

# D9.2 – Formative Evaluation Results

Sara Perry<sup>1</sup>, Maria Economou<sup>2</sup>, Hilary Young<sup>2</sup>, Akrivi Katifori<sup>3</sup>, Maria Roussou<sup>3</sup>

## Reviewed by:

Hara Stefanou (EXUS), Breffni O'Malley (NOHO)

## Abstract

This deliverable presents the formative evaluation results from Year 1 of the EMOTIVE project. It offers a brief overview of user data collected in relation to EMOTIVE's authoring process, including an early draft of the EMOTIVE Design Cards, as well as the EMOTIVE Storyboard Editor. It then presents findings from our evaluations of the alpha versions of three EMOTIVE experiences: the On-site Çatalhöyük Collaborative experience, ChatÇat - the Çatalhöyük Chatbot, and the On-site Hunterian experience.

**Official Submission Date:** 31 January 2018

**Actual Submission Date:** 16 February 2018

**Dissemination Level:** PU

<u>Partner</u>	<u>Estimated Effort (in PMs)</u>
<sup>1</sup> YORK	0.40
<sup>2</sup> UGLA	0.50
<sup>3</sup> ATHENA	0.65



# TABLE OF CONTENTS

1	EXECUTIVE SUMMARY.....	4
2	INTRODUCTION .....	5
3	FORMATIVE EVALUATION OF THE EMOTIVE AUTHORIZING PROCESS .....	6
3.1	EMOTIVE Design Cards .....	6
3.1.1	Overview .....	6
3.1.2	Methodology.....	6
3.1.3	Findings.....	7
3.1.4	Next Steps .....	8
3.2	EMOTIVE Storyboard Editor .....	9
3.2.1	Overview .....	9
3.2.2	Methodology.....	9
3.2.3	Findings.....	10
3.2.4	Next Steps .....	10
4	FORMATIVE EVALUATION OF EMOTIVE EXPERIENCES .....	11
4.1	Çatalhöyük On-site Collaborative Experience .....	11
4.1.1	Overview .....	11
4.1.2	Methodology.....	11
4.1.3	Findings.....	11
4.1.4	Next Steps .....	12
4.2	ChatÇat, the Çatalhöyük Chatbot Experience .....	12
4.2.1	Overview .....	12
4.2.2	Methodology.....	13
4.2.3	Findings.....	15
4.2.4	Next Steps .....	15
4.3	Hunterian On-site Experience .....	15
4.3.1	Overview .....	15
4.3.2	Methodology.....	15
4.3.3	Findings.....	16
4.3.4	Next Steps .....	19
5	CONCLUSION .....	20
6	BIBLIOGRAPHY .....	21
	APPENDIX A: CHATBOT EVALUATION QUESTIONNAIRE.....	22
	APPENDIX B: CHATBOT EVALUATION FINDINGS .....	24
	11/12/2017, U1 Male 45+ (D.N.) .....	24
	13/12/2017, U2 Female 45+ (K.S.) .....	25
	13/12/2017, U3 Female 20-30 (E.K.).....	25
	13/12/2017, U4 Male 20-30 (T.R.) .....	25
	20/12/2017, U5 Male 21 (G.F.).....	26
	21/12/2017, U6 Male 29 (M.G.) .....	26
	31/12/2017, U7 Male 50 (F.P.) .....	27
	Bot of conviction preliminary evaluation .....	28



APPENDIX C: COMPLETED OBSERVATION RECORD FOR HUNTERIAN ON-SITE EXPERIENCE ..... 29

APPENDIX D: COMPLETED EMOTIVE POSTCARD FEEDBACK FOR HUNTERIAN ON-SITE EXPERIENCE ..... 30

APPENDIX E: SEMI-STRUCTURED INTERVIEW QUESTIONS FOR HUNTERIAN ON-SITE EXPERIENCE..... 32

APPENDIX F: TRANSCRIPTION OF SEMI-STRUCTURED INTERVIEW WITH USERS OF HUNTERIAN ON-SITE EXPERIENCE ..... 35

**TABLE OF FIGURES**

Figure 1. Museum professionals using the EMOTIVE Design Cards at IMA/17 in Athens. .... 7

Figure 2. Screenshot from Hunterian On-site story development in the ATHENA Storyboard Editor showing the three strands of the narrative and a branching point highlighted in blue..... 9

Figure 3. Screenshots of distance slab as implemented in first iteration of Hunterian onsite Ebutius’s dilemma experience by EMOTIVE Storyboard Editor..... 17

Figure 4. Screenshots of distance slab with map, 360-degrees view and 3D buttons as implemented in second iteration of Hunterian On-site Ebutius’s dilemma experience through the Visual Scenario Editor. .... 18

**LIST OF ABBREVIATIONS**

- SBE: Storyboard Editor
- VSE: Visual Scenario Editor



# 1 Executive summary

This report presents the formative evaluation results from Year 1 of the EMOTIVE project. The following pages provide details of data collected from user feedback on two of EMOTIVE's in-development authoring methodologies, and three of EMOTIVE's alpha-version storytelling experiences (or 'use cases') for our cultural partners, the archaeological site of Çatalhöyük in Turkey, and the Hunterian Museum's Antonine Wall display in Glasgow, Scotland.

## **SUMMARY OF CONTENTS**

A brief introduction, offered in **Section 2**, explains the background to our formative evaluations. **Section 3** summarizes the formative evaluation of two of EMOTIVE's alpha-version authoring tools (the EMOTIVE Design Cards, and the EMOTIVE Storyboard Editor), providing an overview of each, as well as describing the method deployed to evaluate each, and the findings and next steps for each. In similar fashion, **Section 4** summarizes the formative evaluation of three of EMOTIVE's alpha-version user experiences (the On-site Çatalhöyük Collaborative experience, ChatÇat - the Çatalhöyük Chatbot, and the On-site Hunterian experience). **Section 5** concludes the report, commenting on next steps and links to the larger EMOTIVE Conceptual Framework (see D5.1). In the subsequent **Appendices**, we reproduce (A) the Chatbot's evaluation questionnaire; (B) selected findings from the most recent iteration of the Chatbot's formative evaluation; (C) a completed user observation record as well as (D) feedback postcards from the formative evaluation of the Hunterian On-site experience; and (E) semi-structured interview questions and (F) transcripts of an interview with two users of the Hunterian On-site experience.

## **METHODS OF RESEARCH AND ANALYSIS**

EMOTIVE's formative evaluations reflect the overall vision articulated in the project's Evaluation Framework (see D9.1), while simultaneously recognizing the specificities of each authoring methodology and experience use case (e.g., in terms of intended audience, level of expertise of users, nature of cultural heritage site, expected outcomes, etc.). Workshops, individual and group interviews, focus groups, questionnaires, observations, written records, and system logs are deployed, sometimes over multiple iterations, to gather data (primarily qualitative at this early stage) from approximately 150 users over the first year of the project. These data are analysed largely through thematic hand-coding, allowing us to consider recurring topics including usability, functionality, user engagement, emotional connection, and learning and intellectual stimulation.

## **KEY FINDINGS SUMMARISED & NEXT STEPS**

EMOTIVE's Design Cards show promise as a tool for professionals and students working on matters of experience design, although they are cumbersome in their current formulation, hence demanding redevelopment. EMOTIVE's Storyboard Editor is effective as a first prototype, and formative evaluation results have provided critical understanding to help us reconceptualize the experience design paradigm offered by the tool. The Çatalhöyük On-site Collaborative experience has been studied in depth as part of an MSc dissertation project (Mirashrafi, 2017). Evaluation results validate the emotive impact of its design and provide recommendations to inform its beta release. The Çatalhöyük Chatbot has also been conceived and analysed as part of an MSc dissertation project (Tzouganatou, 2017). Evaluation results have offered the building blocks for further iterations of the bot, including both a more discerning, information-delivery version and a more provocative, reflection-inducing version. Finally, evaluation of two versions of the Hunterian On-site experience demonstrate strong emotional responses by users, therein validating the alpha-release design, and providing key directives for exploring the delivery of more contextual information within all of EMOTIVE's stories.

## 2 Introduction

Year 1 of the EMOTIVE Project has been centred on casting our research interests wide in an effort to examine the many possibilities and impacts of emotive experiences with heritage sites. Such examinations have led us to more broadly explore the interdisciplinary literature on emotional forms of engagement with cultural heritage (see D5.1, EMOTIVE Conceptual Framework), and to narrow our attention on the creation of effective authoring methodologies for emotive stories, as well as specific EMOTIVE use cases for our two cultural partner sites.

Below we begin by reviewing the findings from our evaluations of the alpha-version EMOTIVE authoring methods. From there, we report on three of our alpha-version use cases (whose rationale and dimensions are described more fully in D3.7, Pilot Experience Prototypes), all of which have been subject to formative evaluation over several iterations. Grounded in the interdisciplinary scholarship on affect and emotion, these use cases represent distinct means of emotively connecting people (individuals and groups) to the past: through re-enactment, through critical conversation, and through personal decision making. Each use case responds to particular user needs and requirements for our partner sites, and each has been developed with specific visitor personas in mind (see D3.1, User Requirements & Scenarios – Alpha).

We conclude with general reflection on next steps for further development and summative evaluation of EMOTIVE's experiences.

### 3 Formative Evaluation of the EMOTIVE authoring process

The EMOTIVE project seeks to develop tools to assist cultural heritage professionals in the authoring of EMOTIVE experiences. Below we outline two tools that have been implemented and evaluated over Y1 of the project, including a brief description of the tools themselves, review of the process behind their formative evaluation (primarily qualitative), brief overview of the evaluative data gathered from users, and commentary on next steps for the methods in light of the findings. Note that a third tool, the EMOTIVE Visual Scenario Editor (VSE), is currently under development and its systematic evaluation will be reported in D9.3 or D9.4 (EMOTIVE's Summative Evaluation deliverables).

#### 3.1 EMOTIVE Design Cards

##### 3.1.1 Overview

At the time of the writing of this deliverable, the formative evaluation of EMOTIVE's design cards (described in D3.4, EMOTIVE User Workshop #1) has been carried out at two different settings/events: at the EMOTIVE 1st workshop in February 2017 in Glasgow and at the British Council's International Museum Academy course on Digital Skills in October 2017 in Athens. In the first event, participants were experts in experience design, whereas the latter event was within a more educational setting with less experienced participants.

##### 3.1.2 Methodology

###### ***EMOTIVE USER WORKSHOP - GLASGOW***

At EMOTIVE's User Workshop #1, the Cards' usefulness as a design methodology was explored in two ways: through direct observation of group dynamics during their application in Day 1 of the Workshop, and as feedback provided during debriefing sessions at the end of Day 1 and Day 2. We provide (in Section 3.1.3) a summary of findings, noting that details of the workshop itself and broad conclusions about the cards are reported in D3.4.

###### ***BRITISH COUNCIL'S INTERNATIONAL MUSEUM ACADEMY (IMA'17) - ATHENS***

The EMOTIVE design cards were introduced in the British Council's [International Museum Academy \(https://www.britishcouncil.gr/en/programmes/arts/cultural-skills/transforming-future-museums/international-museum-academy-2017\)](https://www.britishcouncil.gr/en/programmes/arts/cultural-skills/transforming-future-museums/international-museum-academy-2017) course on Digital Skills, which was held on October 18, 2017 in Athens, Greece (Figure 1). The event was targeted at a select group of 20 museum professionals (not all museologists) and included a hands-on prototyping activity. After introducing the EMOTIVE concept and objectives, the Hunterian Antonine Wall material and personas created by UGLA for Workshop #1 were handed out. The full deck of design cards was also made available to four groups, although we had selected a subset of cards which we judged to be more relevant to the guidelines we had given them. The four groups spent an hour thinking of a digital experience based on the material and concepts that were communicated to them earlier that afternoon.



Figure 1. Museum professionals using the EMOTIVE Design Cards at IMA/17 in Athens.

### 3.1.3 Findings

#### *EMOTIVE USER WORKSHOP #1*

Observations of the application of the Design Cards during Day 1 of the first EMOTIVE User Workshop suggested the following about their use and validity:

Participants spent most of their initial time familiarising themselves with Çatalhöyük’s personas and basic information about the site. This was to the detriment of the exploration of the design cards.

The cards provided a tangible referent to basic information (about the knowledge domain).

Cards helped structure some of the design process and/or focus on the issue at hand, but in variable and unintended ways. Those cards, which defined the goal of the experience, were used in quite a systematic, almost prescriptive way. Subsequently, cards focused on interface, experience and mechanics were played in a more random way. Some users would keep cards for later, to bring issues into the discussion. In some cases, a card was discussed and discarded because it was considered irrelevant. As participants came to talk naturally about the different aspects of their proposed experience, the cards were increasingly disregarded. This was exacerbated when users were informed about the remaining time and the necessity to complete their task.

“Players” (group members) mostly used an individual card’s title and inspiring question to guide their work. It is unclear how much the illustrations on each card were used. Support elements on the flip side of each card (concerned with design methodology and suggestions) were less used and in many cases disregarded completely.

The final debriefing on Days 1 and 2 of the Workshop provided the following feedback on the cards:

Given the timeframe, the cards were overwhelming because there were too many. Participants felt a couple of cards per person would have been sufficient to help initiate conversation, and did not want to be bound to ‘playing’ the whole deck. In any case, participants felt they needed more time, even before the meeting, to become familiar with the cards and their mechanics of use.

Groups contained skilled, experienced professionals and therefore issues were discussed naturally; as a result, participants felt stressed because they had not used the cards, yet at the same time many felt they did not need the cards. Participants reported that Day 2 of the workshop proceeded more smoothly because they only had to contend with personas and could physically see the site and its exhibits. In other words, they had fewer resources (and fewer inputs with which they were not familiar) to take into account.

The cards themselves were structured in a two-sided fashion, with suggestions for possible experiences outlined on the reverse side. Participants felt these were confusing because they compelled them to spend a lot of time thinking about how to manage the suggestions 'correctly'. A compromise needs to be found between openness to allow creativity, and structure/suggestions to guide the process and keep it within the limits of effectiveness. It is proposed that the final outcome of this resource might be a written deliverable that helps keep focus but is still flexible.

During the workshop the cards helped: (1) Structure discussion in an easy manner, especially given the fact that team members did not previously know each other, (2) start conversation and/ or resume it when there was silence, (3) focus on specific things and remember them during the design process.

Participants considered cards may be more useful for more advanced phases of design, to develop and refine experiences.

Target users: participants considered that the cards could be useful for meetings between design companies and museum clients, to help focus and structure their initial discussions. The cards could thus be better suited for a particular target audience: those who are less expert in the relevant knowledge domain.

The EMOTIVE Design Cards aim at a more open, long-term design process, whereas participants in Workshop #1 felt they needed to complete a quite complex task in a very limited amount of time. As such, the cards were primarily applied in a checklist fashion, and while they may have structured the activity for some, they were also perceived as overwhelming. Indeed, they tended to be considered as simply another resource requiring study (along with site-specific cards/information and personas) when experts would already, naturally raise the relevant issues. However, the fact that experts had the impression that they had already tackled many of the issues in the cards seems to validate, at least preliminarily, their suitability, especially for less knowledgeable or focused users. Also, comments about the need to become familiar with the team and to designate a moderator, confirm the preliminary validity of some of the cards (those we eliminated from the deck in order to speed up the design process).

#### ***BRITISH COUNCIL'S INTERNATIONAL MUSEUM ACADEMY '17 WORKSHOP***

The key points that emerged from the 4-group (20-participant) design activity at IMA'17 with regards to the cards are as follows:

Instructions on how to use the cards are too complicated and should not be on the cards themselves. We should offer different sets of instructions for different contexts of use: e.g., design sessions by experienced designers, educational activities, etc.

Authoring set-up cards (1-42) should be revised to keep only the basics or remove them completely. It makes the deck huge, it has template cards that will not be used in practice and for experienced designers it is not useful at all. We should revise them if needed for educational purposes.

Evaluation cards should be changed to reflect proposed evaluation methodologies. They are not useful as they are now.

"Format" is actually "experience type". These should be revisited to include others like chatbot, soundscape, immersive VR etc. Each experience type should have its own card. Approach: not clear what it means if you do not look at the back, so improve description using the words at the back on the design.

Personalization should be re-examined - profiling issues are not there.

"World" is also not so clear. What is the difference/interaction with plot? Could this be Setting or Props?

#### **3.1.4 Next Steps**

Based on the results reported above, we are in the process now of fully redesigning the cards, with the aim of tightening their focus and their intended audiences, creating a separate deck of evaluation-specific cards, and testing their validity amongst a range of stakeholders.

## 3.2 EMOTIVE Storyboard Editor

### 3.2.1 Overview

The main EMOTIVE authoring tool tested during the first year of the project was the Storyboard Editor (SBE), which was first formatively evaluated more systematically by the internal User Group of the EMOTIVE consortium from Month 7 onwards.

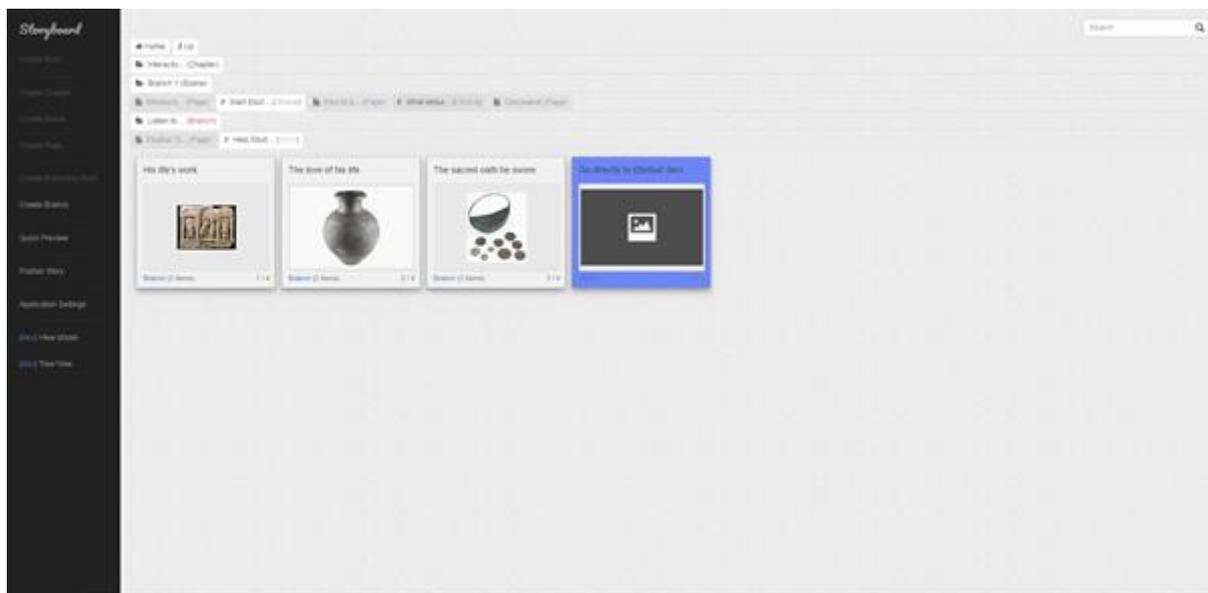


Figure 2. Screenshot from Hunterian On-site story development in the ATHENA Storyboard Editor showing the three strands of the narrative and a branching point highlighted in blue.

The SBE (initially developed by ATHENA) (Figure 2) relies on "a tiered storytelling data model in accordance to the authoring phases, distinguishing between scripting, staging, editing and producing the digital resources" as defined by CHES (Vayanou et al., 2014). The scripting phase of the experience is broken down into script units (or narrative blocks or linking segments) and script branching points, where the user must make a decision, choosing which branch of the story to follow. The script units and branching points are joined together by connecting the modes for either a branching point or linear narrative. There are layers to organise the story within the tool that are unseen by the user: "Chapters" "Scenes" and "Pages" which help to arrange the text. Multimedia assets, such as audio and images, can be uploaded to the Storyboard and attached to the relevant page.

### 3.2.2 Methodology

For the formative evaluation of the EMOTIVE SBE, members of the EMOTIVE Glasgow, York and NOHO teams acted as testers with the Glasgow and York teams representing primarily the domain experts, but also adopting the role of experience designers and storytellers. The NOHO team primarily took on the role of storytellers and experience designers, but also became domain experts to some extent during the design process. The ATHENA team also contributed to the evaluation, offering technical support for the SBE to Glasgow and York. They were thus able to identify technical glitches and difficulties with the interface in their role as developers, but also contributed to experience design and storytelling. DIGINEXT followed the process, including examining the evaluation results, in their role as developers in order to feed this knowledge into the design of the remaining components of the EMOTIVE Authoring Tool (in the first instance, for its alpha release – D4.1), crafting a system which will integrate the digital assets and support all types of authoring users.

This collaborative work between Glasgow, NOHO, ATHENA and DIGINEXT resulted in the first EMOTIVE prototype of the Hunterian Antonine Wall on-site experience which was evaluated at the Scottish European Researchers' Night, Explorathon 2017 in the Hunterian Museum, on 29 September 2017. The evaluation of the authoring of the experience using the SBE, as well as evaluation with end users in the gallery, fed into the final iteration of the alpha release of the EMOTIVE Authoring Tool (EAT).

Formative evaluation, where ATHENA led and Glasgow, York, and NOHO assisted, took place during the development of the alpha release of the EAT (D4.1). This formative phase of study continues, providing insight to the design and development of prototypes. User-centred prototyping has complemented other qualitative evaluation processes, such as think aloud protocols, screen-sharing, analysis of experience usage logs, demos and walkthroughs of tool components, together with regular Skype and email discussions between all the user partners mentioned above.

The formative evaluation of EAT included also the design of the first prototype of the Çatalhöyük collaborative experience which was tested both off site at the University of York as well as on-site at Çatalhöyük in the summer of 2017. This testing aimed to gauge both how users interacted with the prototype, as well as appropriateness of the SBE to support the authoring of the experience (Mirashrafi, 2017: 41). In terms of the York team's authoring role, in this instance, they were acting as a combination of domain experts, experience designers, and storytellers, getting support from ATHENA for both storytelling and experience design, but also from ATHENA's expertise as developers regarding the use of the SBE.

### **3.2.3 Findings**

The EMOTIVE SBE has been effective as a first prototype used in Y1 of the project. The on-going evaluation of the tool allowed the identification of a variety of issues affecting usability and user experience, both in the SBE and the produced experiences.

However, the main outcome of the evaluation activities in this period was the need to re-conceptualize the experience design paradigm offered by the tool. The current approach of structuring the experience into different chapters, which include scenes, seems to create an unnecessary overhead for the author, who does not necessarily want to structure the experience in this way. Furthermore, it is not intuitive for the authors to understand how this structure is reflected in the mobile user experience, nor do they necessarily understand the function of the branching points as they are currently visualized within the tool.

As the EMOTIVE experiences include elements of branching narratives and storytelling through characters, more emphasis should be given in the new design to offer an overview of the graph structure of the experience, combined with a textual description, structured in different parts as needed.

### **3.2.4 Next Steps**

The SBE development team will proceed with the redesign of the conceptual, structural aspects of the tool, according to the evaluation findings. The new design will be tested with internal authors as well as with UX and usability experts before implementation.

## 4 Formative Evaluation of EMOTIVE Experiences

The EMOTIVE project seeks not only to build robust authoring methodologies for the cultural heritage sector, but also to apply those (and other) methods in the development of EMOTIVE experiences for visitors to our cultural partner sites. To date, we have initiated the development of six such experiences (also referred to as use cases), three of which have progressed to the point of formative evaluation, often over multiple iterations: the On-site Çatalhöyük Collaborative experience, ChatÇat - the Çatalhöyük Chatbot, and the On-site Hunterian experience.

### 4.1 Çatalhöyük On-site Collaborative Experience

#### 4.1.1 Overview

As described in previous deliverables, our cultural partner, Çatalhöyük, is a challenging site to interpret for its diverse visiting audiences. Yet these audiences travel great distances at considerable effort to tour its Visitor's Centre, Replica Homes and active excavation areas, seeking to connect with the landscape and its Neolithic past. Our aim with this experience, therefore, has been to aid the visitor in gaining a better understanding of the mindset of the people of Çatalhöyük. With little easily identifiable physical evidence, and even fewer archaeological artefacts to interpret, visitors instead are asked to focus on more abstract themes, like societal structure, to connect themselves directly to the site's former inhabitants. In particular, the archaeological evidence suggests the Neolithic town may have functioned in an egalitarian way, with little to no structured hierarchy or accumulation of power. Different from what many people are accustomed to today, the concept may be a difficult one to fully grasp as a visitor. In EMOTIVE's On-Site Collaborative experience, participants literally go through the motions of an egalitarian society by trading and leaving behind physical artefacts with their tour group. Visitors are consequently encouraged to discuss the way in which past people lived their lives. In other words, the experience is designed to challenge Western preconceptions of ownership and community.

The alpha-version experience is the product of a MSc Dissertation by Sophia Mirashrafi, which is reported in full in Mirashrafi (2017), and detailed in brief in D3.7, Pilot Experience Prototypes. Formative evaluation data have been further analysed by Sierra McKinney, and will be used to inform beta-version development as well as our Off-site Collocated EMOTIVE experience. Here we present a brief summary of Mirashrafi's and McKinney's analyses.

#### 4.1.2 Methodology

The experience was developed through a series of stages of design, crafted directly with members of our user group and with visitors and stakeholders associated with Çatalhöyük. These included:

Development of Çatalhöyük's group personas, tested with user group members at the 1st User Workshop, February 2017.

First drafts of script prepared based on group personas, March-April 2017, then refined through bodystorming session at EMOTIVE Management Meeting, May 2017.

Authoring of experience through EMOTIVE SBE, May-July 2017.

Two pilot tests at York with 14 users, July 2017.

Formative evaluation at Çatalhöyük with 12 users: 6 friendly pairs, August 2017.

#### 4.1.3 Findings

The full methodological process, as well as evaluative data collected throughout alpha-version experience development for this use case, are reported in Mirashrafi (2017). Formative evaluation, conducted on site in summer 2017, entailed thematic analysis of (i) audio-recordings of the tours of each of six pairs of users,

(ii) standardised observational notes produced by an EMOTIVE researcher during each tour, and (iii) 30 to 60-minute audio-recorded interviews conducted with each pair immediately following their tour.

Broadly, user feedback can be divided into two categories: process (usability and functionality of the mobile experience) and experience (emotional and social engagement). With respect to the former, users found the app to be immersive, with the few identified distractions being for just short periods of time. However, users were split regarding the role of the mobile experience in connecting them to the environment. Some indicated it limited their interactions, prompting them to focus more on the device than the room. Others identified multiple instances of engagement with the physical environment through their visual and physical examination of the space.

With respect to experience, users developed feelings of attachment regarding the items used in the activity. They were commonly referred to as ‘mine’, ‘my’ or sometimes by a given name (i.e., Grandma). Users also discussed reluctance in parting with the objects, as well as their connections to what the objects represented. This showcased a strong sense of personal attachment with the items. It is clear that this personal connection was key to the experiences of the visitors. The concept of personal attachment was also discussed in the context of personalization or ownership of belongings.

Many of the users connected strongly with the role that they were assigned in the pre-visit stage of the experience. Users would refer to their role multiple times throughout the interview, often as a defining characteristic of themselves (i.e., “I am a storyteller”). This was also connected to periods of self-reflection, as a user discussed whether the role accurately or non-accurately reflected how they viewed themselves. The experience promoted social engagement both immediately between the two participants and with the larger global community through the users’ interests in engaging in the post-site experience.

The true impact of this use case can be seen most strongly in the frequency that visitors made comments indicating self-reflection and being affected or changed by the experience. Many of the users were prompted to engage in self-reflection. This included direct statements about being compelled to become self-reflective, but also through descriptions of connections to personal experiences, both real and imagined. Most positively, there were frequent statements identifying that users related to the space in one way before, and another after, participating in the experience, including expressing higher levels of empathy with the people of Çatalhöyük.

#### **4.1.4 Next Steps**

The On-site Collaborative experience was, in part, a proof of concept for exploring complex conceptual ideas (in this case, egalitarianism) in the context of a remote, hard-to-interpret, technologically very challenging archaeological site. Our preliminary findings suggest there is tremendous promise in the simple model that we have followed here, and our next steps aim to build on these successes. This will entail fully integrating 3D prints into the experience, elaborating the pre-visit experience with a more complex algorithm for assigning identities and objects (i.e., the 3D prints) to users, implementation of the post-visit experience which has not yet been tested with users, translation of all content to Turkish, pilot testing in York in Spring 2018, followed by on-site testing (pre-booked) using our EMOTIVE evaluation framework in June-July 2018.

## **4.2 ChatÇat, the Çatalhöyük Chatbot Experience**

### **4.2.1 Overview**

A series of formative design and evaluation activities have been carried out since the spring of 2017 for the evaluation of EMOTIVE’s chatbot (<https://m.me/catalhoyukbot>). These include:

Thematic analysis of existing content on Çatalhöyük’s Facebook site, May 2017

Hosting of 5 thematic live chats on Çatalhöyük's Facebook site to develop content for the chatbot, Summer and Autumn 2017

4 versions of a 'traditional' chatbot implemented (adjusting thresholds, e.g. picky/not picky), Autumn and Winter 2017

Formative evaluation of traditional bot with 10 (5 ♀, 5 ♂) non-specialist users, Autumn and Winter 2017

Ongoing development of preliminary patterns for a 'bot of conviction'

The chatbot was initiated as the MSc dissertation project of Angeliki Tzouganatou (2017), and the first two points above are reported in detail therein. Below we discuss points three and four, which reflect work conducted with Tzouganatou following the submission of her dissertation. The final point, concerning the Bot of Conviction, will be reported in future deliverables.

#### 4.2.2 Methodology

A multitude of methods have been considered and used so far for the evaluation of the chatbot.

##### **FORMATIVE EVALUATION WITH NON-SPECIALIST USERS**

For evaluation with end-users, our initial aim has been to evaluate the impact of two bot parameters in the user experience. The first parameter concerns the matching of responses (i.e., how discerning the chatbot is when replying to users' interactions):

**picky:** only matches a response when there is high confidence on similarity (i.e., similarity between existing content populating the chatbot and the content of the user's comments)

**non-picky:** matches a response when there is low confidence on similarity

We have prepared a set of "experiment bots" each of which can be configured separately with respect to these and other parameters. We call them Catallab bots. Our Experiment #1 sought to:

*Evaluate the difference of picky/non-picky matching in the case of non-casual content.*

Catallab1 (non-casual non-picky): <https://www.messenger.com/t/464133980638608>

Catallab2 (non-casual picky): <https://www.messenger.com/t/126600111389041>

Within-groups design where 10 people try both Catallab 1 and 2: 5 people start with Catallab1, 5 other people start with Catallab2, and then they continue with the other version. We ask them to evaluate it in terms of a small number of questions (the same in both steps).

The procedure for Experiment #1 went as follows:

Inform user what their task will be: "Thank you for participating in this study. You will be using Messenger on this phone to obtain information about an important UNESCO World Heritage archaeological site called Çatalhöyük, which is a Neolithic (New Stone Age) settlement dating back 9000 years. The site is located in Turkey and thus hard to get to so your task is to interact with a chatbot in order to learn more about Çatalhöyük."

Ask user to complete a consent form: "We would like to observe your interaction with the chatbot and take some photos which will be used for research purposes."

Ask user to begin by exploring on her/his own: "You may start conversing with the chatbot now in any way you wish. Please be yourself, just like you would be if we were not here."

After max **10** minutes of free exploration time, give user specific tasks.

The tasks (selected to not have related content) included:

**Let's suppose you want to learn where people of Çatalhöyük buried their dead. Ask ChatCat the following:**

**Both:**

**Where did they bury their dead?**

Well, you've picked a fascinating topic - and a bit spooky 🧟 too 😊

But I love chatting about burials cause you can learn SO many things about the past from their remains!! 🐱

Tell me Maria, what do you wanna know about them?

**Let's suppose you want to learn whether people of Çatalhöyük played. Ask ChatCat the following:**

**Did the people of Çatalhöyük play?**

**Catallab1 (non-casual non-picky) answer:**

It's really impossible to say because we interpret the archaeological record with our modern biases

For the most part, our evidence suggests people were fairly egalitarian (meaning they shared things equally) in regards to access to resources!

But we don't have any concrete evidence about their morals or ideals.

Wanna chat about egalitarianism?

**Catallab2:**

I'm just a (young) bot! It's better to focus on the things I studied.

**Did children at the Neolithic site play?**

**Catallab1 (non-casual non-picky) answer:**

The site dates back to the Neolithic and Chalcolithic periods. It was inhabited between 7100-6000 BC by up to 8,000 people who lived in this large Neolithic 'town'.

**Catallab2:**

For now I can do just a few things...Type "help" to get some options

**Did they have a theatre?**

**Catallab1:**

It seems like they didn't have full toilets, they just had open areas where trash (what archaeologists call midden areas) built up. This is where they pooped.

**Catallab2:**

Oh well! Do you know that I'm just a few weeks-old? I need a bit more time before I can really converse with you. Try "help" to see some of the things I can tell you about.

### 4.2.3 Findings

As evidenced from the chatbot's responses, there is much variability, including obvious misunderstandings, in matching responses. The questionnaire, used to gather user input on the chatbot, is reprinted in Appendix A. A selection of data from our formative evaluations is collected and presented in Appendix B.

### 4.2.4 Next Steps

We continue now to enrich the content of the "traditional" bot, to further develop and implement a Bot of Conviction (which aims to subvert the primarily information-driven delivery of content, focusing instead on questioning and provoking users), and to perform an extensive evaluation (targeting 100+ people), off-site & on-site with non-specialist users.

## 4.3 Hunterian On-site Experience

### 4.3.1 Overview

The Hunterian On-site use case is driven by an overall, high-level hypothesis that an EMOTIVE experience will increase or positively affect visitors' engagement and connection with the objects on display at the museum (and more broadly with related themes, historic periods, heritage, museums, and the past). Within our evaluation instruments, the following aspects / components of EMOTIVE experiences and their impact on the user are to be explored in more depth:

Story plot

Characters

User interface

User control of the story development

Mobility within the museum display space

Our lower-level hypothesis relates to the stage of experience development when our formative evaluation was conducted. In other words, when the first release of the story was created in EMOTIVE's Visual Scenario Editor (replacing a previous version created in the SBE), we hypothesized that the new/ added media components and added functionality would increase emotional engagement with the specific collection, the period in the past, the site and its objects.

### 4.3.2 Methodology

The Hunterian On-site experience was designed iteratively by the EMOTIVE user group, including participants in the 1st EMOTIVE User Experience Workshop in February 2017 at the University of Glasgow, and in response to user feedback and testing throughout the first year of the EMOTIVE Project. Overall, this entailed:

Development of Hunterian Museum single personas, tested with user group members at the 1st User Workshop, February 2017.

First drafts of script prepared by the University of Glasgow EMOTIVE team and NOHO, June -September 2017.

Authoring of experience by University of Glasgow team through EMOTIVE Storyboard Editor, September 2017.

Authoring of experience through EMOTIVE Visual Scenario Editor, October 2017 - January 2018 to implement extra media components.

The phase of evaluation discussed below includes formative evaluation gathered from the following evaluation events held in the Hunterian Museum between September 2017 and January 2018:



- i) user testing of the SBE version with 27 school pupils and c.40 adults at the Explorathon event held in September 2017 at the Hunterian Museum in Glasgow.
- ii) user session with c.20 University of Glasgow MSc Museum Studies, Museum Learning and Interpretation students, November 2017.
- iii) formative evaluation sessions using the Visual Scenario Editor version held in January 2018 with 10 student volunteers from Information Studies, University of Glasgow.

As well as the above-mentioned formal evaluation events, the Glasgow EMOTIVE team has also held more informal testing sessions with colleagues in Archaeology, University of Glasgow, and Heritage Environment Scotland. These discussions with peers, who are also involved in cultural heritage interpretation of the Antonine Wall site and display in the Hunterian Museum, were helpful to hone our ideas and thinking in relation to the on-site experience, but are not discussed here.

The methods used have allowed us to create a holistic, triangulated and multilevel approach to evaluation that we were able to adapt after each evaluation session, and which we will continue to refine throughout the life of the EMOTIVE project. The qualitative data that we have collected relate to: usability of the interface, the narrative of the story, users/visitors' reactions, both verbal and physical, to the experience development of the story characters, interaction with the museum objects and the exhibition space, and most importantly, emotional engagement with the objects, the story and the heritage site. Quantitative data take the form of system logs and more quantitative questions in the questionnaire that we will ultimately relate back to the qualitative data. All of the above methods are essential for further development of the EMOTIVE experiences.

### 4.3.3 Findings

Formative evaluation of this use case was conducted on site in the Hunterian Museum in September and November 2017 and January 2018. Evaluation entailed: (i) observation of users (see Appendix C for an example of a completed observation record), (ii) focus group with volunteers, (iii) visitor postcard feedback (see Appendix D for examples of completed postcards), (iv) system logs, and (v) 15 to 20-minute semi-structured interviews with users immediately after the experience (see Appendix E for semi-structured interview questions, and Appendix F for a transcription of one of the semi-structured interviews).

Two versions of the Hunterian On-site experience were tested. The first version, developed using the SBE, was tested on-site in September 2017 with two user groups (school children aged 10-11 years old and general museum visitors) and also in November 2017 with a group of about 20 MSc Museum Studies postgraduate students (Figure 3). These evaluation activities aimed to test the story created and gather feedback related to the users' experiences and engagement in relation to enjoyment of the story, interaction with the objects on display and navigation of the user in the gallery, and personalisation for the Hunterian personas.

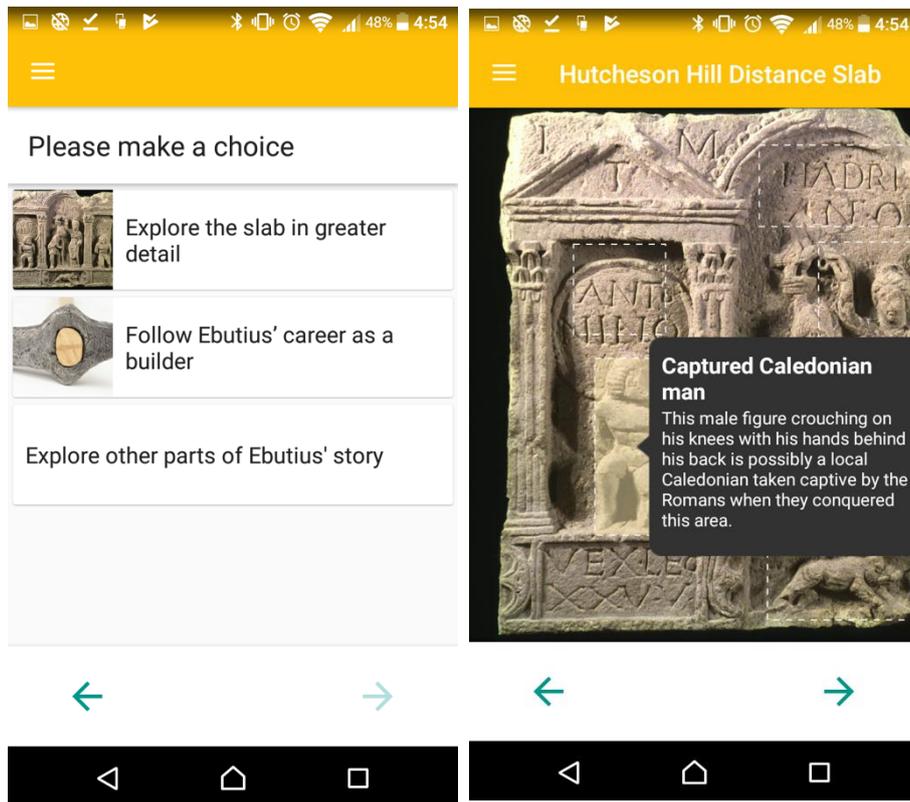


Figure 3. Screenshots of distance slab as implemented in first iteration of Hunterian onsite Ebutius's dilemma experience by EMOTIVE Storyboard Editor

The second version, created using the Visual Scenario Editor (VSE), was tested in January 2018 on-site in the Hunterian Museum with 10 Information Studies postgraduate students (Figure 4). Two Samsung S7 devices were preloaded with this version of the experience. They included new media components that were supported in the VSE, such as: 3D models of the physical objects that feature in the story and can be manipulated by the user; a 360° panorama of the Antonine Wall display space; and a floorplan of the display space with the objects located on the map. The aim of this evaluation was to test the reception of the new media components in the experience in relation to immersion and emotional engagement with the story.

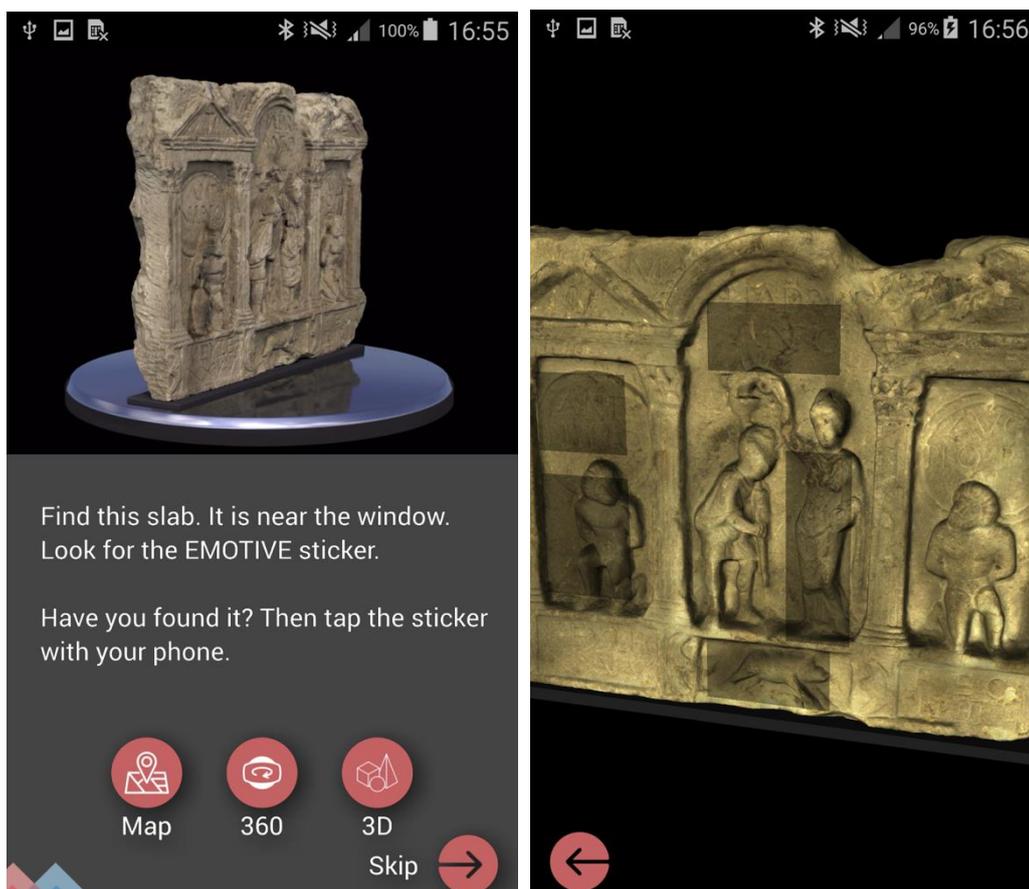


Figure 4. Screenshots of distance slab with map, 360-degrees view and 3D buttons as implemented in second iteration of Hunterian On-site Ebutius’s dilemma experience through the Visual Scenario Editor.

Both prototypes of the Hunterian onsite experience (the Storyboard Editor version and the Visual Scenario Editor version) provoked strong emotional responses from our users as evidenced in the postcard feedback, users’ casual remarks to volunteers after the experience and comments recorded in the semi-structured interviews. Based on the user evaluations of both versions these emotional responses were elicited by a combination of factors including: writing the story in the first person while looking back in time; the way the story was linked with physical objects on display; the final decision that users had to make for the main character of the story and the voice over used to narrate the story. The hotspots with more information on objects were welcomed by most users and were a highlight of both versions tested. Some users requested more of this type of functionality. One concern we had was that this type of “didactic” information would remove users from feeling immersed in the story or experience. But most users reported that this enhanced their experience by being able to explore similar objects not included in the experience. Indeed, the main finding from this phase of evaluation is the need to find a way to deliver more contextual information within the story.

More critically some users commented on the branching narrative used to structure the story, reporting that they got “stuck” or lost within linear branches. Forgetting the point of the story or the experience in relation to a specific branch of the story was also reported by users. The navigation through the “chapters” or levels of the story needs to be refined as people interpret the terms “back” and “skip” differently, thus making it confusing to navigate through the story. The lack of a specific back button to allow users to return to the page they had just visited was also highlighted. Users also commented that they would like to be able to return to specific pages, namely with the 3D models to “play” with them more at the end of the experience. Currently, once the experience is completed, the user is unable to revisit pages completed

during their experience. There were also some issues with usability of the VSE version, where users did not understand that they were required to scroll on the page to keep up with the audio, and inadvertently hit next which took them to the next section of the story. This type of usability issue will be addressed in the next development version of the experience.

#### **4.3.4 Next Steps**

Based on the formative evaluation discussed above we will focus on the following next steps in terms of developing the On-site Hunterian experience:

Address the usability of the interface and navigation including scrolling, next, skip, back

Consider implementing a progress bar or contents map to indicate how far the user is within the story

Develop personalisation of the experience in tandem with the development of the recommender system as part of the Visual Scenario Editor

Integrate the physical 3D model of the hammer into the experience

Further develop a second story that could be used in a group on-site experience

In terms of developing our evaluation instruments and methods for evaluating emotional responses in cultural heritage contexts we plan to:

Refine further our instruments for evaluating emotional engagement and empathy

Examine Matthew Reason's (n.d.) "Where in your body?" method of evaluating experiences of watching performances in relation to cultural heritage to explore further emotional engagement

Explore the potential of physiological measurement methods for recording cultural heritage engagement and emotional response

Investigate the potential of the iShoU application, developed for large-scale interactive immersive visualization environments (Kocsis and Kenderdine, 2014) for evaluating emotional engagement of EMOTIVE experiences

## 5 Conclusion

The first year of the EMOTIVE Project has seen the consortium work together fluidly and highly productively, contributing to the design of multiple pilot experiences and authoring methods. The three experiences and two authoring tools reported above have seen formative evaluation with a combined total of more than 150 users, both specialists and non-specialists. The results of our formative evaluations are promising, not only in terms of the usability and functionality of our outputs, but also in terms of their contribution to the state of the art of digital heritage practice, their capacity to engender visitor engagement with our cultural partner sites, and – most importantly – their potential for emotionally connecting visiting audiences with the distant human past. The research and associated publications of Mirashrafi (2017) and Tzouganatou (2017), coupled with our sustained Hunterian-based EMOTIVE storytelling experiments, offer in-depth understandings of our users' experiences, testifying to the promise of the conceptual model that we are following.

The broad scope of our Year 1 efforts has allowed us to explore and, latterly, hone our approach. Year 2 of EMOTIVE will see the release of beta versions of our experiences, and summative evaluations of those beta versions. Therein, our EMOTIVE Evaluation Framework will be fully implemented, and synthesis of our quantitative and qualitative data, both within and between use cases, can begin. This synthesis will provide the seedbed for larger-scale analysis of the emotive capacities and components of storytelling within cultural heritage.

## 6 Bibliography

- Kocsis, A., and Kenderdine, S. (2014). 'I Sho U: an innovative method for museum visitor evaluation', in *Digital Heritage and Culture: Strategy and Implementation*, H. Din and S. Wu (eds.), World Scientific Publishing Company, Singapore, pp. 245-260.
- Mirashrafi, S. (2017). *A collaborative experience in Çatalhöyük: Conception, Construction, and Evaluation*. MSc in Digital Heritage dissertation, University of York, submitted August 2017.
- Reason, M. (n.d.). 'Where in your body' online application, [www.whereinyourbody.com](http://www.whereinyourbody.com) [Accessed October 30, 2017].
- Tzouganatou, A. (2017). *A chatbot experience for Çatalhöyük*. MSc in Digital Heritage dissertation, University of York, submitted August 2017.
- Vayanou, M., Katifori, A., Karvounis, M., Kourtis, V., Kyriakidi, M., Roussou, M., Tsangaris, M., Ioannidis, Y., Balet, O., Prados, T., Keil, J., Engelke, T., and Pujol, L. (2014). Authoring Personalized Interactive Museum Stories. *International Conference on Interactive Digital Storytelling*, 3-6 November, 2014, Singapore.

# APPENDIX A: Chatbot Evaluation Questionnaire

The questionnaire is available on-line at:

[https://docs.google.com/forms/d/1ZkQHfycLBOeFJTurfTUz\\_sH7aaDjf376ddS39\\_s-7Pk/edit](https://docs.google.com/forms/d/1ZkQHfycLBOeFJTurfTUz_sH7aaDjf376ddS39_s-7Pk/edit)

## DEMOGRAPHICS

- Age
- Gender
- Have you interacted with a chatbot in the past? Yes/No

## ENGAGEMENT AND ABSORPTION LEVEL

1. Did you enjoy conversing with the bot?  
 Not at all | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Very much
2. While interacting with the bot you felt...  
 Uninterested | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Interested  
 Indifferent | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Curious  
 Bored | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Excited

## PERSONAL RESONANCE AND EMOTIONAL CONNECTION

3. Did you feel that the bot was...  
 Stupid | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Clever  
 Impolite | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Nice  
 Conservative | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Open-minded  
 Boring | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Interesting  
 Predictable | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Provocative
4. Did you feel that the bot could understand you?  
 Not at all | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Yes, completely

## LEARNING AND INTELLECTUAL STIMULATION

5. Conversing with the bot helped me better understand the site.  
 Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree
6. Conversing with the bot helped me learn something new about Çatalhöyük  
 Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree
7. Conversing with the bot got me thinking about things differently  
 Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree
8. While conversing with the bot, I felt challenged and provoked  
 Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree
9. While conversing with the bot, my eyes were opened to new ideas  
 Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree



### **ADDITIONAL INTERVIEW QUESTIONS**

1. Briefly describe what you remember from the interaction with the bot.
2. Which was the most interesting part (or parts) of this interaction? (if any)
3. Do you think other archaeological sites or museums should adopt chatbots? Why?
4. Elaboration on the specific evaluation objective: for example picky vs not picky
5. What would you like to say to the designers of the bot that would improve your experience?

## APPENDIX B: Chatbot Evaluation Findings

A selection of user responses to the questionnaire and interview are included in this section.

---

*11/12/2017, U1 Male 45+ (D.N.)*

14:21 - 14:28 - **free interaction C1**

Starts with casual talk. Then the bot responds something about funerary customs.

Help -> tickets : "OK, I missed that!"

### Specific Tasks

#### **Did children at the neolithic site play?**

C1: Tries to do voice recording with this question. The bot does not understand so he types the question without a question mark. He gets an irrelevant answer.

C2: OK, I missed that, try to get help for some options

#### **Do they have a theater?**

C1: does not understand

### **Additional interview questions**

1. Briefly describe what you remember from the interaction with the bot.

*That it didn't understand my questions. It was unpredictable but in a negative way because the responses were random).*

2. Which was the most interesting part (or parts) of this interaction? (if any)

*None*

3. Do you think other archaeological sites or museums should adopt chatbots? Why?

*No; if I can find the info on a website...*

4. Elaboration on the specific evaluation objective: for example picky vs not picky

*One answered randomly and the other told me to ask for help. Preferred the 2nd one.*

5. What would you like to say to the designers of the bot that would improve your experience?

*Make it work... (Make it less stupid). Also, different spelling of words and ways of writing should not return different responses from the bot. E.g., my autocorrect changed spelling of theater to theatre and neolithic to Neolithic. Or, I didn't put a question mark at the end of the questions and it seemed to change the response.*

### Comments

Check whether different conventions influence the response of the chatbot (they shouldn't). E.g. [\*Texting insincerely: The role of the period in text messaging\*](#) or [\*Punctuation in text messages may convey abruptness. Period\*](#)

"What are the paw-emojis?" (The name of the bot in Messenger should change: Maybe make it chatcat1, 2, 3, etc...?)

Add tasks / questions that work, not just the ones that don't!

---

13/12/2017, U2 Female 45+ (K.S.)

16:20 - 16:28 - free interaction with C1

#### Additional interview questions

*Loved it. Super idea. If I ask "What colour did they dye their hair" I don't expect it to answer so what I saw was great.*

*I definitely prefer C2 (I don't know responses). It's more honest, it doesn't say nonsense (δεν λει αρλουμπες), which make me lose trust in it.*

---

13/12/2017, U3 Female 20-30 (E.K.)

16:20 - 16:29 - free interaction with C1

#### Additional interview questions

1. Briefly describe what you remember from the interaction with the bot.

*It was interesting. At the beginning I thought that it would be very dumb, but finally I found it to be funny so I liked the experience. It grew on me, it made me want to ask it things. The experience was built slowly. And you learn some things too.*

2. Which was the most interesting part (or parts) of this interaction? (if any)

*The emoticons.*

3. Do you think other archaeological sites or museums should adopt chatbots? Why?

*Yes, maybe for museums or sites with history but not available to see. The details. It makes me understand things better when someone is telling me or when I discuss with someone rather than having to, for example, read information on a website. You feel as if you are conversing with someone.*

4. Elaboration on the specific evaluation objective: for example picky vs not picky

*I don't know which of the two versions I prefer. On the one hand, C2 ("I don't know") is good but it also annoys me when I get so many I don't answers all the time. But C1 makes me want to make fun of it and (τρολλάρω) after a while.*

5. What would you like to say to the designers of the bot that would improve your experience?

*I was generally covered by the answers I got from the bot. Maybe just make it answer more things.*

---

13/12/2017, U4 Male 20-30 (T.R.)

17:07-17:14 - free interaction with C2

#### Additional interview questions

1. Briefly describe what you remember from the interaction with the bot.

*It's very nice! I liked that it was humorous. I liked the parts with the toilets and the graves.*

2. Which was the most interesting part (or parts) of this interaction? (if any)

*I really liked the suggestions. And that I could select one, it would answer and then it would ask me if I wanted it to tell me something else.*

3. Do you think other archaeological sites or museums should adopt chatbots? Why?



*Yes. It is very fun to obtain information through such an interaction. It is more accessible, like an everyday conversation, like talking with a friend of yours.*

4. Elaboration on the specific evaluation objective: for example picky vs not picky

*I didn't like any of the two versions more than the other. I would probably prefer something in between. For example, not to tell you something irrelevant nor tell you "I don't know" but to offer you another suggestion when it doesn't know the answer. I didn't like that I had to press 'help' every time. And I didn't like that the continuous "I don't know" either.*

5. What would you like to say to the designers of the bot that would improve your experience?

*I didn't feel that I could ask anything I wanted. Since it started providing me with suggestion buttons, I assumed that it doesn't know much, so I didn't ask it further. I guess that I would want it to be somehow smarter.*

---

**20/12/2017, U5 Male 21 (G.F.)**

15:02-15:10 - free interaction with C2

#### **Additional interview questions**

1. Briefly describe what you remember from the interaction with the bot.

*Its responses are very "formal", as if it was presenting me with a text. Most of the times it did not understand what I was asking. When I was trying to get answers from it, it was losing its flow. It didn't give me the impression that I was conversing with someone.*

2. Which was the most interesting part (or parts) of this interaction? (if any)

*It's very original to use a chatbot for such a purpose (or for anything, for that matter). But it would be good if there was a conversation. I wanted to ask about the use of tools at Catalhoyuk but there was no way, I had to start the "communication thread" again. It was a like a set of links rather than a conversational bot. In this sense, I'd prefer to have a website where I can see all at once the available options than having a chatbot which does not understand.*

3. Do you think other archaeological sites or museums should adopt chatbots? Why?

*Yes, but not in the form that it is now.*

4. Elaboration on the specific evaluation objective: for example picky vs not picky

*I prefer C1 even though it does not answer your question. I prefer it because it doesn't force you to go back to the beginning if you want to learn something. It doesn't make sense for me to type. If it says "I don't know" (C2), it discourages you from using it. Whereas if it says something else, even if not relevant, the experience continues.*

5. What would you like to say to the designers of the bot that would improve your experience?

*If they can make it, to have a conversation.*

---

**21/12/2017, U6 Male 29 (M.G.)**

11:00-11:10 - free interaction with C1

#### **Additional interview questions**

1. Briefly describe what you remember from the interaction with the bot.



EMOTIVE

*It was interesting in the sense that it had the bubbles to guide me to the different topics because the problem many times with these chatbots is that I don't know what to ask. Also the information presented was kind of "juicy" information. It wasn't academic. The answers had to do with "poop" and how they buried people, which are inherently more interesting facts to know because they are fun facts. And the informal way was nice. The emoticons made sense. Trying to engage with me.*

2. Which was the most interesting part (or parts) of this interaction? (if any)

*It was interesting in the sense that it had the bubbles to guide me to the different topics because the problem many times with these chatbots is that I don't know what to ask. Also the information presented was kind of "juicy" information. It wasn't academic. The answers had to do with "poop" and how they buried people, which are inherently more interesting facts to know because they are fun facts. And the informal way was nice. The emoticons made sense. Trying to engage with me.*

What not?

*When there was a lot of text, the speed I was reading it was not fast enough and so the top parts of the text would not fit in the visible part of my screen (I had to scroll back up to see). This was not something that disengaged me but it was annoying. I was, like, what is happening now, where is my text?*

Was it the information that was interesting and/or that the information was conveyed in a different way?

*Both. I liked the informal way and the emoticons used. The emotions used were good because they made sense; they were not for instance a smiley just for the sake of a smiley. What I liked was that the chatbot was trying to engage with me through the questions it was asking me.*

3. Do you think other archaeological sites or museums should adopt chatbots? Why?

*I think that, in general, there should be a hierarchy of information giving; in order to engage users, use something informal that will capture their interest so that people start understanding what is going on, and then go deeper and deeper to get the more formal knowledge. In guided tours you get all of the formal information that the curators want to give you, which can be interesting but in the end you will be bored. If you start with something that is more informal and then go to the formal it's a top down approach, I would say, you start with the more general and then go to the more specific. Even if people do not care for the specifics, they will go out of the experience knowing a few fun facts which in the end is more than what they originally knew before getting into the experience.*

4. Elaboration on the specific evaluation objective: for example picky vs not picky

*I prefer the picky one (C2) because I think that the non-picky one tries to distract you, I feel that it's not transparent enough, and that makes me feel uncomfortable. If, however, the picky one responded "I don't know what you're asking but here is something about the toilets which you may find interesting" and then provide this I would prefer it.*

5. What would you like to say to the designers of the bot that would improve your experience?

*Interface-wise it's the standard Messenger so it's good. But I'm not sure whether it is clear to all that you can scroll right (horizontally) to see that there are more bubbles. In the desktop application I could see the arrow but in the mobile I could not, it's just implied. A suggestion would be to not have more than three buttons placed horizontally. The non-picky variation is annoying because of it not being transparent.*

---

31/12/2017, U7 Male 50 (F.P.)

10:50-10:57 - free interaction with C2



EMOTIVE

### **Additional interview questions**

1. Briefly describe what you remember from the interaction with the bot.

*That the excavation took place in the 60s, the burials were within the houses, they had an egalitarian system, hmm what else?*

Anything else that has to do with your interaction with the bot?

*There was a problem at the beginning in that I couldn't read because the text was going off my screen and I had to swipe up in order to read the whole text. Also, in free-text questions of mine the bot would get confused but that's normal*

2. Which was the most interesting part (or parts) of this interaction? (if any)

*I don't know. In general it was a nice way, a somewhat interactive way, to learn a few things.*

3. Do you think other archaeological sites or museums should adopt chatbots? Why?

*Yes. Because they're easy and quick and interactive.*

4. Elaboration on the specific evaluation objective: for example picky vs not picky

*In one I wrote 'theater' and it didn't understand anything and in the other I wrote 'theater' and it understood 'toilet'. I prefer the one that says about toilets because I learned something about their toilets even though I hadn't asked; better than not learning anything... I don't mind that the bot's response was irrelevant to what I'd asked, as long as the response provides me with some (any) information.*

5. What would you like to say to the designers of the bot that would improve your experience?

*To add more images.*

---

### **Bot of conviction preliminary evaluation**

**27/01/2018, CC\_U6 MALE 29 (M.G.)**

*In the first part (where the bot asks the users about where they bury their dead, etc.) there was a buildup of interest by provocation. This part was very good, I liked that the bot was asking me questions that were a bit provocative and caught my interest.*

*The second part where the bot gave me some info about how they buried their dead at Catalhoyuk was not so good. I felt that the conversation ended (rather suddenly) and I was simply presented with information. Ideally, I would've liked to continue the conversation or at least have the illusion that I was still in conversation with the bot.*

# APPENDIX C: Completed Observation Record for Hunterian On-site Experience

Shared experience: 2 Female users, both university students

Observation Sheet – EMOTIVE Antanine Wall

Observer: *Mekala Ferguson*      Date: 29/9/17      Time: 2:35

Group composition:	# Adult Females	2	#Girls	(ages)
	#Adults Males		#Boys	(ages)

Time entered exhibition: *2:35*      Time left exhibition: *2:53*

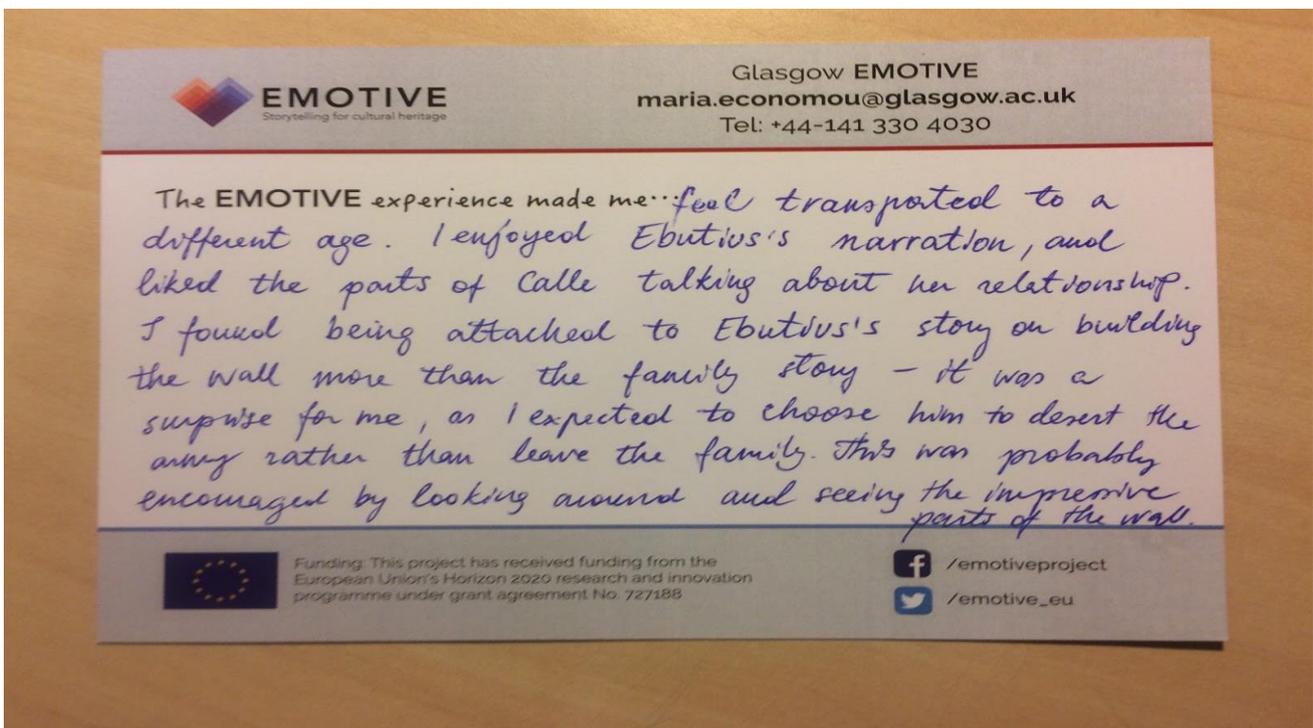
Mark on plan below with a C where group has conversation, R where they stop to read any text. Note any intriguing conversation or interesting behaviour in margins (e.g. H-off and H-on for Headphones). If you can discern facial expressions, use E for excitement and F for frustration. Later tally up the behaviours and expressions in table below.

① x RC Start

	Adult Fem	Adult Male	Girls	Boys
Convers				
Read				
Excitem				
Frustr				

## APPENDIX D: Completed EMOTIVE Postcard Feedback for Hunterian On-site Experience

See below the front of the postcard, followed by two examples of the reverse side of the postcard, completed by two different users.



The EMOTIVE experience made me...

- Understand the backstory, lives and feelings of the people living at the time.

A very creative way to learn about the exhibits!



Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 727188



/emotiveproject



/emotive\_eu



B6. Did you do the EMOTIVE experience:

a) on your own b) shared a device with someone else c) other, please specify \_\_\_\_\_

**C. ENGAGEMENT**

C1. Did you enjoy the experience?

Not at all | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Very much

C2. I wanted more control over how the story developed

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree

Please explain further: \_\_\_\_\_

C3. During the experience, which of the following did you feel? (Circle all that apply)

Indifferent Interested Uninspired Bored

Excited Captivated Engaged Disappointed

Satisfied Neutral Frustrated Other \_\_\_\_\_

C4. I felt like I was transported to another world and lost track of time

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree

**D. EMOTIONAL CONNECTION**

D1. I felt empathy for the characters in the story

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree

D2. I found the experience emotionally engaging

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree

D3. Some aspects of the experience seemed relevant to my own life

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree

D4. The experience brought the past to life for me.

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree

D5. The experience made me connect with the objects on display

Completely disagree | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | \_\_\_\_\_ | Completely agree



D6. I will be thinking about the experience for some time to come

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

**E. LEARNING AND UNDERSTANDING**

E1. The experience helped me better understand the Antonine Wall.

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

E2. The experience helped me learn something new about the Antonine Wall.

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

E3. The experience helped me engage with the objects on display

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

E4. During the experience, I felt challenged and provoked

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

E5. During the experience, my eyes were opened to new ideas

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

E5.1 If you agreed with the previous statement, please explain what these new ideas were

---

---

---

E6. The experience gave me a new understanding for people who are different than me/ for a culture other than my own.

Completely disagree | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | \_\_\_\_ | Completely agree

Thank you for taking part in our EMOTIVE evaluation!

Would you like to participate in future evaluation of EMOTIVE application?    YES    No

## APPENDIX F: Transcription of Semi-Structured Interview with Users of Hunterian On-site Experience

*User 07: Female, age 20, Postgraduate Museum Studies Student*

*User 09: Male, age 24, Postgraduate Classic Studies Student, also a Hunterian MUSE guide*

Interviewer: So I'll just talk to you both together if that's okay. So firstly how would you describe the experience that you have just participated in?

User 09: Do you want to go first?

User 07: Describe? Oh okay, sure, I mean, I guess it was pretty, how'd you say inclusive, immersive experience. I mean it was pretty acted out as well, so yeah definitely immersive, and quite entertaining in the end, as well as informative.

INTERVIEWER: What did you find entertaining?

User 07: Entertaining, well the acting and the characters and the creation of the story. I didn't know, so I didn't know about the centurion before, I was wondering at the end if it was actually a true story or if it was completely fictionalized just to get an introduction to the actual objects. But then reading as well, reading around the objects as well I saw that it was true. But the immersion in the story was entertaining.

INTERVIEWER: Okay

User 09: For me I'd probably say more enjoyable like again I'm really sorry to keep harping on about the tours but when I do, I really like to focus on the life element rather than strictly looking at what you know measurements and strict statistics and data because I don't think people respond to that. So I really thought it was enjoyable to have such an interesting story kind of weaved around and I really liked it because normally I think everybody immediately looks at the distance slabs, the bigger things and it was nice to have your attention drawn to the kind of coin purse, nails and I think it brings on, it forces you to think about another aspect rather than the decorative feature you think of the actual construction, I think you think about that, you think about the people building it, and that leads you pretty nicely you into the idea of thinking about the people along the wall not just the wall itself. And in educational as well because when I do the tours I don't always focus on the nails and it's nice to revisit [recording interrupted]

INTERVIEWER: There we go.

User 09: And it was kind of like what you said there it was really well, the voice acting, I thought was really nice as well and not at times it can be over animated and hard to follow but was a good balance of engaging but also not losing the information there I really liked it and it kept my focus there.

INTERVIEWER: What was the most interesting part of it for you? If you had to choose something that you thought was the most interesting part of it?

User 07: Well for me I think as a specific object I think the window, I've never actually seen a proper window from that time so that story in particular I found was pretty fun and then just at the end the decision in the end, discovering what the outcome was.



User 09: I really liked the decision element it kind of, it was interesting as well if you followed it through, it kind of felt like you were informed to make the decision, oh I kind of know more what I want to do, and I really liked at the bit at the end the inclusion of social media to maybe interact with the museum, like tell us your, like what decision you made, especially it could be engaging with a younger audience as such a topic that more and more people are feeling out of touch with, I really liked. I also really liked the window as well, I never do the window on my tours, but my tours always run over time, but I never think of glass that much in the ancient world, I always think of stone and carvings, but no I liked that.

INTERVIEWER: And so that element of the decision part, you mentioned the social media aspect. [User 07], would you have liked to have seen what other people had chosen or would you have like to have, in terms of known what other people who had done the app had thought?

User 07: Yeah probably because the last question is probably the whole conclusion to that, just to see what other people got out of the experience, I don't know whether you choose for him to stay or for him to leave probably kind of informs on which part of the tour people responded to more, it was about Calle, the son, yeah I guess a statistic in the end why not.

User 09: I also think it would be quite interesting if you could have maybe a statistic at the end, how many people chose what, if you could see was it split 50 – 50, is it possible to leave behind why you chose your decision so you could, you may a lot of people may have chose for him to leave but they could all of had different reasons, maybe a logical one, maybe like everyone... I think that would be really interesting because you could think of it quite educational, oh I think he should leave for this reason, oh I agree but I think it was more this reason and it makes you think of another element which would be really interesting.

INTERVIEWER: Okay. That's great. So you both did the experience by yourself can you say a bit about that and how you found it, but obviously you don't necessarily have anything to compare in terms of doing it with somebody else. But how did you find, finding the objects or navigating the actual tool, the app and phone itself?

User 09: For me because I know the layout but I really liked, there was first of all the map and then the 360 so you could really easily find it, I think. Some of the objects, like the coin purse I think you could easily walk past, and the nails and the hammer. So I really liked it was very clear where it was going to be.

INTERVIEWER: So you used the map and the 360 panorama. Did you, I know you know the objects on display quite well but how quickly did you realize that those elements, those media elements were embedded within the app. Did you use them straight away? Or...

User 09: It was more I saw them and I wondered, map, oh it must just be of the exhibit perhaps. I was like I'll just test and if it's not what I think it is it's fine. So I just kind of clicked them all. I really liked the 3D image, you could flip and rotate it around I know there's often a difficulty in museums, like looking behind the case. It was interesting to have that. And I liked how it was on the distance slab it had it the most, the different highlighted bits and it broke down what you could see, I thought that was really nice.

User 07: Well, I'd say I was a bit disconcerted at the beginning on how exactly because I didn't really know what the format of the application was going to be. So I did skip directly without first going to the object. But then when I did realize how it worked, I mostly used the map setting to actually find the object. I tried the 360 but it wasn't as direct as the map I guess. The 3D tool was pretty ideal although I think I thought it was also another way to find the object whereas it was actually maybe considered maybe additional

information of the object so I didn't really know why it was next to the map and finding the object elements maybe. And also I don't know if there is a kind of a, so there was a kind of linear structure, at first you had to go through, ehm? I was a bit confused on the navigation as well. I think, I think we went different routes I think...?

User 09: We did the same option first and then

User 07: Right, so I mean, there is a possibility to navigate through the different chapters but I didn't really, I managed to have a clear idea of the mapping of the app I think.

INTERVIEWER: You did or you didn't?

User 07: I didn't. But otherwise it is also very possible to go just kind of go through it linearly. I mean it took me every time to the right place so it might that wasn't that much of a problem in the end.

INTERVIEWER: I think you did actually go through it quite linearly.

User 07: yeah that's it. So in the end it wasn't that much of a problem.

INTERVIEWER: No. Okay. [User 09] did you have any problems with anything at all?

User 09: Tech wise?

INTERVIEWER: Hmm or just understanding the usability?

User 09: Ehm it was fairly explanatory because it greyed everything out so you wouldn't click the same thing, which I liked and then once you got as far as you could within a certain segment you couldn't go any further so you just clicked the back button. But one thing I think might have been a glitch every time it came the shoe, every time I click it, it played the audio for the pots one.

INTERVIEWER: Oh that is a problem. So did you not hear the audio for the shoe?

User 09: No so I just, I read it. I should have maybe looking back taken the headphones out and just read it.

INTERVIEWER: Oh that is a problem but we have identified it and so it needs to be sorted.

User 09: But apart from that accessibility. But I think that's a really good point about the 360, not the 360 but the 3D model because as you said it doesn't help you find the object but it's useful still. I really like it but maybe separated from them.

User 07: Well under, for example when we hear the story of the object, maybe put it there.

User 09: One thing that confused me but I don't know it might be something reminiscent of an older patch of the app they don't say it but it is written down that you're supposed to touch the sticker. So I kind of did but I was like...what happens?

INTERVIEWER: It's okay. The audio and the text don't match and the stickers are not activated. So we're kind of slightly ahead of ourselves and slightly behind ourselves all at the same time. Well spotted. Okay I've already asked you but for the purposes of the taping, what did you decide for Ebutius to do when you got to the final decision?

User 09: I decided for him to leave but it was an investment of both the heart and the head I think, because although I think he should stay and support his family I was like realistically he would probably because I know the legions most of them were from, most of them got pulled up to Scotland from the UK, like England as we would say now, he wouldn't be so far away on a different country, although we can't really call it a country for, it's a bit confusing, but then in my mind I was like oh well he might be near the end of his service and if that was the case for the time he could have supported his family, sent some money and then gone up. But if he had acted impulsively and left he might have been killed and then the wife and the child would have been left on their own. So it was a balance for me of head and for heart where I was trying to play mediator for both sides.

INTERVIEWER: And what about yourself?

User 07: Well I have to say that I was really quite invested in the decision in the end and quite torn between the two possibilities and hesitated for long seconds. But I chose for him to leave just because of the objects, well there was the whole emotional side there but also talked about the idea of the functional uses of all of the objects and it's also the thought and the impression of the era, probably I don't know out of stereotype I don't know people probably had a more, bigger sense of duty and function within especially within the army. So that was the experience I got out of it.

User 09: I was quite torn as well. I think the reason I tried to do a split was because I kept going back and forth between the two but I was quite surprised how invested I was trying to make that decision.

INTERVIEWER: Is there anything else that I've not covered in this conversation that you want to draw our attention to?

User 09: I thought maybe kind of nice, I thought there wasn't much attention for the slabs and for the stuff in front of the window and then that much along this wall. I think, I know for a while, I think I remember maybe you were you saying Verecunda because I think that is such an interesting figure if that could be integrated that would be good.

User 07: I think more about the context of the, the whole historical aspect because I have not much of an idea of the whole historical background so at the beginning I was looking for yeah more historical context as to why yeah a bit more history.

INTERVIEWER: Okay. That's brilliant. Okay I'll pause that there then.

[end of recording]