer	Observation
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle		Flip of right hand, starting with palm up position	
A20	Goodbye	Wave	Praying or Wave Gesture			

Table E.6 – Observation compilation of User six's experience.

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		
T1	Grab Microphone	Close hand in a fist	Open hand slightly	YYN	Both hands closing	
T1.1	Talk			YN		
A3	Give Microphone	Drop/throw micro-phone back	Drop/Throw micro-phone back	YY		Drop

Table E.6 continued from previous page

Task ID	Task Action	or PT	Salish	User 6	Pre-Test Answer	Observation
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YY		
T4	Look at reservation pictures			YY		
T4.1	Interact with pictures	Swipe left or right	Swipe left or right	NYN	Grab them, in turns, and switch their place	Close all fingers but no motion
A4	Look at car and bear demonstration			YY		
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YYN	Grab them, in turns, and switch their place	

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	NY Y	Hold it and stretch it to make it bigger, and smash it with the palm to make it smaller	
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NY Y	Hold it with the index, middle, and thumb fingers	
T7.1	Draw bird	Drawing motion	Drawing motion	Y Y Y	Hold it the same way and move the hand	

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A6	Look up-wards when narrator points to the sun			YN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	NYN	Stretch back and forth the upper limbs	Stretched upwards
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in-/out with both hands	YYN	Hold it and stretch it to make it bigger, and smash it with the palm to make it smaller	
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	YYN	Throw it frontwards	

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
T10.1	Look at the Geese in the sky			YY		
A7	Look at Luli at the beach			YY		
T11	Look at light leading to the sky			YY		
T11.1	Look at Geese flying back			YY		
A8	Look at Geese and bear in lake			YY		

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYN	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right.	Thumbs up
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A12	Look at the mountains snow coming down like a blanket			YY		
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYN	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right.	Thumbs up
A13	Look at Geese flying at different speeds			YY		

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A14	Look at goose go- ing out of forma- tion and recovering			YY		
A15	Look at the Geese forma- tion's leader ro- tating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A17	Look at a goose falling and two others staying with him until he recovers			YY		
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right.	

Table E.6 continued from previous page

Task ID	Task or Action	PT	Salish	User 6	Pre-Test Answer	Observation
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	YYY	Hand to the back to refer to the past, and hand to the front to refer to the future	
A20	Goodbye	Wave	Praying or Wave Gesture	YY		

Table E.7 – Observation compilation of User seven's experience.

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		
T1	Grab Microphone	Close hand in a fist	Open hand slightly	YYY	One hand closing	
T1.1	Talk			YY		
A3	Give Microphone	Drop/throw microphone back	Drop/Throw microphone back	YY		Drop

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YY		
T4	Look at reservation pictures			YY		
T4.1	Interact with pictures	Swipe left or right	Swipe left or right	YYN	Look at one, and then the other	
A4	Look at car and bear demonstration			YY		
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YYN	Look at one, and then the other	
A5	Look at coyote			YY		

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	NY Y	Position the object further away to make it smaller, and vice versa	
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NY Y	Hold it with the index, middle, and thumb fingers (with cap removal)	
T7.1	Draw bird	Drawing motion	Drawing motion	YY Y	Hold it the same way and move the hand (but with the palm if bigger object)	

Table E.7 continued from previous page

Task ID	Task Action	or PT	Salish	User 7	Pre-Test Answer	Observation
A6	Look up-wards when narrator points to the sun			YN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	NYN	Position the object further away to make it smaller, and vice versa	Moved up-wards like holding a ball
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in-/out with both hands	YYY	Position the object further away to make it smaller, and vice versa	
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	YYN	Throw it frontwards	

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
T10.1	Look at the Geese in the sky			YY		
A7	Look at Luli at the beach			YY		
T11	Look at light leading to the sky			YY		With delay
T11.1	Look at Geese flying back			YY		
A8	Look at Geese and bear in lake			YY		
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Thumbs up; Aloud with slight upwards movement of the shoulders	Thumbs up

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		
A12	Look at the mountains snow coming down like a blanket			YY		
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Thumbs up; Aloud with slight upwards movement of the shoulders	Thumbs up

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
A13	Look at Geese fly- ing at different speeds			YY		
A14	Look at goose go- ing out of forma- tion and recovering			YY		
A15	Look at the Geese forma- tion's leader ro- tating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
A17	Look at a goose falling and two others staying with him until he recovers			YY		
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Thumbs up; or Aloud with slight upwards movement of the shoulders	

Table E.7 continued from previous page

Task ID	Task or Action	PT	Salish	User 7	Pre-Test Answer	Observation
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	YYY	Hand to the back to refer to the past, and hand to the front to refer to the future	
A20	Goodbye	Wave	Praying or Wave Gesture	YY		

Table E.8 – Observation compilation of User eight’s experience.

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		The user says the name as soon as the question is asked.
T1	Grab microphone	Close hand in a fist	Open hand slightly	YYY	One hand closing	

Table E.8 continued from previous page

Task ID	Task Action	or	PT	Salish	User 8	Pre-Test Answer	Observation
T1.1	Talk				YY		Talks again
A3	Give microphone	Mi-	Drop/throw micro-phone back	Drop/Throw micro-phone back	YY		Drop
T2	Look at the V-shaped peak				YY		
T3	Look back at narrator				YY		
T4	Look at reservation pictures				YY		
T4.1	Interact with pictures		Swipe left or right	Swipe left or right	YYN	Grab them at the same time, and switch their place	
A4	Look at car and bear demonstration				YY		
T5	Look at art pictures				YY		

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	YYN	Grab them at the same time, and switch their place	
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	NYN	Both hands in and out (from the inside to the out-side, and vice versa)	Uses only one hand to grab it (palm facing upwards)

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NY Y	Hold it with the index, middle, and thumb fingers (flipping in the right direction if needed)	
T7.1	Draw bird	Drawing motion	Drawing motion	Y Y Y	Hold it the same way and move the hand	
A6	Look upwards when narrator points to the sun			Y N		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	Y Y N	Pinch (index and thumb fingers) with both hands (in and out)	Pinch with single hand

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in-/out with both hands	YNN	Both hands in and out (from the inside to the outside, and vice versa)	
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	NY Y	Throw it upwards (with the palm)	Hand is not really closed to be considered grab-bing, but the user throws it upwards
T10.1	Look at the Geese in the sky			YN		
A7	Look at Luli at the beach			YY		

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
T11	Look at light leading to the sky			YN		
T11.1	Look at Geese flying back			YY		The user looks when the geese audio starts
A8	Look at Geese and bear in lake			YY		
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	NYN	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right	The user said yes aloud and used thumbs up.

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		
A12	Look at the mountains snow coming down like a blanket			YN		

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	NYN	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right	The user only spoke this time.
A13	Look at Geese flying at different speeds			YY		
A14	Look at goose going out of formation and recovering			YY		

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
A15	Look at the Geese formation's leader rotating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		
A17	Look at a goose falling and two others staying with him until he recovers			YY		

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	NYN	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right	The user only spoke.
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	YYY	Hand to the back to refer to the past, and hand to the front to refer to the future	Used the standard ‘change’ gesture.

Table E.8 continued from previous page

Task ID	Task or Action	PT	Salish	User 8	Pre-Test Answer	Observation
A20	Goodbye	Wave	Praying or Wave Gesture	YN		

Table E.9 – Observation compilation of User nine's experience.

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		
T1	Grab Microphone	Close hand in a fist	Open hand slightly	YNN	Both hands facing with the palm upwards, and closing together	The user says the name pre-emptively
T1.1	Talk			YY		The user says the name again
A3	Give Microphone	Drop/throw microphone back	Drop/Throw microphone back	YN		

Table E.9 continued from previous page

Task ID	Task Action	or	PT	Salish	User 9	Pre-Test Answer	Observation
T2	Look at the V-shaped peak				YY		
T3	Look back at narrator				YY		
T4	Look at reservation pictures				YY		
T4.1	Interact with pictures		Swipe left or right	Swipe left or right	YNN	Walk to one and then the other	
A4	Look at car and bear demonstration				YY		
T5	Look at art pictures				YY		
T5.1	Interact with pictures		Swipe left or right	Swipe left or right	YNN	Walk to one and then the other	
A5	Look at coyote				YY		

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
T6	Look at Luli's book cover			YY		
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	YNN	Walk away to make it smaller, and vice versa	The user was not interacting much until this point
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NYN	Hold it with the index, middle, ring, and thumb fingers (with cap removal)	
T7.1	Draw bird	Drawing motion	Drawing motion	YYY	Hold it the same way and move the hand	
A6	Look upwards when narrator points to the sun			YN		

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	YNN	Feed him (if alive), or, walk away to make it smaller, and vice versa (for inanimate objects)	
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in-/out with both hands	YNN	Walk away to make it smaller, and vice versa	
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	YNN	Place some wings on its back, while holding the object	
T10.1	Look at the Geese in the sky			YN		
A7	Look at Luli at the beach			YY		

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
T11	Look at light lead-ing to the sky			YN		
T11.1	Look at Geese flying back			YN		Does a slight head movement after hear-ing the geese but not enough to see them.
A8	Look at Geese and bear in lake			YY		

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right	
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
A12	Look at the mountains snow coming down like a blanket			YY		
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right	
A13	Look at Geese flying at different speeds			YY		

Table E.9 continued from previous page

Task ID	Task Action	or PT	Salish	User 9	Pre-Test Answer	Observation
A14	Look at goose go- ing out of forma- tion and recovering			YY		
A15	Look at the Geese forma- tion's leader ro- tating back when it gets tired			YY		Nods head as if in agreement
A16	Look at the Geese flying and motivating the ones in front to continue			YY		

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
A17	Look at a goose falling and two others staying with him until he recovers			YY		
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Aloud, always accompanied with nodding up and down; or Aloud accompanied with nodding left or right	

Table E.9 continued from previous page

Task ID	Task or Action	PT	Salish	User 9	Pre-Test Answer	Observation
A19	Free exploration	“Change” gesture, as in rotating with the index fingers facing each other multiple times	Gesture with both hands to change season, doing a semi-circle	YNN	Today as both hands with palm facing upwards in front, Thumb finger to the back to refer to the past, and and standard ‘change’ to refer to the future	
A20	Goodbye	Wave	Praying or Wave Gesture	YN		

Table E.10 – Observation compilation of User ten’s experience.

Task ID	Task or Action	PT	Salish	User 10	Pre-Test Answer	Observation
A1	Look around			YN		
A2	Hello	Wave	Wave	YN		

Table E.10 continued from previous page

Task ID	Task Action	or PT	Salish	User	Pre-Test Answer	Observation
T1	Grab microphone	Mi-	Close hand in a fist	Open hand slightly	YY	One hand closing
T1.1	Talk			YN		The user hesitated but didn't really talk.
A3	Give microphone	Mi-	Drop/throw micro-phone back	Drop/Throw micro-phone back	YY	Drop
T2	Look at the V-shaped peak			YY		
T3	Look back at narrator			YY		
T4	Look at reservation pictures			YY		
T4.1	Interact with pictures		Swipe left or right	Swipe left or right	YYN	Look at one, and then the other
A4	Look at car and bear demonstration			YY		

Table E.10 continued from previous page

Task ID	Task or Action	PT	Salish	User	Pre-Test Answer	Observation
T5	Look at art pictures			YY		
T5.1	Interact with pictures	Swipe left or right	Swipe left or right	NYN	Look at one, then the other	Used index finger to swipe
A5	Look at coyote			YY		
T6	Look at Luli's book cover			YY		Looked at the sky just before that
T6.1	Resize Luli's book	Pinch with both hands (forming a square)	Pinch with one hand	YNN	Pinch (index and thumb fingers)	
T7	Hold brush	Grab gesture to closed fist	Grab gesture (slightly)	NYN	Hold it with the index, middle, and thumb fingers	
T7.1	Draw bird	Drawing motion	Drawing motion	YYY	Hold it the same way and move the hand	

Table E.10 continued from previous page

Task ID	Task or Action	PT	Salish	User	Pre-Test Answer	Observation
A6	Look up-wards when narrator points to the sun			YN		
T8	Make Luli grow	Pinch gesture or Zoom with both hands	Grow gesture (as if a flower)	YNN	Pinch with one hand moving hand upwards	
T9	Adjust/Resize map with migration animation	Grab and Pinch	Zoom in-/out with both hands	YNN	Pinch (index and thumb fingers)	
T10	Make Luli fly	Grab it with one hand and move it upwards	Lift with one or two hands, from below	YNN	Two hands moving as if they were wings	
T10.1	Look at the Geese in the sky			YY		The user starts looking around more often, afterwards

Table E.10 continued from previous page

Task ID	Task Action	or PT	Salish	User 10	Pre-Test Answer	Observation
A7	Look at Luli at the beach			YY		Looks also at the sun at this point
T11	Look at light leading to the sky			YY		Not fully to the end
T11.1	Look at Geese flying back			YY		Looks up again when starting hearing the Geese sound
A8	Look at Geese and bear in lake			YY		
T12	Yes/No choice for further info on insights from the environment	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right	

Table E.10 continued from previous page

Task ID	Task or Action	PT	Salish	User 10	Pre-Test Answer	Observation
A9	Look at the sun and planets models			YY		
A10	Look at the earth globe			YY		
A11	Look at the animals disappearing (hibernating)			YY		
A12	Look at the mountains snow coming down like a blanket			YY		
T13	Yes/No choice for further info on geese migrations	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right	

Table E.10 continued from previous page

Task ID	Task Action	or PT	Salish	User	Pre-Test Answer	Observation
A13	Look at Geese fly- ing at different speeds			YY		
A14	Look at goose go- ing out of forma- tion and recovering			YY		
A15	Look at the Geese forma- tion's leader ro- tating back when it gets tired			YY		
A16	Look at the Geese flying and motivating the ones in front to continue			YY		

Table E.10 continued from previous page

Task ID	Task or Action	PT	Salish	User	Pre-Test Answer	Observation
A17	Look at a goose falling and two others staying with him until he recovers			YY		
A18	Look at the Geese joining back the same flock			YY		
T14	Yes/No for free exploration	Ring with index and thumb; Nodding	Nodding, or thumbs up/down	YYY	Nodding up and down; or Nodding left and right	

Table E.11 continued from previous page

[illegible]

Table E.11 continued from previous page

Task ID	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
A14	YY	YY	YY	YY	YY	YY	YY	YY	YY	YY
A15	YY	YY	YY	YY	YY	YY	YY	YY	YY	YY
A16	YY	YY	YY	YY	YY	YY	YY	YY	YY	YY
A17	YN	YY	YY	YY	YY	YY	YY	YY	YY	YY
A18	YY	YY	YY	YY	YY	YY	YY	YY	YY	YY
T14	YYY	YYY	YNN	YYY	YYY	YYY	YYY	NYN	YYY	YYY
A19	YYY	NYN	NYY	YNN	NYY	NYY	YNN	YYY	YNN	NYY
A20	YY	YY	YY	YY	YN	YY	YY	YN	YN	YN

E.2 Data Analysis

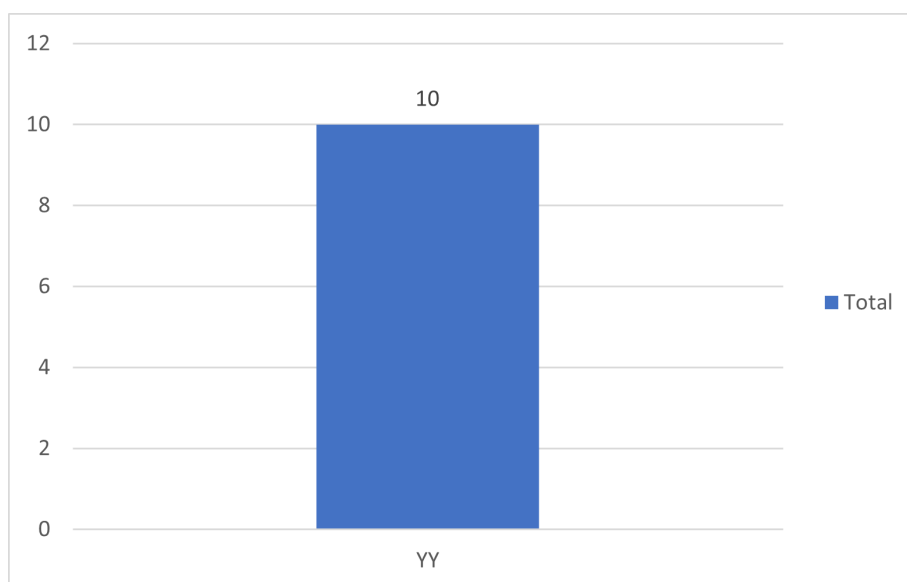


Figure E.1 – Task T4 - Look at reservation pictures.

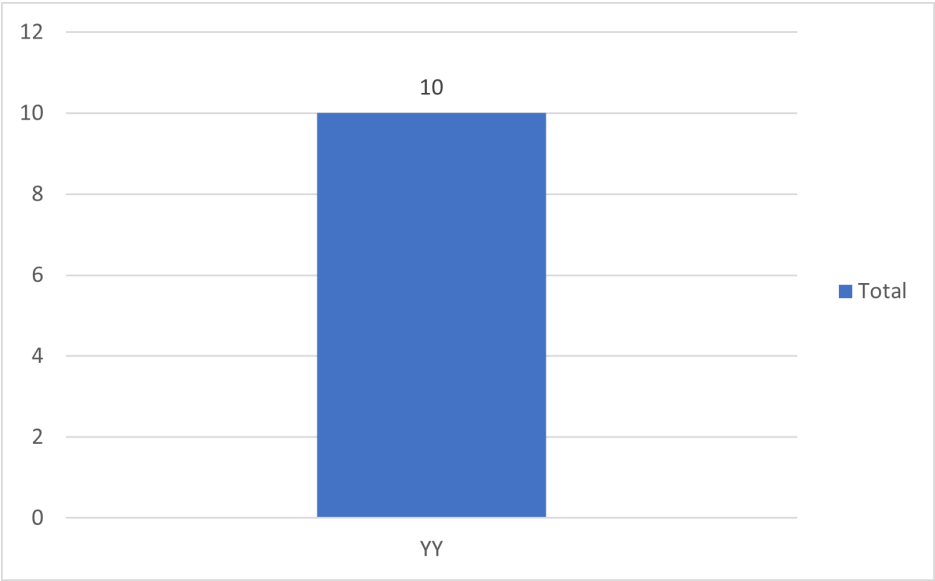


Figure E.2 – Action A4 - Look at car and bear demonstration.

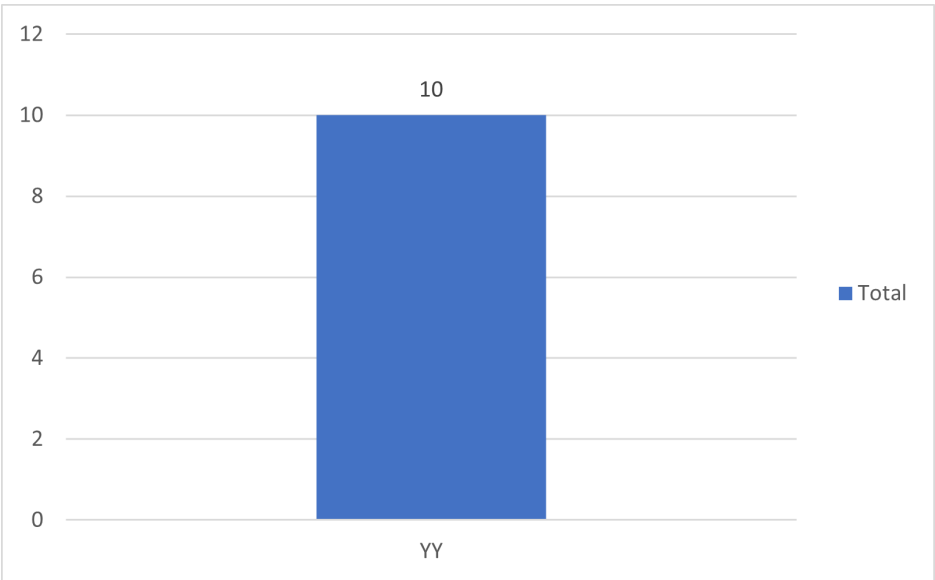


Figure E.3 – Task T5 - Look at art pictures.

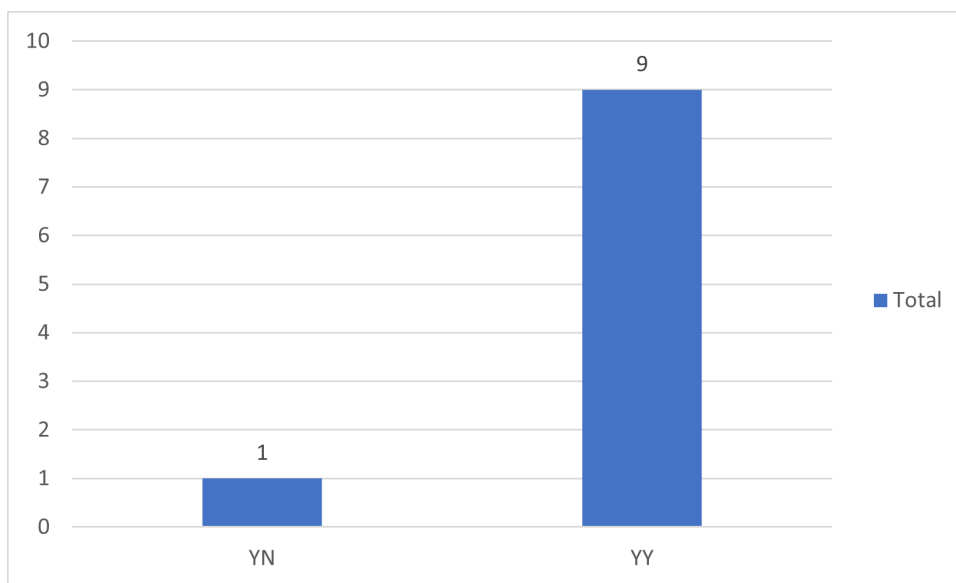


Figure E.4 – Action A5 - Look at coyote.

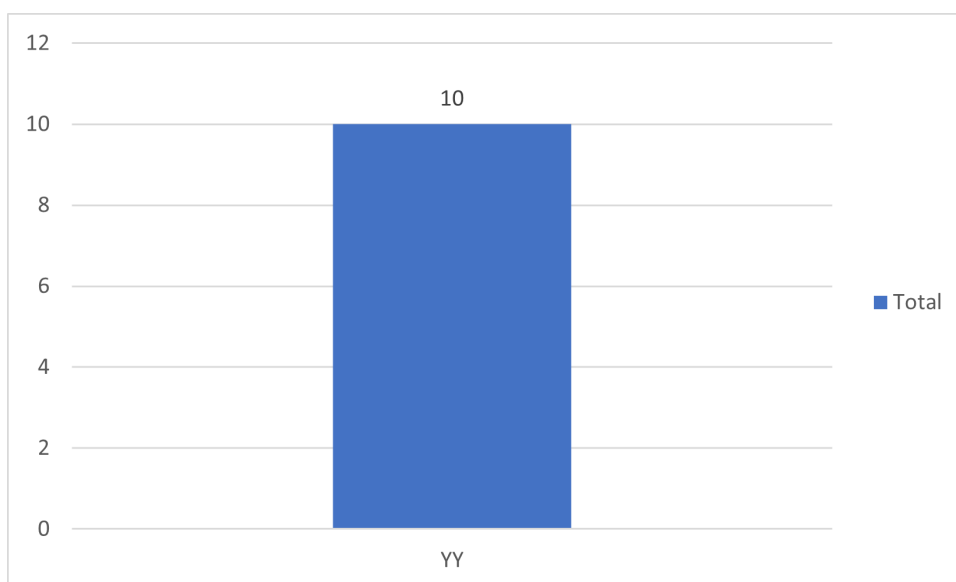


Figure E.5 – Task T6 - Look at Luli's book cover.

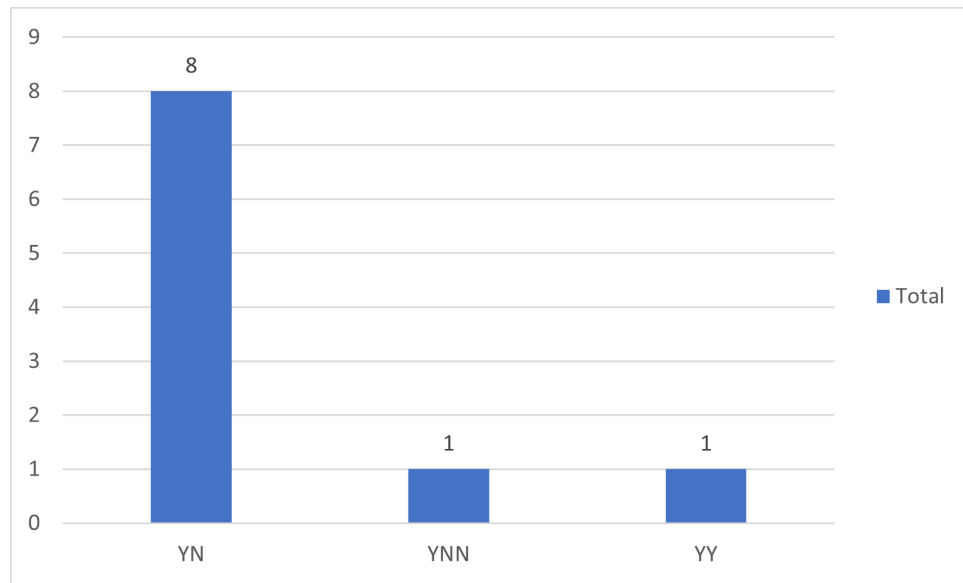


Figure E.6 – Action A6 - Look at Luli's book cover.

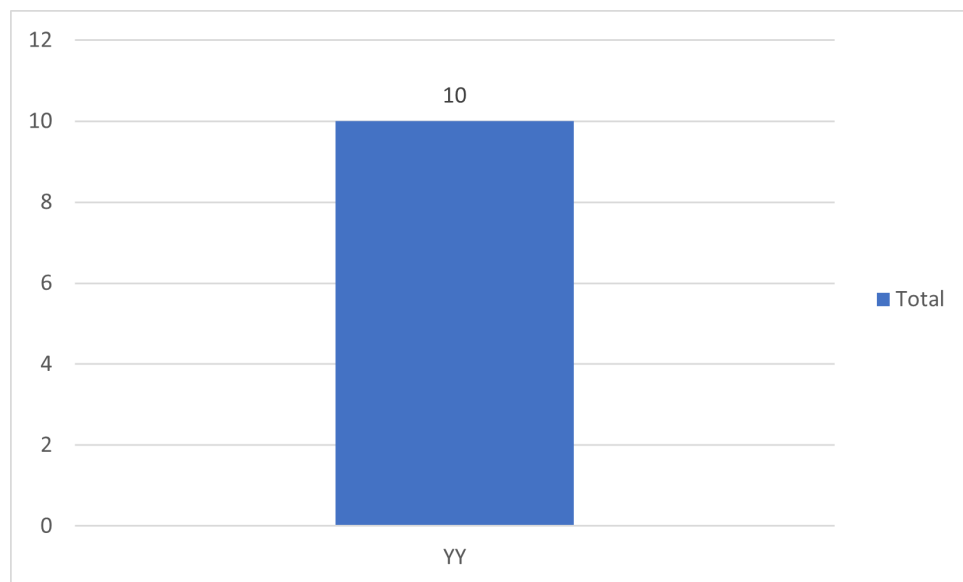


Figure E.7 – Action A7 - Look at Luli at the beach.

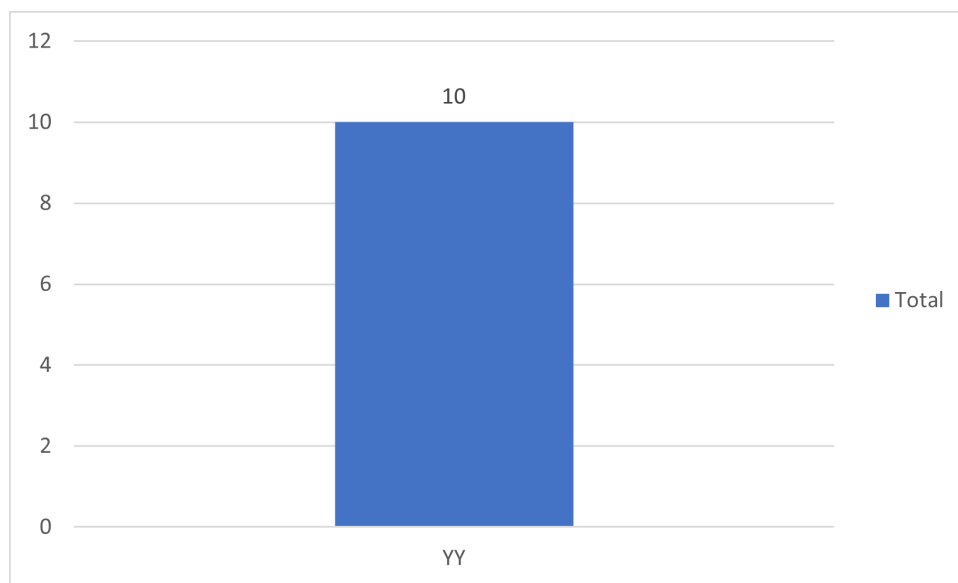


Figure E.8 – Action A8 - Look at Geese and bear in lake.

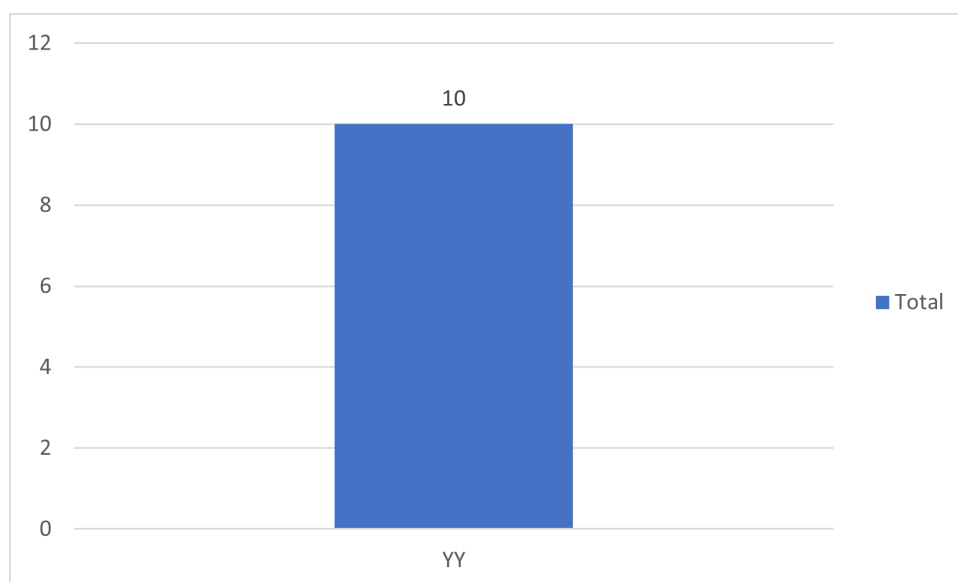


Figure E.9 – Action A9 - Look at the sun and planets models.

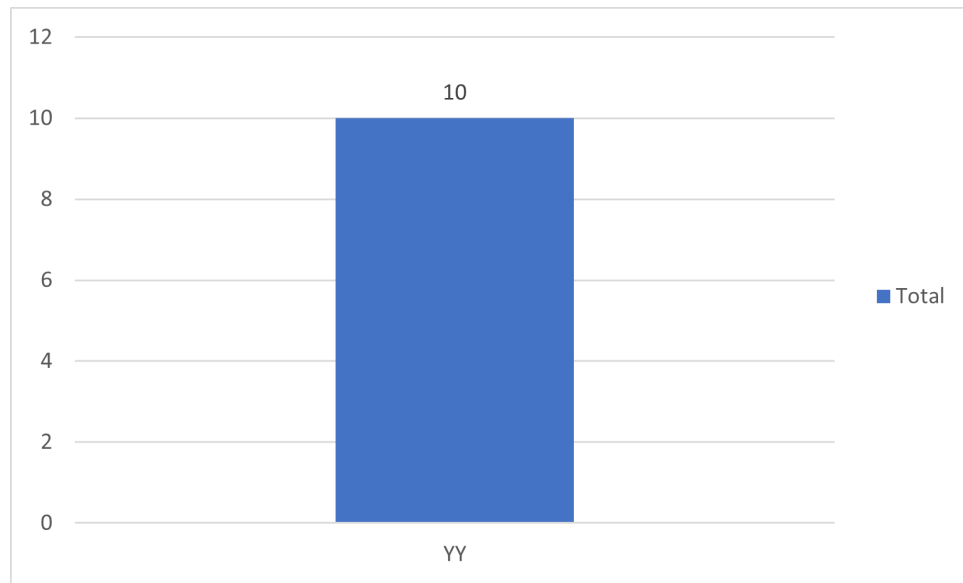


Figure E.10 – Action A10 - Look at the earth globe.

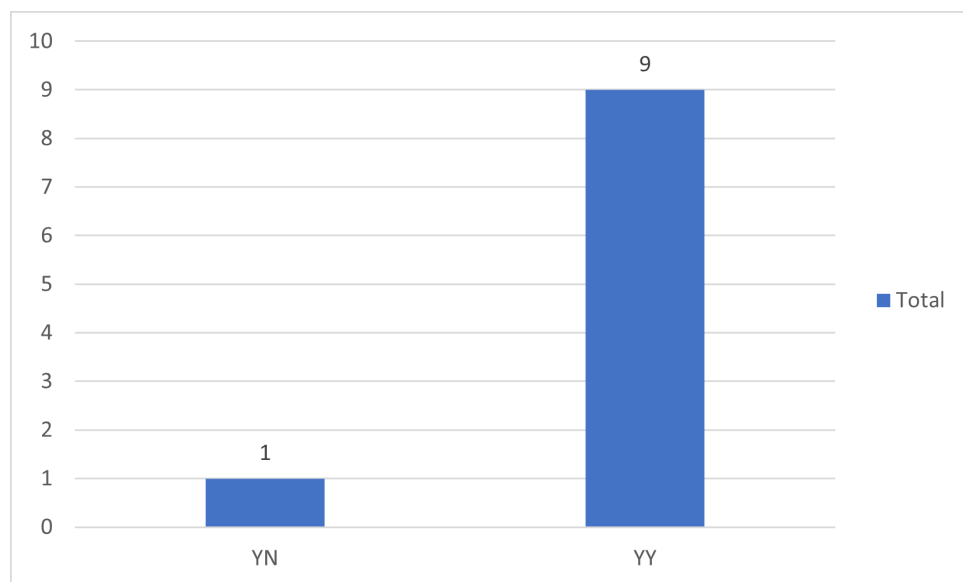


Figure E.11 – Action A11 - Look at the animals disappearing (hibernating).

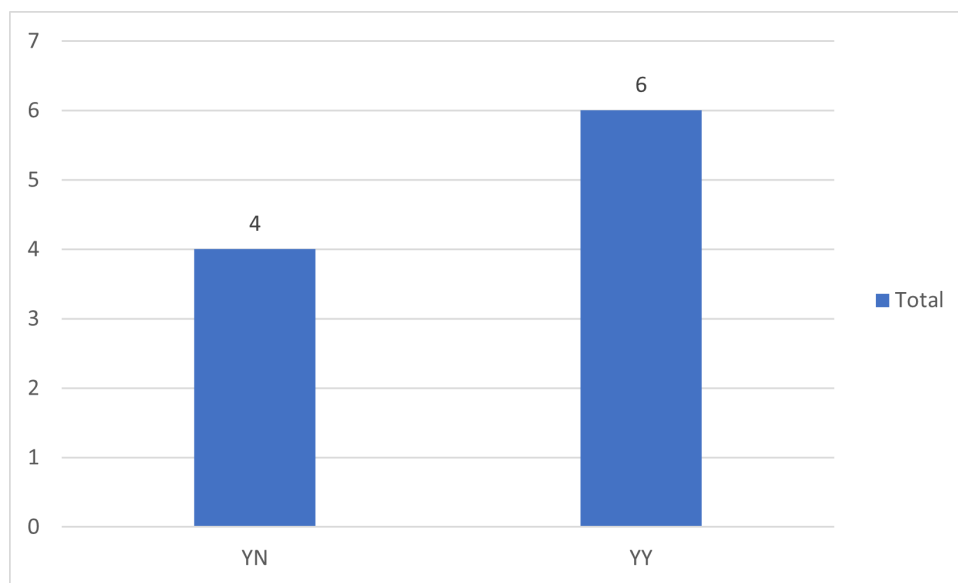


Figure E.12 – Action A12 - Look at the mountains snow coming down like a blanket.

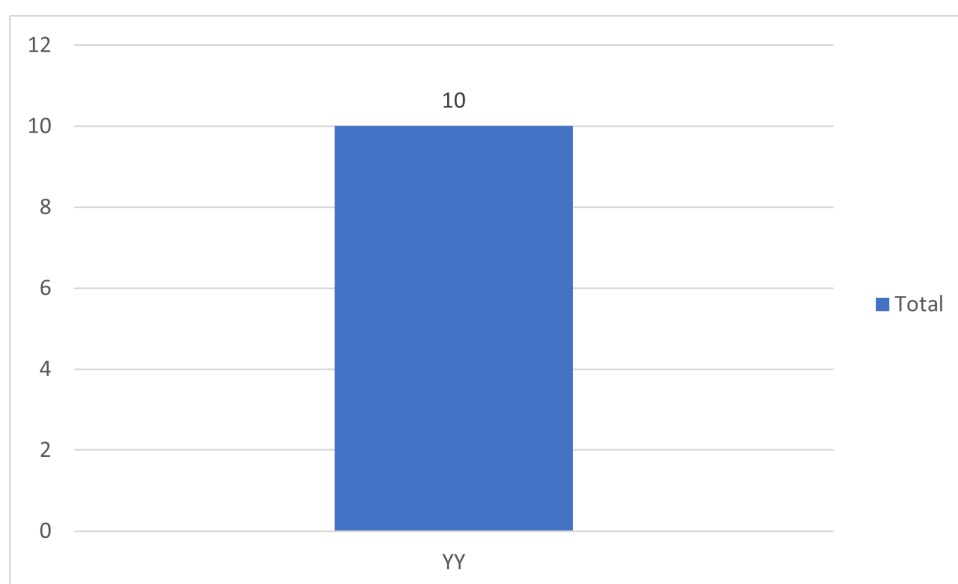


Figure E.13 – Action A13 - Look at Geese flying at different speeds.

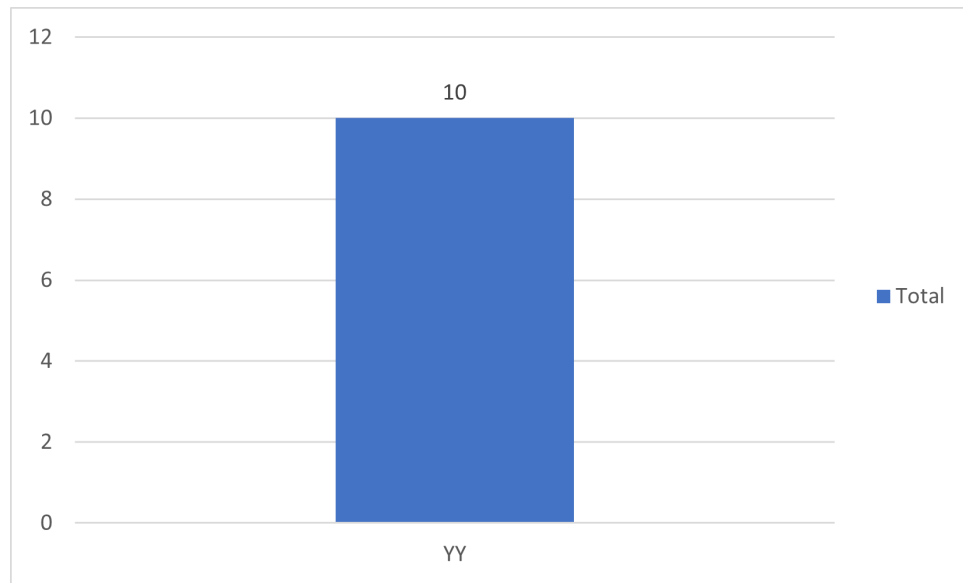


Figure E.14 – Action A14 - Look at goose going out of formation and recovering.

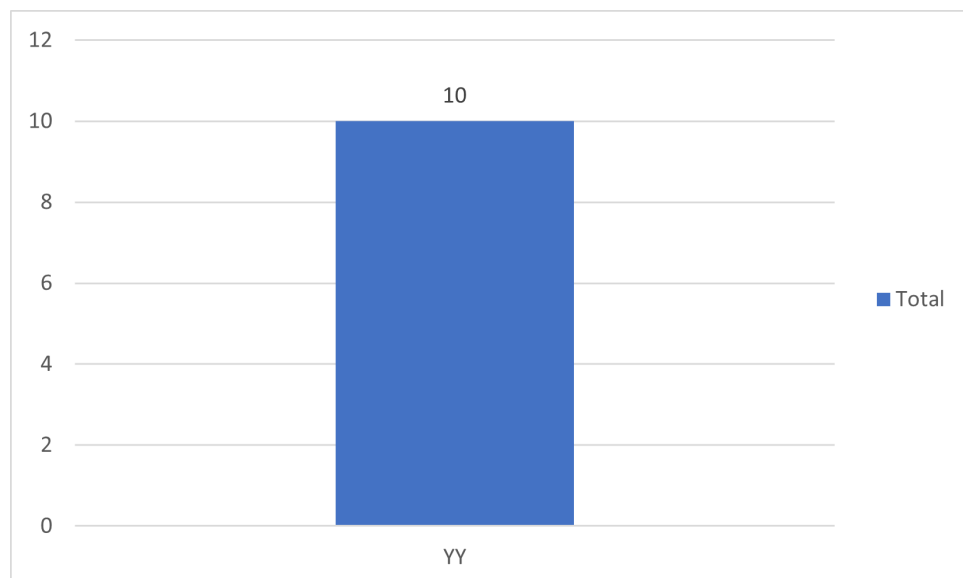


Figure E.15 – Action A15 - Look at the Geese formation's leader rotating back when it gets tired.

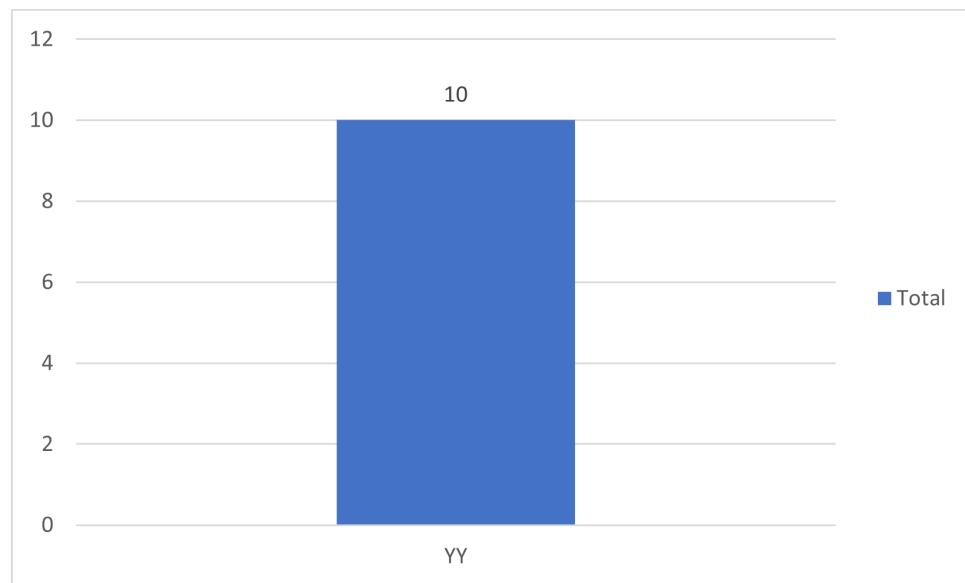


Figure E.16 – Action A16 - Look at the Geese flying and motivating the ones in front to continue.

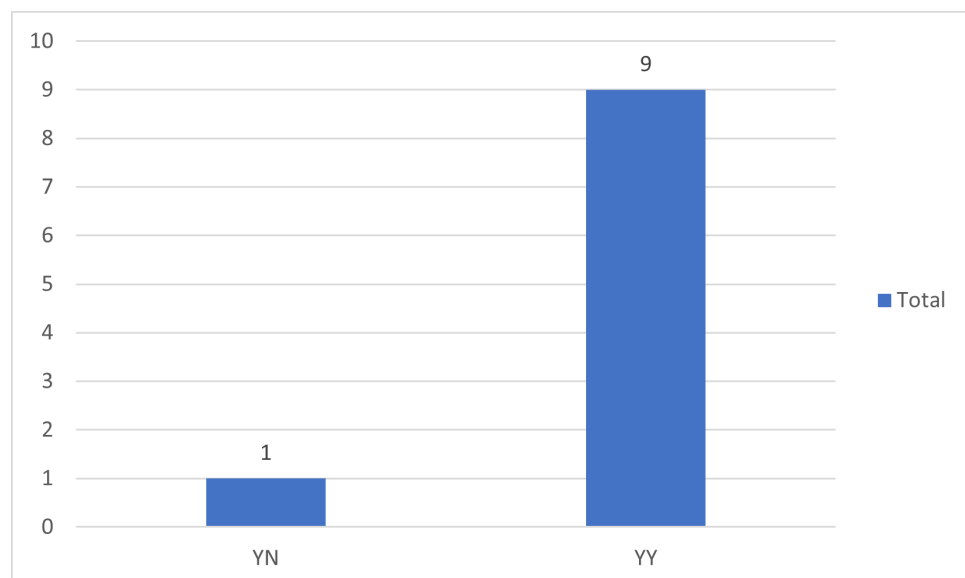


Figure E.17 – Action A17 - Look at a goose falling and two others staying with him until he recovers.

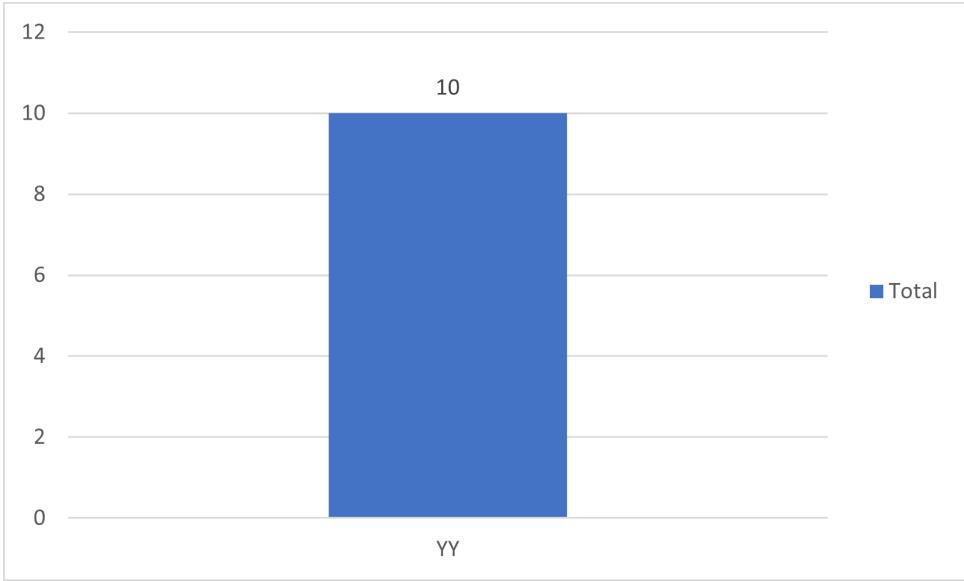


Figure E.18 – Action A18 - Look at the Geese joining back the same flock.

Acronyms

2D	Two-dimensional. 2 , 10 , 14 , 25 , 40 , 45 , 110 , 111 , 218 , 220
3D	Three-dimensional. 2 , 10 , 14 , 25 , 26 , 32 , 38 , 40 , 43 , 45 , 48 , 49 , 51 , 52 , 54 , 60 , 67 , 68 , 110 , 140 , 145 , 187 , 189 , 191 , 218
3DoF	3 Degrees of Freedom. 27 , 38 , 58
4C/ID	Four-Component Instructional Design. 15 , 16 , 69 , 111
6DoF	6 Degrees of Freedom. 31 , 43 , 58 , 59 , 169 , 218 , 226
AI	Artificial Intelligence. 123
ANT	Actor-Network Theory. 23
API	Application Programming Interface. 59 , 61 , 121 , 183
AR	Augmented Reality. 18 , 27 , 35 , 39 , 40 , 43 , 45 , 54 , 56 , 58 , 67 , 68 , 96 , 97 , 101 , 103 , 110 , 111 , 171 , 175 , 183 , 199 , 218 , 242
AT	Activity Theory. 23
BAU	Bahçeşehir University. 141
BCL	Base Class Library. 182
BDUF	Avoid doing a big design upfront. 118

CAVE	Cave Automatic Virtual Environment. 27, 38, 51, 58, 59
CHAT	Cultural-Historical Activity Theory. 23, 68
CNN	Convolutional Neural Network. 230, 231
COVID-19	Coronavirus disease 2019. 32, 141, 217, 218, 242–244
CPS	Complex Problem Solving. 15–17, 69, 111
CPU	Central Processing Unit. 128, 223
CRUD	Create, Read, Update, and Delete. 198
CSKT	Confederated Salish and Kootenai College. 3, 140, 141, 143–145, 201, 202, 221, 229, 244, 278, 279
DoF	Degrees of Freedom. 43
DRY	Don't Repeat Yourself. 118
EBSE	Evidence-based Software Engineering. 33
ECG	Electrocardiography. 49
FAU	Friedrich-Alexander University. 140
FCT	Foundation for Science and Technology (Portugal). 140, 201
FK	Forward Kinematics. 221
GIS	Geographical Information System. 26
GPS	Global Positioning System. 41, 43, 45, 49, 55, 128
GPU	Graphics Processing Unit. 223
GUI	Graphical User Interface. 195
HASS	Humanities, Arts, and Social Sciences. 12, 15, 90, 91, 113, 146
HCD	Human-centred Design. 29
HCI	Human-Computer Interaction. 1, 2, 27–31, 72, 74, 75, 131, 132, 138, 140, 141, 167, 172, 182, 226, 227
HD	High Definition. 55

HDD	Hark Disk Drive. 223
HMD	Head-mounted Display. 27, 38, 45, 50–52, 58, 59, 71, 127, 141, 167, 169, 172, 174, 183, 199, 221, 222, 227, 228, 237, 242, 247, 251, 275
IK	Inverse Kinematics. 221
IMU	Inertial Measurement Unit. 43, 49
IoC	Inversion of Control. 196, 197
IPD	Interpupillary Distance. 251
IR	Infrared. 57
IRB	Institutional Review Board. 140
IVR	Immersive Virtual Reality. 26
LED	Light-emitting Diode. 53
LiDAR	Light Detection and Ranging. 26
LoD	Law of Demeter. 118
MR	Mixed Reality. 18, 35, 39, 45, 54, 56, 58, 67, 71, 96, 110, 171, 242
NN	Neural Network. 230, 231, 240, 242
NPC	Non-player Character. 62, 68
NUIs	Natural User Interfaces. 28
ONNX	Open Neural Network Exchange. 229, 233, 236, 240, 242
ORM	Object Relational Mapping. 198, 199
OSC	Open Sound Control. 228
PBL	Problem-based Learning. 15, 69, 71
PC	Personal Computer. 222
PI	Place Illusion. 25, 32
POI	Point of Interest. 31, 54, 58, 71

Psi	Plausibility Illusion. 25, 26, 32
QAQs	Quality Assessment Questions. 42
QR	Quick Response code. 41, 50, 55, 58, 60
RAM	Random Access Memory. 223
ReLU	Rectified Linear Unit. 231, 232, 240
RFID	Radio-frequency identification. 41, 43, 49, 55
RPC	Remote Procedure Call. 124
SCs	Sensorimotor Contingencies. 25, 26, 32
SDK	Software Development Kit. 174
SKC	Salish Kootenai College. 141, 143, 201, 218
SLR	Systematic Literature Review. 24, 33, 34, 76, 92, 101, 108, 112, 140, 141
SOC	Self-organised Criticality. 13
SSD	Solid-state Drive. 223
STEM	Science, Technology, Engineering, Mathematics. 15, 48, 67, 90, 113
SVM	Support Vector Machine. 240, 241
TA	Thematic Analysis. 34, 46, 76, 96
TV	Television. 40
UDP	User Datagram Protocol. 228
UI	User Interface. 128
US	United States. 140, 144, 201, 202, 219
USA	United States of America. 140, 143
UX	User Experience. 30, 104, 218
VFTs	Virtual Field Trips. 23, 62, 66

VFX	Visual Effects. 221
VLEs	Virtual Learning environments. 18, 19
VR	Virtual Reality. 3, 27, 31, 32, 35, 38, 39, 43, 57, 58, 68, 71, 75, 96, 101, 111, 141, 145, 171, 172, 182, 188, 189, 202, 218, 221–224, 226–228, 230, 242, 251, 252, 273, 278
WPF	Windows Presentation Foundation. 183, 189, 227, 237
XR	Extended Reality. 35, 39
ZPD	Zones of Proximal Development. 16, 68, 73

Abbreviations

Abbreviation	Meaning(s)
e.g.	<i>exempli gratia</i> ; for example
et al.	<i>et alia</i> ; and others (co-authors)
ibid.	<i>ibidem</i> ; in the same place
p. (and pp.)	<i>pagina</i> (<i>paginae</i>); page(s)
i.e.	<i>id est</i> ; that is
ca.	<i>circa</i> ; approximately
cf.	<i>confer</i> ; compare; see, by way of comparison
q.v. (and qq.v.)	<i>quod vide</i> (<i>quae vide</i>); which see; on this (matter) go see
etc.	<i>et cetera</i> ; and so forth
vs.	<i>versus</i> ; against

References

- Agrawal, S., Simon, A., Bech, S., Bærentsen, K., and Forchhammer, S. (2019). Defining immersion: Literature review and implications for research on immersive audiovisual experiences. In *147th Audio Engineering Society Convention*. Audio Engineering Society. 24, 25
- Allen, R. J. (1997). A formal approach to software architecture. Technical report, DTIC Document. 117
- Alves Fernandes, L., Morgado, L., Paredes, H., Coelho, A., and Richter, J. (2019). Immersive learning experiences for understanding complex systems. *iLRN 2019 London - Workshop, Long and Short Paper, Poster, Demos, and SSRiP Proceedings from the Fifth Immersive Learning Research Network Conference*, pages 107–113. Accepted: 2019-07-01T12:08:29Z Publisher: Verlag der Technischen Universität Graz. 2, 10, 23, 24, 27, 30, 33, 201
- Alves Fernandes, L., Nunes, R. R., Matos, G., Azevedo, D., Pedrosa, D., Morgado, L., Paredes, H., Barbosa, L., Fonseca, B., Martins, P., and et al. (2015). Bringing user experience empirical data to gesture-control and somatic interaction in virtual reality videogames: an exploratory study with a multimodal interaction

- prototype. In *SciTecIn15-Conferência Ciências e Tecnologias da Interação 2015*. 30
- Alves Fernandes, L. M. (2016). Wearable devices for blind orientation and navigation. Master's thesis, University of Trás-os-Montes and Alto Douro. 21, 174
- Alves Fernandes, L. M., Cruz Matos, G., Azevedo, D., Rodrigues Nunes, R., Paredes, H., Morgado, L., Barbosa, L. F., Martins, P., Fonseca, B., Cristóvão, P., and et al. (2016). Exploring educational immersive videogames: an empirical study with a 3d multimodal interaction prototype. *Behaviour Information Technology*, 35(11):907–918. Publisher: Taylor Francis. 30, 169, 228
- Argyris, C. and Schön, D. A. (1974). *Theory in practice: increasing professional effectiveness*. Jossey-Bass Publishers, illustrate edition. Series Title: Management Series. 15, 17
- Aristotle (4th century B.C.). *Metaphysics, Book VIII, 1045a.8–10*. 12
- Axtell, R. E. (1997). *Gestures: The Do's and Taboos of Body Language Around the World*. Wiley. Google-Books-ID: 7ayesslFHf0C. 29
- Bauer, C. and Dey, A. K. (2016). Considering context in the design of intelligent systems: Current practices and suggestions for improvement. *Journal of Systems and Software*, 112:26–47. Publisher: Elsevier. 20, 23
- Beck, D., Morgado, L., and O'Shea, P. (2020). Finding the gaps about uses of immersive learning environments: A survey of surveys. *JOURNAL OF UNIVERSAL COMPUTER SCIENCE*, 26(8):1043–1073. 24, 25
- Bedrossian, J., Blumenthal, A., and Punshon-Smith, S. (2020). The batchelor spectrum of passive scalar turbulence in stochastic fluid mechanics at fixed reynolds number. *arXiv:1911.11014 [physics]*. arXiv: 1911.11014. 18
- Berlin, M., Gray, J., Thomaz, A., and Breazeal, C. (2006). Perspective taking: An organizing principle for learning in human-robot interaction. pages 7. 20

- Birch, D. P. (2008). Factors influencing academics' development of interactive multi-modal technology-mediated distance higher education courses. Publisher: Queensland University of Technology. 32
- Blaauw, G. (1972). Rechner-architektur/computer architecture. *it-Information Technology*, 14(1-6):154–159. 116
- Bock, A., Hansen, C., and Ynnerman, A. (2018). Openspace: Bringing nasa missions to the public. *IEEE Computer Graphics and Applications*, 38(5):112–118. Publisher: IEEE Computer Society. 14, 32
- Boroditsky, L., Fuhrman, O., and McCormick, K. (2011). Do english and mandarin speakers think about time differently? *Cognition*, 118(1):123–129. 22
- Bowman, N. D., Pietschmann, D., and Liebold, B. (2017). The golden (hands) rule: Exploring user experiences with gamepad and natural-user interfaces in popular video games. *Journal of Gaming Virtual Worlds*, 9(1):71–85. 28
- Braun, V. and Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597. Publisher: Routledge. 46
- Braun, V., Clarke, V., Hayfield, N., and Terry, G. (2019). *Thematic Analysis*, pages 843–860. Springer, Singapore. 46
- Bronislaw, M. (1923). *The Problem of Meaning in Primitive Languages*, pages 296–336. Harcourt, Brace World, Inc., 8 edition. 22
- Brooks, F. (1975). *The Mythical Man-Month - Essays on Software Engineering*. Addison-Wesley Publishing Company, Inc. 116
- Brooks, F. and Iverson, K. (1969). *Automatic data processing: System/360 edition*. Wiley.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31:21–32. Place: US Publisher: Harvard Education Publishing Group. 23

- Bunt, H. (1998). Issues in multimodal human-computer communication. In Bunt, H., Beun, R.-J., and Borghuis, T., editors, *Multimodal Human-Computer Communication*, Lecture Notes in Computer Science, pages 1–12. Springer. 28
- Caivano, J. L. (1998). Color and semiotics: A two-way street. *Color Research Application*, 23(6):390–401. 33
- Carvalho, P. (2019). Experiment in human-computer interaction - evaluation of a shamanic interface for interaction with cultural gestures in virtual environments. Master’s thesis, University of Porto. 2, 3, 10, 31, 168, 202, 227, 243, 277
- Checkland, P. (2012). Four conditions for serious systems thinking and action. *Systems Research and Behavioral Science*, 29(5):465–469. Publisher: John Wiley Sons, Ltd. 9
- Chi, M. T. H., Roscoe, R. D., Slotta, J. D., Roy, M., and Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science*, 36(1):1–61. 1, 9
- Chui, G. (2000). ‘unified theory’ is getting closer, hawking predicts. *San Jose Mercury News*, pages 29A. 9
- Cilliers, P. (1998). *Complexity and Postmodernism: Understanding Complex Systems*. Routledge. Google-Books-ID: OeCEAgAAQBAJ. 13
- Cilliers, P. (2013). *Understanding Complex Systems* — *SpringerLink*, pages 27–38. Springer, New York, NY. 11, 13
- Cleary, S. (2014). *Concurrency in C# Cookbook*. O’Reilly Media, Inc. 195
- Cooper, R. G. (2021). Accelerating innovation: Some lessons from the pandemic. *Journal of Product Innovation Management*, 38(2):221–232. 217
- Crouch, C. H. and Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9):970–977. Publisher: American Association of Physics Teachers. 15

- Cruz, A., Paredes, H., Fonseca, B., Martins, P., and Morgado, L. (2016). Collaboration in 3d virtual worlds: a protocol for case study research. *EAI Endorsed Transactions on Future Intelligent Educational Environments*. 145
- Cruzes, D. S. and Dybå, T. (2011). Research synthesis in software engineering: A tertiary study. *Information and Software Technology*, 53(5):440–455. 33
- Dey, A. K., Abowd, G. D., and Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, 16(2–4):97–166. Publisher: Lawrence Erlbaum Associates, Inc. 20, 23
- Dijkstra, E. W. (1968). The structure of the “THE”-multiprogramming system. *Commun. ACM*, 11(5):341–346. 116
- Dijkstra, E. W. (1982). *On the Role of Scientific Thought*, pages 60–66. Springer New York, New York, NY. 121
- Doolittle, P. E. (2014). Complex constructivism: A theoretical model of complexity and cognition. *International Journal of Teaching and Learning in Higher Education*, 26(3):485–498. Publisher: International Society for Exploring Teaching and Learning. Web site: <http://www.isetl.org/ijtlhe>. 12, 14, 15, 16
- Drew, N. and Hendley, B. (1995). Visualising complex interacting systems. In *Conference Companion on Human Factors in Computing Systems*, CHI ’95, pages 204–205. Association for Computing Machinery. 14
- Dreyfus, H., Dreyfus, S. E., and Athanasiou, T. (2000). *Mind Over Machine*. Simon and Schuster. 16, 22
- D’Mello, S. and Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition and Emotion*, 25(7):1299–1308. Publisher: Routledge. 19
- Esposito, D. and Saltarello, A. (2014). *Microsoft. NET-Architecting Applications for the Enterprise*. Microsoft Press, 2 edition. 196

- Figueiredo, A. D. d. (2005). Learning contexts: A blueprint for research. *Interactive educational multimedia: IEM*, (11):127–139. 14
- Figueiredo, A. D. d. and Afonso, A. P. (2006). *Context and Learning: A Philosophical Framework*, pages 1–22. nformation Science Publishing. xxvi, 14, 19, 20, 23, 133
- Fischer, A., Greiff, S., and Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving*, 4(1):19–42. 12, 13, 15, 16
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3):288–299. Place: US Publisher: American Psychological Association. 15
- Fowler, M. (2002). *Patterns of enterprise application architecture*. Addison-Wesley Longman Publishing Co., Inc. 198
- Freksa, C., Klippel, A., and Winter, S. (2007). A cognitive perspective on spatial context. pages 16. 21
- Garau, M., Slater, M., Pertaub, D.-P., and Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 14(1):104–116. 27
- Garcés, L., Ampatzoglou, A., Avgeriou, P., and Nakagawa, E. Y. (2017). Quality attributes and quality models for ambient assisted living software systems: A systematic mapping. *Information and Software Technology*, 82:121–138. 41
- Garlan, D. and Perry, D. E. (1995). Introduction to the special issue on software architecture. *IEEE Trans. Software Eng.*, 21(4):269–274. 117
- Garlan, D. and Shaw, M. (1993). An introduction to software architecture. *Advances in software engineering and knowledge engineering*, 1(3.4). 115, 116, 117, 119, 121
- Godden, D. R. and Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66(3):325–331. 19

- Goldstein, E. B. (2010). *Sensation and perception*. Cengage Learning. 17
- Greeno, J. G. and Moore, J. L. (1993). Situativity and symbols: Response to vera and simon. *Cognitive Science*, 17(1):49–59. Publisher: John Wiley Sons, Ltd (10.1111). 19
- Guba, E. G., Lincoln, Y. S., et al. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194):105. 133
- Han, I. (2020). Immersive virtual field trips in education: A mixed-methods study on elementary students' presence and perceived learning. *British Journal of Educational Technology*, 51(2):420–435. 14, 23
- Heald, J. B., Ingram, J. N., Flanagan, J. R., and Wolpert, D. M. (2018). Multiple motor memories are learned to control different points on a tool. *Nature Human Behaviour*, 2(4):300–311. Number: 4 Publisher: Nature Publishing Group. 17
- Hendley, R. J. and Drew, N. S. (1995). Visualisation of complex systems. pages 10. 14
- Henricksen, K. (2003). *A framework for context-aware pervasive computing applications - UQ eSpace*. PhD thesis, Queensland. 20
- Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, 19(2):4. xxx, 135, 136
- Holbert, N. R. and Wilensky, U. (2014). Constructible authentic representations: Designing video games that enable players to utilize knowledge developed in-game to reason about science. *Technology, Knowledge and Learning*, 19(1–2):53–79. 2, 10
- Holland, J. H. (2000). *Emergence: From Chaos to Order*. Oxford University Press. Google-Books-ID: VjKtpujRGuAC. 24
- Houliston, T., Fountain, J., Lin, Y., Mendes, A., Metcalfe, M., Walker, J., and Chalup, S. K. (2016). Nuclear: A loosely coupled software architecture for humanoid robot systems. *Frontiers in Robotics and AI*, 3(20). 124

- Hylving, L. (2017). Sociomaterial quasi-objects: From interface to experience. *AIS Transactions on Human-Computer Interaction*, 9(3):202–219. 23
- IAWGroup et al. (2000). Ieee recommended practice for architectural description of software-intensive systems. Technical report, Technical Report IEEE Std 1471-2000, IEEE Computer Society. 117
- Ifenthaler, D. and Eseryel, D. (2013). *Facilitating Complex Learning by Mobile Augmented Reality Learning Environments*, pages 415–438. Springer, Berlin, Heidelberg. 2, 10, 13, 14, 15, 16, 18, 19, 23
- ISO, IEC, and IEEE (2011). Ieee: 42010: 2011 systems and software engineering, architecture description. *International Standard*, pages 1,46. 117
- Jensen, L. and Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4):1515–1529. 27
- Jewitt, C. E. (2011). *The Routledge handbook of multimodal analysis*. Routledge/-Taylor & Francis Group. 147
- Johnson, N. (2009). *Simply Complexity: A Clear Guide to Complexity Theory*. Simon and Schuster. Google-Books-ID: SBq9DwAAQBAJ. 11
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux. Google-Books-ID: ZuKTvERuPG8C. 17
- Kauffman, S. A. (2008). *Reinventing the Sacred: A New View of Science, Reason, and Religion*. Basic Books. 12
- Kelley, T. (2001). Prototyping is the shorthand of innovation. *Design Management Journal (Former Series)*, 12(3):35–42. 217
- Kelly, L. B. and Kachorsky, D. (2021). Text complexity and picturebooks: Learning from multimodal analysis and children’s discussion. *Reading & Writing Quarterly*, pages 1–18. 147

- Kitchenham, B. and Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 34, 41
- Kwapien, J. and Drozd, S. (2012). Physical approach to complex systems. *Physics Reports*, 515(3–4):115–226. Publisher: North-Holland. 12, 13
- Körding, K. P. and Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10(7):319–326. 17
- Laha, B., Sensharma, K., Schiffbauer, J. D., and Bowman, D. A. (2012). Effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):597–606. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 14
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press. Series Title: Learning in Doing: Social, Cognitive and Computational Perspectives. 19
- Liebold, B., Bowman, N. D., and Pietschmann, D. (2020). Natural in the eyes of the (be)holder: A survey on novelty and learning effects in the enjoyment of naturally mapped video game controllers. *Psychology of Popular Media*, 9(2):255–265. Place: US Publisher: Educational Publishing Foundation. 28
- Lombard, M. and Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(JCMC321). 26
- Lugmayr, A., Sutinen, E., Suhonen, J., Sedano, C. I., Hlavacs, H., and Montero, C. S. (2017). Serious storytelling - a first definition and review. *Multimedia Tools and Applications*, 76(14):15707–15733. 20
- Mahdavi-Hezavehi, S., Durelli, V. H. S., Weyns, D., and Avgeriou, P. (2017). A systematic literature review on methods that handle multiple quality attributes in architecture-based self-adaptive systems. *Information and Software Technology*, 90:1–26. 34, 42, 77
- Malinowski, B. (1967). The problem of meaning in primitive languages. *Meaning in Meaning*. 20

- Malizia, A. and Bellucci, A. (2012). The artificiality of natural user interfaces. *Communications of the ACM*, 55(3):36. Publisher: ACM. 28
- McMahan, A. (2003). *Immersion, Engagement, and Presence: A Method for Analyzing 3-D Video Games*, pages 67–86. Routledge, 1st edition. 26
- Medvidovic, N. and Taylor, R. N. (2010). Software architecture: foundations, theory, and practice. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2*, pages 471–472. ACM. 121
- Meier, J., Homer, A., Hill, D., Taylor, J., Wall, L., Bansode, P., Bogawat, A., and Boucher Jr, R. (2008). Mobile application architecture guide. *Microsoft patterns & practices*. 118, 128
- Moreno, R. and Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3):309–326. Publisher: Springer. 32
- Morgado, L. (2014). Cultural awareness and personal customization of gestural commands using a shamanic interface. *Procedia Computer Science*, 27:449–459. Publisher: Elsevier. 30, 168
- Morgado, L., Cardoso, B., Carvalho, F. d., Alves Fernandes, L., Paredes, H., Barbosa, L., Fonseca, B., Martins, P., and Nunes, R. R. (2015a). Separating gesture detection and application control concerns with a multimodal architecture. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 1548–1553. 30, 167, 169, 172
- Morgado, L., Rodrigues, R., Coelho, A., Magano, O., Calçada, T., Cunha, P. T., Echave, C., Kordas, O., Sama, S., Oliver, J., and et al. (2015b). Cities in citizens' hands. *Procedia Computer Science*, 67:430–438. Publisher: Elsevier. 2, 10, 23, 25, 27, 168
- Morgado, L., Varajão, J., Coelho, D., Rodrigues, C., Sancin, C., and Castello, V. (2010). The attributes and advantages of virtual worlds for real world training. 23

- Morin, E. (2011). *Introduction to complex thought*. Média Diffusion. Google-Books-ID: qNQHCwAAQBAJ. 12, 14
- Muijres, F. T. and Dickinson, M. H. (2014). Fly with a little flap from your friends. *Nature*, 505(7483):295–296. 158
- Naur, P. and Randell, B. (1969). Software engineering: Report of a conference sponsored by the nato science committee, garmisch, germany, 7-11 oct. 1968, brussels, scientific affairs division, nato. 115
- Nilsson, N. C., Nordahl, R., and Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, 12(2):108–134. 24, 25
- Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books. 19, 22, 29, 33
- Norman, D. A. (2010). Natural user interfaces are not natural. *interactions*, 17(3):6–10. 28
- Norman, D. A. (2014). Gestural control: The good, the bad, and the ugly. 29
- Obukhova, L. F. and Korepanova, I. A. (2009). The zone of proximal development - a spatiotemporal model. *Journal of Russian East European Psychology*, 47(6):25–47. Publisher: Routledge. 16
- Ogden, C. K. and Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*. Harcourt, Brace World, Inc., 8 edition. 22
- Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37(1). 34
- Papert, P. S. (1999). What is logo? and who needs it? pages 5. 2, 10
- Papert, S. and Seymour (1980). *Mindstorms : children, computers, and powerful ideas*. Basic Books. 18

- Parnas, D. L. (1978). Designing software for ease of extension and contraction. In *Proceedings of the 3rd international conference on Software engineering*, pages 264–277. IEEE Press. 116
- Parnas, D. L., Clements, P. C., and Weiss, D. M. (1984). The modular structure of complex systems. In *Proceedings of the 7th international conference on Software engineering*, pages 408–417. IEEE Press. 116
- Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77. xxx, 135, 136, 137
- Peirce, C. S. (1931-1935). In c. hartshorne & p. weiss. *Collected papers of Charles Sanders Peirce*, 8. 33
- Perry, D. E. and Wolf, A. L. (1992). Foundations for the study of software architecture. *ACM SIGSOFT Software Engineering Notes*, 17(4):40–52. 116, 119
- Piaget, J. (1967). Logique et connaissance scientifique. 133
- Picciano, A. (2009). Blending with purpose: The multimodal model. *Journal of the Research Center for Educational Technology*, 5(1):4–14. 32
- Pierre, P. (2010a). *Blue Jay Brings the Chinook Wind*. Salish Language Revitalization Institute, Arlee, Montana, USA. 149, 219
- Pierre, P. (2010b). *Luli's Journey*. Salish Language Revitalization Institute, Arlee, Montana, USA. 143, 144, 148, 202
- Pinto, T. S., Coelho, A., Lukosch, S., and Morgado, L. (2015). A game as a tool for empirical research on the shamanic interface concept. *SciTecIN'15-Sciences and Technologies of Interaction 2015*. 3, 202, 277
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25. 17
- Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4):515–526. Publisher: Cambridge University Press. 30

- Prince, M. J. and Felder, R. M. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2):123–138. 15, 16
- Qian, J., Ma, J., Li, X., Attal, B., Lai, H., Tompkin, J., Hughes, J. F., and Huang, J. (2019). Portal-ble: Intuitive free-hand manipulation in unbounded smartphone-based augmented reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 133–145. xxxii, 228, 229, 236, 237
- Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K. E., and Ansari, R. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction*, 9(3):171–193. Publisher: ACM. 28
- Rechtin, E. (1991). *Systems Architecting: Creating and Building Complex Systems*. Prentice Hall. 116
- Repetto, A. and Catalano, C. E. (2015). A semantic layer for knowledge-based game design in edutainment applications. In *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*, pages 186–194. 18
- Roque, L. (2006). *Context Engineering for Learning: A sociotechnical approach*, pages 40–61. IGI Global. 23
- Roque, L., Almeida, A., and Figueiredo, A. D. (2004). Context engineering: an is development research agenda. *ECIS 2004 Proceedings*, pages 111. 19, 20, 23
- Roth, W.-M. (2003). From epistemic (ergotic) actions to scientific discourse: Do gestures obtain a bridging function. *Pragmatics and Cognition*, 11:139–168. 29
- Roth, W.-M. and Lawless, D. V. (2002). Signs, deixis, and the emergence of scientific explanation. *Semiotica*, 2002(138). 28
- Roth, W.-M. and Lee, Y.-J. (2007). “vygotsky’s neglected legacy”: Cultural-historical activity theory. *Review of Educational Research*, 77(2):186–232. Publisher: American Educational Research Association. 23

- Routhier, P. H. (2018). The immersive experience classification system: A new strategic decision-making tool for content creators. *SMPTE Motion Imaging Journal*, 127(10):46–54. [24](#)
- Sankey, M. D., Birch, D., and Gardiner, M. W. (2011). The impact of multiple representations of content using multimedia on learning outcomes across learning styles and modal preferences. *International Journal of Education and Development using Information and Communication Technology*, 7(3):18. Publisher: University of the West Indies. [32](#)
- Schmeder, A. and Freed, A. (2008). Implementation and applications of open sound control timestamps. In *Proceedings of the ICMC*, pages 655–658. [228](#)
- Schön, D. A. (1983). *The Reflective Practitioner How Professionals Think in Action*. [18](#), [20](#), [134](#)
- Seinfeld, S. and Müller, J. (2020). Impact of visuomotor feedback on the embodiment of virtual hands detached from the body. *Scientific Reports*, 10(1):22427. Number: 1 Publisher: Nature Publishing Group. [27](#)
- Serres, M. (1982). *The Parasite*. The Johns Hopkins University Press. Google-Books-ID: xS90DwAAQBAJ. [23](#)
- Shaw, M. (1990). Toward higher-level abstractions for software systems. *Data & Knowledge Engineering*, 5(2):119–128. [116](#), [117](#), [121](#)
- Shaw, M. and Clements, P. (1997). A field guide to boxology: Preliminary classification of architectural styles for software systems. In *Computer Software and Applications Conference, 1997. COMPSAC'97. Proceedings., The Twenty-First Annual International*, pages 6–13. IEEE. [117](#), [119](#)
- Shaw, M. and Garlan, D. (1996). *Software Architecture: Perspectives on an Emerging Discipline*. An Alan R. Apt book. Prentice Hall. [117](#)
- Sicat, R., Li, J., Choi, J., Cordeil, M., Jeong, W.-K., Bach, B., and Pfister, H. (2019). Dxr: A toolkit for building immersive data visualizations. *IEEE Transactions on*

- Visualization and Computer Graphics*, 25(1):715–725. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 14
- Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning (IJITDL)*, 2(1). 15
- Silva, H. A. P. G. d. (2008). Uma arquitetura de software dinâmica para a criação de ambientes de interacção social regulada na web. 123
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557. Publisher: Royal Society. 24, 25, 27
- Slater, M., Khanna, P., Mortensen, J., and Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications*, 29(3):76–84. Conference Name: IEEE Computer Graphics and Applications. 26
- Snoswell, A. J. and Snoswell, C. L. (2019). Immersive virtual reality in health care: Systematic review of technology and disease states. *JMIR Biomedical Engineering*, 4(1):e15025. Company: JMIR Biomedical Engineering Distributor: JMIR Biomedical Engineering Institution: JMIR Biomedical Engineering Label: JMIR Biomedical Engineering Publisher: JMIR Publications Inc., Toronto, Canada. 27
- Sousa, A. (2017). *Adaptação dinâmica e sensível ao contexto de interfaces móveis em ambiente ubíquo*. PhD thesis, University of Trás-os-Montes and Alto Douro. 21
- Steenbeek, H. and van geert, P. (2013). The emergence of learning-teaching trajectories in education: a complex dynamic systems approach. *Nonlinear dynamics, psychology, and life sciences*, 17(2):233–267. 13
- Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2–3):291–330. Publisher: John Wiley Sons, Ltd. 15, 17, 18, 19, 31

- Sterman, J. D. (2006). Learning from evidence in a complex world. *American journal of public health*, 96(3):505–14. Publisher: American Public Health Association. 18, 19, 20
- Sturmberg, J. P., Martin, C. M., and Katerndahl, D. A. (2014). Systems and complexity thinking in the general practice literature: an integrative, historical narrative review. *Annals of family medicine*, 12(1):66–74. Publisher: American Academy of Family Physicians. 12, 13, 22
- Suarez, D. (2009). *Daemon*. Dutton. 168
- Suarez, D. (2010). *Freedom (TM)*. Daemon Series. Penguin Publishing Group. 168
- Taylor, R. N., Medvidovic, N., and Dashofy, E. M. (2009). *Software Architecture: Foundations, Theory, and Practice*. Wiley Publishing. 119, 120, 121, 122, 123, 124, 125, 126, 127
- Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, 36:189–195. 28, 32
- van Merriënboer, J. J. G., Clark, R. E., and de Croock, M. B. M. (2002). Blueprints for complex learning: The 4c/id-model. *Educational Technology Research and Development*, 50(2):39–61. Publisher: Kluwer Academic Publishers. 2, 10, 14, 16
- Van Merriënboer, J. J. G. and Kester, L. (2014). *The four-component instructional design model: Multimedia principles in environments for complex learning*, pages 104–148. Cambridge handbooks in psychology. Cambridge University Press. 15
- Vance, A. (2015). *Elon Musk: Tesla, SpaceX, and the Quest for a Fantastic Future*. HarperCollins. 217
- von Foerster, H. (2003). *On Self-Organizing Systems and Their Environments*, pages 1–19. Springer New York. 13
- Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press. Google-Books-ID: Irq913lEZ1QC. 16
- W. Morris, C. (1938). Foundations of the theory of signs. 33

- Wieman, C. E. (2007). Aps news—the back page. the” curse of knowledge” or why intuition about teaching often fails. *American Physical Society News*, 16(10). 15
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19):7780–7785. Publisher: National Academy of Sciences Section: Social Sciences PMID: 17470790. 22
- Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240. 24
- Xu, Z. and Woodruff, E. (2017). Person-centered approach to explore learner’s emotionality in learning within a 3d narrative game. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference, LAK ’17*, pages 439–443. Association for Computing Machinery. 19
- Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., and Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews Neuroscience*, 20(6):364–375. Number: 6 Publisher: Nature Publishing Group. 19
- Yu, I., Mortensen, J., Khanna, P., Spanlang, B., and Slater, M. (2012). Visual realism enhances realistic response in an immersive virtual environment - part 2. *IEEE Computer Graphics and Applications*, 32(6):36–45. Conference Name: IEEE Computer Graphics and Applications. 26