

# Understanding by Doing – Guided Decision-Making as a Tool for User Engagement with Reconstructions

Marleen de Kramer, C<sup>2</sup>DH and IHIST, University of Luxembourg

## Introduction

The London Charter for the Computer-Based Visualisation of Cultural Heritage (2009) lays out best-practice guidelines for producing virtual reconstructions of historical buildings. This study seeks to further objective 3: “[To] ensure that computer-based visualisation processes and outcomes can be properly understood and evaluated by users”, meaning that “the outcomes of research that include computer-based visualisation should accurately convey to users the status of the knowledge that they represent, such as distinctions between evidence and hypothesis, and between different levels of probability.” (Denard 2009, p 4)

While users are often fellow cultural heritage professionals or other academics who are well versed in source criticism, knowledgeable about the limitations inherent in researching history, and know how scientific theories are formed and tested, we must also consider a potentially much larger group of users: the non-expert consumers of cultural heritage, such as museum visitors, the audience of televised documentaries or the readers of popular history books.

It is especially important for us as professionals to maintain the principles of “intellectual integrity, reliability, documentation, sustainability and access” (Denard 2009, p 2) in relation to these groups, as they rarely have access to the tools and training required to separate fact from fiction or archaeological evidence from speculation, or even know that they should. This is further reinforced by the “museum effect” (Putnam 2001, p34) – how something being displayed in a venue that carries the “Voice of Institutional Authority” (Weschler 1996, p 101) gives the impression of truth and authenticity, regardless of its actual proof and provenance.

To this end, it is important to understand how users think about historical reconstructions, so they can be presented accordingly. This study seeks to address the following questions:

- Do users gain confidence in their decision after being guided through the reconstruction process?
- Can they apply the process independently?
- How do they handle ambiguity or conflicting sources?
- Which arguments do they use to support their decisions, and which type of arguments influence them?
- How do they evaluate existing reconstructions?

## **Study Design**

Originally, this questionnaire was intended as part of a project at a heritage site, in which school groups explore a 3D model semantically enriched with metadata about the buildings and paradata about the reconstruction process, discuss it, research some of the buildings and annotate it with further information of their own. Some users were to complete a version of this questionnaire before viewing this model, others after. This way, the groups could be compared to investigate how being guided through reconstruction decisions influenced their comprehension. It has been adapted and expanded to stand alone due to the pandemic-related impossibility of working directly with large groups, and youth and adult versions were provided in English and German.

## Questionnaire Description

The core of this study is an online survey that demonstrates how the reconstruction process works, presented as a series of questions to be answered by often incomplete or contradictory data. The goal was to show that while one result may be more likely, there is not usually an absolutely “right” answer – though after the intense frustration of not knowing whether they are correct expressed by the testers, a most likely solution is presented after each set of decisions.

The questionnaire utilises the author’s ongoing work into reconstructing historical Larochette, Luxembourg. It is divided into two main phases – the introductory phase dealing with the city’s lower gate, and the more complex, abstract phase dealing with one of the buildings in the castle proper.

In the first phase, there are three sets of choices to be made about the building – positioning, shape, and direction. As the survey progresses, these become slightly more complex because they are more abstracted. The first question is a simple binary choice where respondents are told exactly what they are looking at and what it means, the second provides multiple sources from which to draw information, and the third clearly describes the two options but does not explain how to find an answer.

The images themselves contain no text to make it easier to translate the survey into multiple languages; as the responses remain the same, they can be combined into one master analysis or looked at separately to gain insights into differences between German and English speakers.

For purposes of this study, the real interest lies not in making users choose a version of the gate, but in the questions about their confidence in their answers – whether they felt more or less confident after making their decisions based on data, and what would make them change their minds again. These are asked at the end of the first section to ensure that users still remember their choices.

The second section asks users to make decisions about a specific building in the castle itself by comparing multiple sources. While there is guidance given on what to look for, the answers are no longer binary but involve several choices. They are then asked to analyse several existing reconstructions based on their mental image or decisions. These did not include the digital reconstruction, to ensure that the results were not skewed by users knowing it to have been made by the author of the study and choosing it to please them. The optional introductory text included an image showing a volumetric model of the castle and town, with an incorrect gatehouse based on an erroneous map drawn by an earlier scholar, but refuted by the historical images. Several users cited it in their reply, assuming it must be correct as it was chosen for inclusion, thereby indirectly validating this approach.

The demographic questions help group similar users for the analysis. Among common demographic survey questions, gender, nationality, and especially ethnicity were not deemed relevant, so they were not included – after all, best practice is to collect only necessary personal data, not as much data as possible. Age and educational background help gauge a respondent's experience in source criticism and reasoning processes, while users working in cultural heritage and related fields will tend to be better at analysing maps and historic images and have more exposure to reconstruction methodology in general.

## **Test Phase**

The questionnaire was tested and refined as part of the User Experience Design course taught through the psychology department. Unfortunately, due to the social distancing situation, it was not possible to invite users to test it in the lab and observe them directly. Instead, a meta-survey was incorporated into the questionnaire, with questions about the reconstruction questions, how they made users feel and whether they understood the tasks. It was tested by 19 users, either participants in the course or recruited through social media.

All testers enjoyed the exercise and the general concept, but most complained about the lack of a “correct” solution, so a de-briefing was added after each section. To

balance the risk of information overload and need for metadata, a brief history was added to the website as a pdf, so that reading it was optional. As they felt there was too much time-consuming free-form writing involved, several questions were updated to multiple choice. Over ¼ of users struggled to interpret some content, so images were adapted to make it clearer where the focus lay, and architectural terms were better explained for lay people. Some survey respondents still struggled, especially with the word “gable”, but these respondents often commented that they were not native speakers. Others had trouble distinguishing between the town walls or gate and the castle itself, conflating the two.

## Evaluation

In total, the English adult version had 264 responses, of which 259 were valid, the English youth version had three responses, all valid, the German adult version had 105 responses, of which 101 were valid, and the German youth version had two responses, both valid.

The low participation rate for the youth versions meant that they could not be evaluated; this was most likely due to issues around getting parental consent outside a school setting where formal systems are in place.

Of the 9 invalid responses, 6 were accidental double submissions, 2 failed the screening question whether they had been answered seriously, and one was deemed invalid as most responses were non-answer. This left 360 responses for evaluation.

How frequently the survey page was accessed was not monitored, so it is not known how often it was started but not completed.

The responses were encoded for quantitative analysis, but also considered individually for qualitative analysis. Results of both will be shown below.

## Qualitative Analysis

In this step, specific individual responses were analysed and grouped by theme to show the broad range of responses and their sometimes contradictory nature. These have been anonymised and are attributed using a code made up of the respondent language and number. They have also been lightly edited for spelling and grammar to aid in comprehension and to make them more uniform. German quotes were translated into English by the author.

## Historical Understanding

The responses indicating historical knowledge – or lack thereof – are of particular interest for future iterations of this survey or its application at heritage sites, as they show where there is potential for confusion or bias.

Some respondents were entirely aware of their limitations when it came to historical understanding, and clearly able to identify where those ideas had been formed. Computer games and movies were a common theme – a field of research among medievalists in its own right. (Jiménez 2010, Spring 2015)

*“Most of my ‘knowledge’ in this regard stems from video games (like ‘Kingdom Come Deliverance’ as a current example).” – EN 200*

*“Maybe I just have a Disney idea of castles. (US citizen)” – EN 186*

*“I am so biased by former knight’s tales (e.g. public history ;- ) to guess that it might be used partly as a toilet – EN 132*

Others clearly had misconceptions – whether lacking understanding of the sequence of events, not realising that our modern landscapes are largely man-made, or failing to understand that basic architecture transcends epochs.

*“Round towers are Roman.” – EN 27*

*“A panzerfaust would probably light the [wooden parts of the castle] on fire, though at the time inexperienced kids operated them, so...” – EN 195*

*“Landscapes like rivers don't get up and move.” – EN 120*

Finally, some respondents used the comments to demonstrate or report their detailed historical knowledge, while being aware of their own limitations:

*“It's an aesthetic response. It looks like the early figurative work of Kandinsky. A mix of National Romantic (called Arts and Crafts in Britain) sensibility with a bit of antiquarian knowledge of the building. As an architect myself, I would not be able to trust the reconstruction because I know (and enjoy) the temptation to make things up for completeness when the evidence is not quite there.” – EN 190*

*“I am active in Living History at a museum level, and while I haven't occupied myself specifically with castle construction, I believe that I have above-average knowledge of the subject. Which does not preclude misinterpretations, as became clear over the course of this survey.” – DE 6*

## Bias and Self-Confidence

Their self-perception of their previous knowledge may have influenced respondents' confidence in their own answers, which varied across a very wide spectrum, from “none” to “absolute conviction that their personal theory is correct”.

*“I need to say here I'm not comfortable in claiming any certainty about anything! (...) Given my level of uncertainty, a sudden noise in the street [could make me change my mind]”. – EN 37*

*“Pretty close to what it actually looked like.” – EN 29*

Some users assumed ulterior motives on the part of the author and made their choices accordingly. This may be due to the “quiz” title under which the survey was distributed, which was chosen to encourage participation. Perhaps this made them want to “win”, despite having been told in the introduction that all replies were valid.

*“You got me on the first quiz, the round was so obvious I thought it had to be rectangular.” – EN 212*

*“I did not change my opinion because it clearly depends on which map you saw last”. – DE 57 EN 204 “*

*“This could be a trick question, because the old structure could have been torn down between 1778 and 1824.” – EN 195*



A large number of respondents argued from personal conviction, often ignoring the new data presented if it did not match their initial guess. Some simply stated that they were unwilling to deviate from their initial choices.

*"[I chose this source because] its opinion is identical to my decision." – DE 3*

*"Because I'm stubborn." – EN 86*

*"The reasoning I made up makes sense." – EN 92*

*"That's what I think – I might be wrong, but that is my opinion." – EN 82*

*"I'm sticking to my choice because I like it." – EN 8*

Some cited a pre-conception, without explaining how it was formed.

*"Left is best." – EN 248*

*"Round towers are better..." – EN 243*

*"With the French name for the town, I felt that a soft, round turret would be used." – EN 202*

Some had a picture of what a medieval building "should" look like.

*"It's suitable to my opinion of what a castle from that time and place should look like but I would add round features." – EN 123*

*"Robust, functional buildings, that's how I imagine the medieval mentality." – DE 28*

*"Corresponds with my internal model of how a castle looks." – EN 147*

Others were aware of their own biases, but consciously maintained their previous theories.

*“It appears from the photos to be round. But I am sticking with my first guess. Wrong or not.” – EN 175*

*“I am biased by the idea that it's on the left, to be honest.” – EN 132*

*“My answer is mostly linked to a feeling rather than a real justification.” – EN 88*

Finally, many respondents reported that they gradually changed their minds while aware that they were doing so – as this was the aim of the exercise, these responses were particularly gratifying.

*“I found it interesting that my functional reasoning was apparently contradicted by reality.” – DE 65*

*“Very interesting how the gradual presentation of extra material in the first part made me change my opinion of the tower design.” – EN 112*

*“I want to believe that my earlier guess was right but it looks round on these images.”- EN 190*

*“I'm afraid that my initial enthusiasm for the positioning on the left is being weakened by this very clear image. Presumably, the gate also did not go through the tower. I was wrong about that, too.” – DE 30*

*“I tried to lose my pre-conception that I wanted to be ‘right’ on the previous one. Then I looked at both maps and tried to see what seemed more ‘logical’ to me.” – EN 46*

## Certainty vs. Ambiguity

Another kind of certainty to be considered is not users' confidence in their own choices or in the work of experts, but how certain one can even be about various aspects of a reconstruction. Due to the sparse and varied nature of sources available to document the historical state of a building, there will always be parts of a reconstruction that are speculative or otherwise ambiguous. The only way to achieve absolute certainty would be to have the building fully preserved and unaltered – which would render a reconstruction unnecessary, and still be lacking data for some aspects, such as its day-to-day use.

This ambiguity, and how to manage it, is a bone of contention in the digital reconstruction community. For physical reconstruction and restoration, the rules are more clear – do nothing that can't be undone, preserve the original, and clearly mark what is new for posterity (ICOMOS 1964) – but because digital representations don't interfere with the original substance and can co-exist, the freedom and temptation to speculate are much larger. Attitudes range from “everyone will know it's only a theory” to “show nothing you can't prove”. However, such representations establish seeing habits – there is no way to not make a statement. Many reconstructions borrow from architects' volumetric models and show all-white, abstracted shapes – but as can be seen from depictions of classical Greek and Roman architecture, this leads to the assumption that the buildings were white. Showing raw stone, on the other hand, because we do not know exactly what was painted on the render leads to the impression that castles were plain and grey. Some artists' impressions are designed to be evocative rather than purely factual, prioritising emotion and a sense of presence over proven fact; other representations err so far on the side of caution that they are nothing more than a sober stock-take of the ruins.

How to balance these conflicting needs is a question that has not yet been answered conclusively, with many researchers working in parallel toward a satisfactory solution. This study is, in itself, a contribution to this collective effort.

Interestingly, the comments from survey respondents mirror this wide range of possible responses, from rejecting ambiguity to celebrating it, from valuing certainty over completeness to fully embracing speculation.

*“For a layman like myself it’s all about impressions rather than knowledge”. – EN 249*

*“For a virtual reconstruction, don’t fall into the ‘horror vacui’ trap and attempt to reproduce everything photorealistically; instead, use our 500-year-old seeing and interpretation habits of abstracted graphics for virtual reconstructions as well. Clearly show what we know and what we don’t, using transparency, for example.” – DE 95*

*“[The physical reconstruction is best] because you can imagine the rest yourself – like the difference between a book and a film. The other three constructs don’t allow for any alternative interpretations.” – DE 62*

*“Considering the sources available, this reconstruction is the most pragmatic. However, it is missing the colour scheme of the model, if it is attestable. Minimalism, especially in castle reconstructions, tends to generate an incorrect impression – like missing render on the exterior walls.” – DE 33*

The sources shown in the survey were specifically chosen for their differences, forcing respondents to choose between two options. Some respondents attempted to reconcile the ambiguity caused by conflicting sources by finding explanations for the conflict.

*“It looks rectangular with a round end, but you didn’t offer that as a choice.”- EN 185*

*“The first two pictures, especially the first one, give a strong impression of a round tower. Are we sure that the ruins were still standing in 1883? Could it have been replaced with a foundation of another building, perhaps built from the tower stones? This is really a baffling one.” – EN 73*

*“The previous question is hard to answer the way it is worded. Because the maps are from different times, they could both be correct. The layout of that area could have changed.” – EN 6*

*“The views are from different locations, elevations, and times of day as well as years. (...) The different perspectives also show that the artists had slightly different intent, one is a street scene emphasising the near buildings, one from a rocky outcrop with a view across the river valley to the castle. (...) So I see different artist’s intent, locations, time of day and town development. I wouldn’t rule out eyesight either.” – EN 64*

Others rejected having to make a decision at all, uncomfortable with having to commit without feeling that they were fully informed. Eliciting this discomfort was a specific aim of the study, to recreate the position in which those making scientific reconstructions found themselves and to demonstrate that certainty is not possible with the limited data available.

*“The question of good, better, best is qualitative. The alternatives offered for comparison are too different to be subjected to a uniform qualitative verdict.” – DE 79*

*“So far, the illustrations are too vague to make a decision.” – DE 33*

*“Normally I would trust a more recent cadastral map more than a more ancient map. The sketches in which the roadway seems to go through other features makes me confused, though. If I would really have to decide I would want to reference both maps to the existing situation and see what seems to fit best. Though that might even further complicate things.” – EN 91*

*“I had to make a decision, and the red marking saying this was a required question was right underneath this answer, so I picked it. I did not have enough information to base a real decision on.” – DE 80*

*“I liked this exercise, but I would have liked a few more ‘I don’t know’ options! (but I realise that too many people would probably have picked it, so forcing an answer at least forces one to rationalise to some degree.)” – EN 168*

## **Purpose**

An important part of source criticism is considering the *purpose* for which a depiction was created. The respondents’ very different views on this became especially clear in the question of which reconstruction is “best”, ranging from outright rejection of the principle to embracing all possibilities.

*“I am against reconstructions.” – DE 8*

*“We don’t know what was there, so the reconstruction can be chosen aesthetically. (...) I think it is better to be honest about it being a 21st century reconstruction.” – EN 19*

*“I have chosen every one of the above models and rejected each to finally settle on the physical reconstruction. Each poses problems. Each proposes possibilities, and I’m not convinced by any, but the physical one is doing the least alteration.” – EN 214*

*“A drawing can be used to communicate one’s ideas and dispute them with others without interfering with the original object. Mistakes are easier to rectify, and to admit to, than if you are standing in front of a completed project. Today, the castle is a tourist attraction, whether authentic or not, that is useful to the community. The reconstruction drawing can be augmented with a 3D model to show all perspectives. The drawing still serves better for drawing attention to scientific findings and expounding on them.” – DE 36*

Another point of contention was whether speculation was to be welcomed or avoided.

*“Do we even need a visual reconstruction? Aren’t the preserved ruins and one’s own imagination enough?” – DE 80*

*“Good balance between information from actual ruins and speculation.” – EN 31*

*“It is best because it leaves the least room for speculation.” – DE 98*

Many respondents believed that the physical reconstruction had a purpose beyond built heritage conservation, from tourism to education. They conceded that this might come at the cost of scientific accuracy if it captured the imagination; this mirrors the approach used by some documentary film-makers and their “the creative treatment of actuality”, which can introduce artistry, plot and dramatic tension. (Kerrigan & McIntyre, 2010)

*“‘Best’ is not the same as ‘most accurate’ for me. The best one is the one that is in the physical location, and engages with the town and the environment, preserving the structure (or a reasonable approximation of it) for posterity.” – EN 34*

*“It’s history you can touch. It’s one thing only to read about it and have to imagine it yourself, and quite another to really be able to go there and see it with your own eyes”. – DE 15*

*“A material reconstruction must be examined very critically, because often, modern materials are used to cover the original remains. This aspect alone is in no way related to the original! But at least the attempt is made to preserve the building and continue using it, which I would gladly see done all over the country! Interested parties can learn what is original and what is fantasy by taking part in guided tours of the site or by reading.” – DE 36*

However, proponents of other reconstructions also had compelling arguments for their purpose, often incorporating an element of documenting metadata, which they felt drawings and models were better able to accomplish than the in-situ reconstruction. This is reflected in the fact that it was not the physical reconstruction, but John Zimmer's perspective drawing that proved the most popular among respondents.

*"[This drawing] gives the easiest to understand perspective on the entire site. The model might be better if you can see it in person, but a 2D photo of a 3D model is not as good as the drawing." – EN 252*

*"[John Zimmer's drawing], because it's the simplest. The reconstruction also has a high value. However, it's hard to show on site what we know and what we don't, so the impression remains that 'THIS is what the castle was like'." – DE 103*

*"A drawing is a reproduction of an idea, and can be adapted to new findings. Any physical reconstruction, whether a model or 1:1, pretends to be authentic to the original." – DE 28*



## Metadata & Source Criticism

The intentionally leading question “what could make you change your mind” about the respondents’ chosen final reconstruction offered several options that corresponded to the codes consensus, authority, and metadata, but also allowed users to add their own response. The overwhelming majority of the “other” responses called for more metadata. However, this was often mediated through an authority who communicated the data, rather than through direct access to data. This is an interesting insight into the wants of users – someone they can trust giving a substantiated explanation, rather than users interpreting data themselves.

*“A time-traveller who explains exactly how it looked back then.” – EN 43*

*“A friend who is a history nerd picking something else based on the maps and drawings and explaining it.” – EN 62*

*“Both 3D model and drawing + explanation if they were created by a historian. In this list we don’t know who created them – could be a games designer for all I know.” – EN 71*

In general, users seemed to trust those in authority to have made the correct choices, even without being given metadata to corroborate those choices.

*“The source is known, and therefore I trust it more than the anonymous scale model.” – EN 205*

*“The reconstruction will have been preceded by all the drawings we have seen, and probably much more. And historians, archaeologists, people who knew what they were talking about. So while the drawings are fine, the physical reconstruction goes one step further and will be the most right. Otherwise, they would not have transferred their findings into stone and into reality!” – DE 37*

Others, however, worried about the availability, reliability and use of metadata, showing an awareness of the need for source criticism and documentation.

*“My original degree was classical archaeology. It seems even harder to clarify later architecture because there are maybe too many sources?” – EN 148*

*“I would view most of these options critically: does the historian have proper evidence to back their claim up? How well grounded is the explanation with the drawing? How natural is the style of the medieval painting? But if the arguments are backed up convincingly, then yes.” – EN 37*

*“I like it, but they created something without a body of evidence that it’s the same as the original.” – DE 28*

Finally, the replies included a comment from a respondent who had clearly watched the TEDx talk linked to from the “more information” section of the survey website.

*“I vaguely remember you complaining about this scale model before, though I don’t recall the details.” – EN 97*

Considering that the complaint in question was about the lack of metadata and paradata and the importance of including the reasoning for decisions in a reconstruction, the irony here perfectly encapsulates the issue that this study is designed to help solve.

## Feedback

Respondents also left feedback about the design and usefulness of the questionnaire itself.

Some made it clear that the purpose of the questionnaire had not been explained well enough. While the description was kept deliberately vague to avoid bias, it was made clear that there were no “wrong” answers, and that it was not meant to crowd-source the “correct” answer. However, some users clearly did not understand this, and were bewildered as to the purpose behind it.

*“This is so bizarre – I have no archaeological evidence to work from, and I am not an expert in French castles to make even a slightly educated guess. What are we gaining here?” – EN 191*

*“This questionnaire is confusing. I don't understand how and why you are asking the opinion of a layman.” – EN 15*

*“I did my best to make it useful for your research, but as I saw, I made a lot of mistakes.” – EN 259*

However, this experience was not common to all respondents.

*“I very much enjoyed this. It was presented in a way that allowed me to feel that my ideas and opinions mattered even though I am not very learned in history or architecture.” – EN 202*

Others were frustrated by the inherent ambiguity and being forced to make choices – this was intentional, to demonstrate the impossibility of giving absolute answers, but left some users feeling unsatisfied with the process.

*“You should allow for 'can't say' answers in the Likert scale items. I answered all of them seriously except the last two Likert scales. I picked a number that 'felt right' because I had nothing to go on.” – EN 98*

*“Nice questionnaire, but it ends with a cliffhanger. What do you think is the best representation based on your research?” – EN 235*

On the other hand, several respondents, both professional and non-professional, explicitly stated that this questionnaire helped them understand reconstruction processes better, the stated aim of the exercise.

*“I found it fascinating to be able to be able to comprehend a reconstruction by re-tracing it.” – DE 64*

*“It's very interesting to think about what goes into recreating structures for which we have limited information.” – EN 105*

*“This really makes you examine your reasoning based on evolving evidence and review of facts.” – EN 138*

*[responding to the question “which of these could make you change your mind? ] “The process of this questionnaire!” – EN 87*

Several respondents also agreed on its general applicability and usefulness as a teaching tool for use with reconstructions.

*“This is an excellent way to encourage people to think about old buildings and will make a visit to the site much more meaningful for them.” – EN 54*

*“Nice survey. Lots of fun to take. I wish other students of history would take this approach to surveys and research as it was both educational and challenging at the same time.” – EN 79*

*“Keep it up! Intriguing questions concerning authenticity and perception of reconstructions.” – DE 92*

## Professionals vs. non-professionals

Of the demographic information collected – age, education and professional background – the latter was the most interesting for the analysis, with age and education primarily being used to ensure diversity among the respondents.

The provided choices of relevant professions – architecture, history, cultural heritage and archaeology – were supplemented by those deemed relevant during the encoding process, and included professional costume reconstruction, civil engineering and stonemasonry, among others. Their common denominator was familiarity with skills such as reading maps and technical drawings, constructing buildings, or evaluating historical data. Both the responses from German and English-speaking professionals were combined for evaluation, as were those of non-professionals.

In some cases, the results hardly varied between groups, in others, there were significant differences. Some were to be expected: professionals were more likely to value metadata, more sceptical of choices made without them but more confident in their choices after they had viewed the sources provided, less likely to make unsubstantiated guesses and more likely to see the significance of context and historical artefacts or base their decisions on specific details. Unsurprisingly, they also relied on their previous knowledge more often and provided more specific sources or direct comparisons, being more familiar with the subject matter.

However, some differences were unexpected. Professionals were, on the whole, less likely to indicate that they had changed their minds, and were more likely to accept authority as a valid argument, though the latter may be due to their confidence in their colleagues' good research than in their lack of a need to see data for themselves. They were also significantly less likely to argue for what makes a good reconstruction by citing realism or authenticity.

While they were much less likely to use aesthetics as a basis for their reasoning, professionals were also significantly more likely to employ emotional arguments for what makes the “best” reconstruction– perhaps demonstrating their passion for the

subject matter at hand. Some also showed a willingness to ignore data, accuracy, or certainty in favour of subjective taste:

*“If I was working on the building, [the Zimmer reconstruction] would be useful but I would rather put the Koenig drawing on my wall.” – EN 190, an architect*

*“I chose the physical reconstruction because I see it as more impressive, more tangible, more immersive even. To be honest, it might be the inner child speaking more than the historian (re the accuracy of the reconstruction), the emotional / impressionable side taking over the reasonable one.” – EN 44, a historian*

*“It's the prettiest. I know it's probably the least accurate, but 'best' to me means 'I like it the best'. – EN 204, a civil engineer*

*“In my opinion, no reconstruction is ‘the best’, each is a product of its time and of particular desires and visions. Zimmer’s drawing is presumably the most ‘accurate’ reconstruction drawing, but the works of Koenig and the model builders express an equally interesting, and therefore ‘good’ desire for idealised embellishment.” – DE 100, a historian*

# Quantitative Analysis

## Encoding

To make it possible to perform a quantitative analysis on free-text answers, show general trends and correlations, and visibly represent changes in attitude, the responses were encoded with key words. To reduce the influence of the author's own bias, the first 10% of responses were encoded by multiple researchers, then compared and a final code set generated from their combined lists.

In total, this list comprises 36 codes in six categories:

- Knowledge – which type of information informs the user's decision
- Reasoning – how the user reaches a decision
- Decision – the type of decision made
- Argument – which type of argument they present for a decision
- Reflection – user demonstrates that they are thinking about the process
- Issues – any problems that arise during the survey

These each convey a different type of information for analysis, from whether the user was able to complete the survey without difficulty to which type of arguments they employ to whether they are thinking about the reconstruction process or the survey on a meta-level. "Frustration" was not counted among the issues but rather as a type of reflection, as it is a reaction to ambiguity or their own perceived lack of information or knowledge. Several respondents used emoji in their comments; these were encoded as "emotion".

The code "evolution" was later added to avoid mis-using the codes "time" and "change" when "change over time" was given as an argument, or to dilute "context" by using it for temporal context as well. The term "detail" was also challenging to apply consistently to refer to users arguing by referencing a specific detail, rather than when terms like "more detailed" was used, which falls under the code "medium". To ensure quality and consistency, all codes were double-checked on completion.

## Full Code List

<b>Knowledge</b>	prior	References their own perceived prior knowledge
	source	Cites a specific example
	authority	Appeals to an expert or official
	consensus	Agreement of multiple indicators or sources
	emotion	Shows emotional response
<b>Reasoning</b>	context	Physical context, surroundings, geometry
	function	Fulfils a specific requirement, logistics, purpose
	inference	Attempt at logical deduction from own prior knowledge
	intuition	Feeling or guess with no reason given
	interpretation	Deduction from data provided
	artefact	Specifies historical thing that still exists
	comparison	Weighs relative merits of two possible solutions
	evolution	suggests change over time
<b>Decision</b>	rejection	Rejects one of author's theories or facts presented
	approval	Confirms theory offered
	sceptical	Questions theory offered
	confirmation	Repeats own previous answer, doubles down
	change	Changes mind about previous answer
	bias	Refuses to examine new data and repeats previous theory
<b>Argument</b>	detail	Specific artistic detail as indicator of accuracy
	time	Later or earlier time as indicator of accuracy
	aesthetics	Appearance, artistic interpretation, or cultural implications
	realism	Classified as realistic/not realistic or accurate/inaccurate
	personal	Reasons particular to the user (bias but not rejecting data)
	medium	How something is made or shown
	purpose	Specific intent by creator of a representation
	physicality	Material properties
<b>Reflection</b>	self-awareness	Identifies own personal bias or limitations
	contradiction	Contradicts own statements
	ambiguity	Points out or comments on ambiguity or lack of certainty
	metadata	Wants more, specific information
	frustration	Expresses annoyance or frustration with survey or self
<b>Issues</b>	non-answer	Response not a valid answer
	comprehension	Problems with understanding, something is unclear
	invalid	Responses not useable – remove this respondent
	technical	Technical issues experienced



## Reasoning and Arguments

Which types of arguments respondents use to justify their choices, or the reasoning used to make them, is especially important for understanding how to communicate with users – which type of arguments are they especially receptive to, and how do those arguments map to the types of data available?

The first point of interest was whether respondents would agree with general consensus, defer to a perceived authority, or require knowledge of the metadata behind the information they were given – though of course, this type of survey only yields their own perceptions and not an objective analysis of whom they believe.

Besides the codes assigned to comments, this was also explicitly asked in one multiple-choice question:

<b><i>Which of these could make you change your mind?</i></b>	<b>Corresponding code</b>
<i>Most people getting a different result.</i>	consensus
<i>A 3D model that looks very real.</i>	authority
<i>A historian writing something different in a book.</i>	authority
<i>A reconstruction in a museum that shows a different gate.</i>	authority
<i>A drawing that shows a different gate, with an explanation of why they drew it that way.</i>	metadata
<i>Seeing a medieval painting where the gate looks different.</i>	metadata
<i>An archaeological excavation that found remains of something different.</i>	metadata
<i>Other [free text]</i>	[individual]

The second subset of codes to be investigated more closely is those corresponding to the degrees of certainty assigned to different parts of the reconstruction (unbeknownst to the respondents) in the virtual model. These degrees, from highest to lowest certainty, are

*Relict – Interpolated – Extrapolated – Speculative*

“Relict” covers elements for which evidence survives from the time of their creation. “Interpolated” refers to consulting several nearby data points, e.g., filling a gap in a wall along an existing foundation. Where this “interpolated” result is a line between two points, an “extrapolated” one is a vector, using a solid point of reference augmented with secondary and tertiary sources. “Speculative” results are obtained using only secondary and tertiary sources, e.g., comparing with similar sites or using engineering knowledge to estimate heights of walls. (de Kramer 2020)

They are associated with the following codes, denoting different types of reasoning commonly used by respondents to explain their choices.

<b>Code</b>	<b>Certainty</b>
<b>artefact</b>	relict
<b>interpretation</b>	interpolated
<b>context</b>	extrapolated
<b>function</b>	speculative
<b>inference</b>	speculative
<b>intuition</b>	guess (outside certainty chart)

Finally, the key to assessing respondents’ understanding of how reconstructions were made was to analyse which arguments they would use to choose the “best” out of four reconstructions. These had minimal metadata – a year, an author, and the author’s profession – and users were asked to give their impression of each and gauge its accuracy before choosing one, then explaining that choice.

Once the responses were encoded, the codes could be analysed to show general trends. As there were clear differences between the versions, they were investigated as a combined total, but also as professional<sup>1</sup> vs. non-professional. Some interesting differences between German and English speakers have also been noted, though the cause of these differences is unclear. 33.6% of German respondents and 19.7% of English respondents indicated that they worked in a relevant field, for a combined total of 23.6%, or 85 vs. 275 users.

---

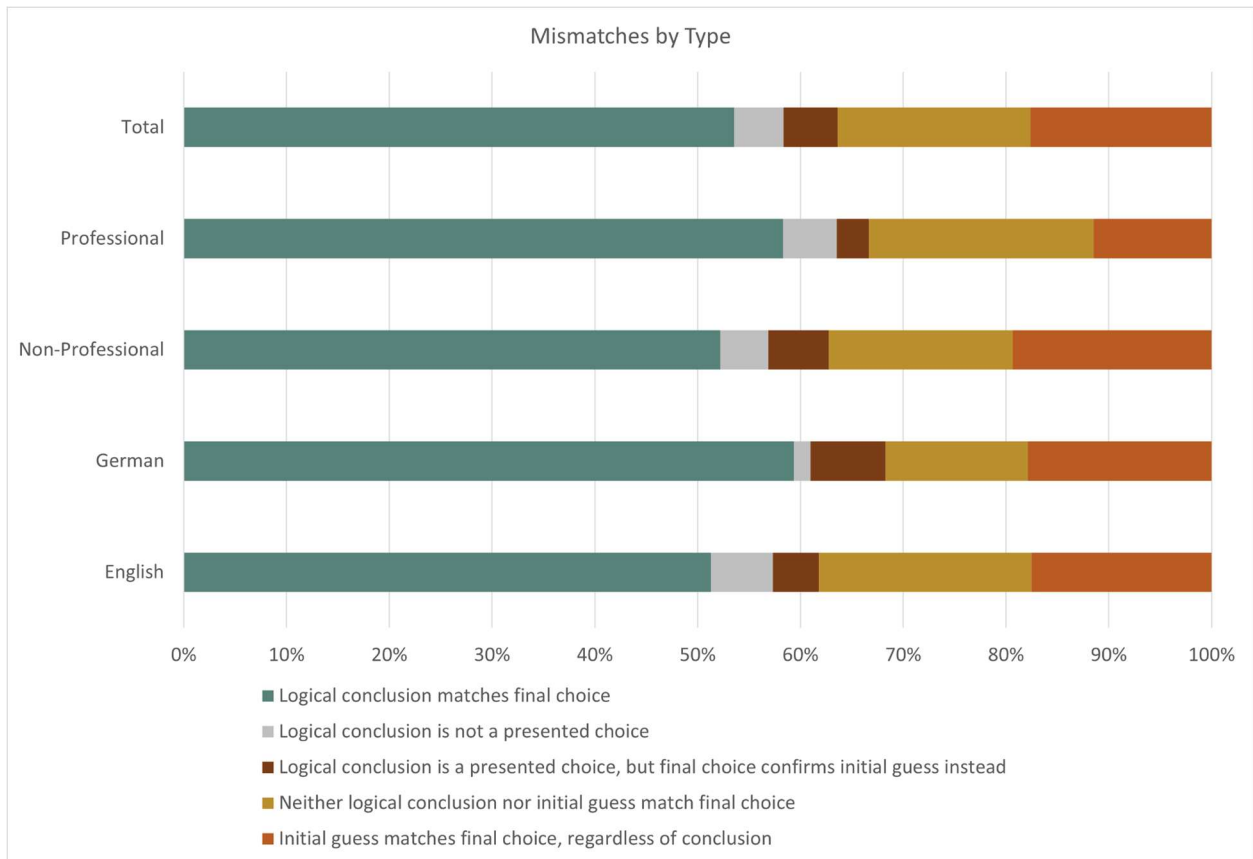
<sup>1</sup> Respondents who worked in architecture, history, cultural heritage, or archaeology were assumed to have a different understanding of the issues involved and be better at interpreting the visual sources.

## Mistakes and Mismatches

The survey makes no claim to be unbiased, but instead subtly steers participants toward one “correct” solution, i.e. the conclusion its designer reached through analysis of the same data, which were then chosen to develop this viewpoint – after all, the purpose is guiding users through the process, not crowd-sourcing a solution. Examples of such consciously designed bias toward one solution include not even offering one possible solution pair as a choice (round or rectangular tower with gate on the left), requesting confirmation of the placement on the left with new, contradictory data, but not for placement on the right, and showing historical images with a tower on the right side of the bridge even after the user has chosen “left” for other steps.

Therefore, it is surprising that only a little over half of users chose the “correct” combination in the end. Furthermore, many respondents chose a final version of the gate that did not correspond to the sum of their choices, meaning they had either forgotten which decisions they had made or decided to ignore them in the final step. These decisions were further investigated to explore whether the respondents simply confirmed their original choice despite their data-driven decisions (showing bias) or had made an entirely new decision. 65% of users made a final choice that matched their decisions, 6.4% confirmed their initial choice, 5.8% had to choose something different because their logical conclusion did not match any of the versions offered, and 22.8% chose something that matched neither their initial guess nor was a logical result of the decisions they had made – which may mean that these users did not properly understand the task they were assigned.

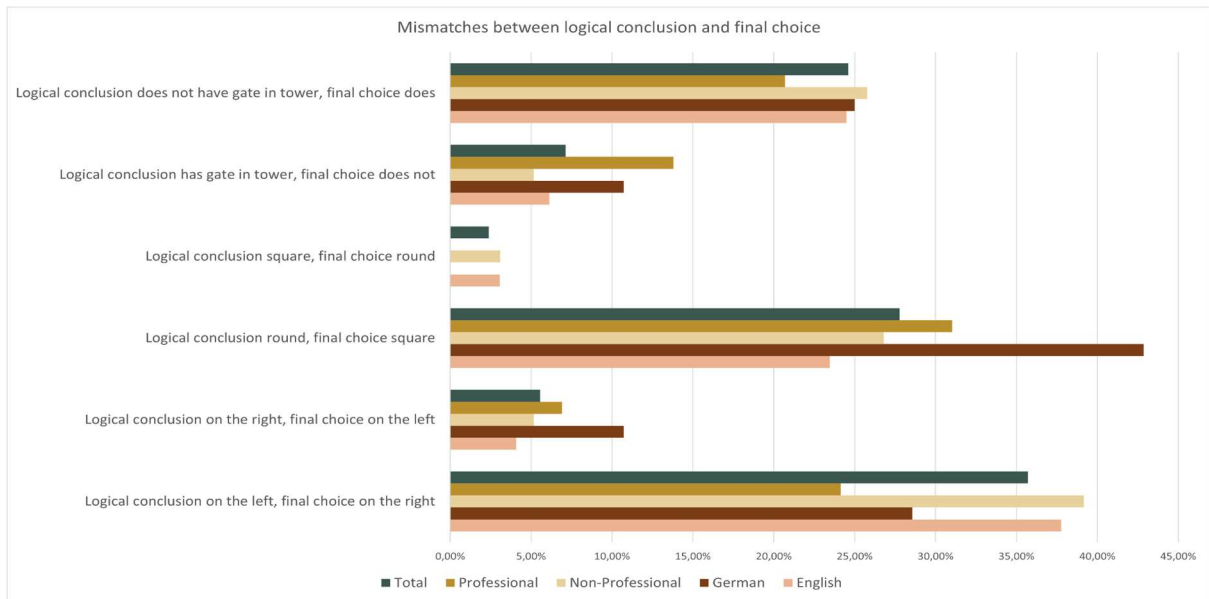
While the overall rate of mismatches varied little between professionals and non-professionals, it was significantly lower among German speakers (27.7%) than English speakers (37.8%). The reason for this difference is unclear – if it were caused by a poor translation, the higher error rate would be expected in the target language rather than the original English.



In just over 35% of cases, the issue was that left and right had been swapped – and in the vast majority of these, the “correct” final choice of the tower on the right was made, despite their conclusions that the tower was placed on the left, which suggests they were influenced by depictions in later choices.

This phenomenon was much more prevalent among non-professionals than professionals (39.2% vs 24.1%). Professionals were also more likely to make the “mistake” of choosing a rectangular tower when they had concluded it was round, at a rate of 31.0% of professionals vs. 26.8% of non-professionals.

For all groups, the third most common error at 24.6% was coming to the logical conclusion that the gate was not in the tower but the wall, then choosing a final option that had the gate in the tower. Notably, the error of choosing a round tower when their conclusions should have led them to a square one was the least common overall, and was not made by professionals.



There was little difference between non-professionals and professionals in making decisions not adding up to any of the variants offered, at 5.9% and 5.9% respectively. Non-professionals were 41% more likely to show bias than non-professionals (31.6% vs. 22.4%).

Among those whose final choice matched their initial guess (regardless of whether it matched their decisions), 40% had shown bias elsewhere – many more than the 6.4% who confirmed their guess despite their logical conclusions. This might mean that they made choices that confirmed their guess, consciously or subconsciously, then argued for them. Professionals had the lowest overall rate (18.2%) of confirming their initial theory in their final choice while also displaying at least one instance of bias – an expected result, due to their training. Interestingly, they did not exceed the total average when it came to making a final choice that was a logical conclusion of their decisions.

## **Bias, Metadata, and Self-Awareness**

At the opposite end of the spectrum, 35% of respondents showed awareness of their own limitations or biases at least once, and 29.2% expressed that they wanted more metadata to make a decision. Interestingly, these two phenomena did not show a strong correlation, with only 13.9% of respondents displaying both.

Despite less than 1/3 of total respondents explicitly wanting more metadata, in answer to the question what would make them change their minds, 98% indicated that they would be convinced by metadata, with 42.8% saying they would believe an authority and only 5.6% being swayed by consensus.

However, among all 8834 codes analysed, the argument of “consensus” was used 56% more often than “authority” (253 vs. 162). 256 responses indicated prior knowledge<sup>2</sup>, though only a further 55 cited specific sources. 163 codes showed emotional arguments, 338 were for personal reasons, and 136 instances of clear bias were recorded.

---

<sup>2</sup> Excluding those who indicated sources; “knowledge” is taken to mean their own perceived general knowledge of the matter.

## Summary of All Responses

Across all responses, the most frequently encoded were realism (601 or 6.8%) and function (593 or 6.71%) followed by interpretation of data shown (540 or 6.11%). The arguments most consistently applied across all categories were (geometrical) context, scepticism at theories presented, inference from their (perceived) previous knowledge and intuition or guesses. A sceptical response to the theories presented was the most similar across all categories, making up 2.4% of total codes in all cases. The largest difference between the groups was found in their use of function as an argument, which accounted for 7% of non-professional, but only 5.8% of professional codes. Conversely, professionals (3.7%) were significantly more likely to argue using the purpose for which a reconstruction was made than non-professionals (2.5%). Professionals (3.1%) and non-professionals (3.1%) did not differ in their use of physicality as an argument, but professionals (2.0%) were somewhat more likely to cite artefacts than non-professionals (1.7%).

<b>Analysis All Codes</b>	<b>Code</b>	<b>EN %</b>	<b>DE %</b>	<b>PRO%</b>	<b>NO %</b>	<b>Total %</b>
<b>Knowledge</b>	prior	2,84%	3,04%	3,51%	2,70%	2,90%
	source	0,60%	0,68%	1,05%	0,48%	0,62%
	authority	1,78%	1,96%	2,60%	1,58%	1,83%
	consensus	3,32%	1,72%	2,33%	3,05%	2,86%
	emotion	2,16%	1,04%	1,78%	1,87%	1,85%
<b>Reasoning</b>	context	4,52%	4,67%	4,38%	4,63%	4,56%
	function	7,85%	3,84%	5,84%	7,01%	6,71%
	inference	4,69%	4,51%	4,84%	4,58%	4,64%
	intuition	4,82%	5,07%	3,93%	5,22%	4,89%
	interpretation	5,67%	7,23%	6,21%	6,09%	6,11%
	artefact	1,80%	1,72%	2,01%	1,70%	1,78%
	comparison	2,72%	3,28%	2,65%	2,95%	2,88%
	evolution	2,16%	2,28%	1,92%	2,29%	2,20%
<b>Decision</b>	rejection	2,37%	3,72%	2,92%	2,70%	2,75%
	approval	8,37%	5,03%	6,16%	7,85%	7,43%
	sceptical	2,40%	2,40%	2,60%	2,34%	2,40%
	confirmation	1,47%	0,84%	0,64%	1,51%	1,29%
	change	0,57%	0,68%	0,50%	0,63%	0,60%
	bias	1,78%	0,92%	1,10%	1,69%	1,54%
<b>Type of argument</b>	detail	2,51%	2,92%	3,24%	2,43%	2,63%
	time	1,47%	0,84%	1,37%	1,27%	1,29%
	aesthetics	4,71%	6,63%	4,88%	5,38%	5,25%
	realism	6,14%	8,47%	6,75%	6,83%	6,80%
	personal	3,79%	3,92%	4,20%	3,71%	3,83%
	medium	5,32%	7,19%	6,30%	5,71%	5,85%
	purpose	2,42%	3,72%	3,65%	2,50%	2,79%
	physicality	3,30%	2,56%	3,06%	3,11%	3,09%
<b>Reflection</b>	self-awareness	2,61%	1,68%	2,19%	2,40%	2,34%
	contradiction	0,00%	0,04%	0,05%	0,00%	0,01%
	ambiguity	2,15%	2,24%	2,92%	1,93%	2,17%
	metadata	1,63%	2,68%	2,74%	1,66%	1,92%
	frustration	0,38%	0,36%	0,55%	0,32%	0,37%
<b>Issues</b>	non-answer	0,79%	1,12%	0,50%	1,01%	0,88%
	comprehension	0,69%	0,60%	0,59%	0,69%	0,67%
	invalid	0,05%	0,16%	0,00%	0,02%	0,08%
	technical	0,14%	0,28%	0,05%	0,17%	0,18%



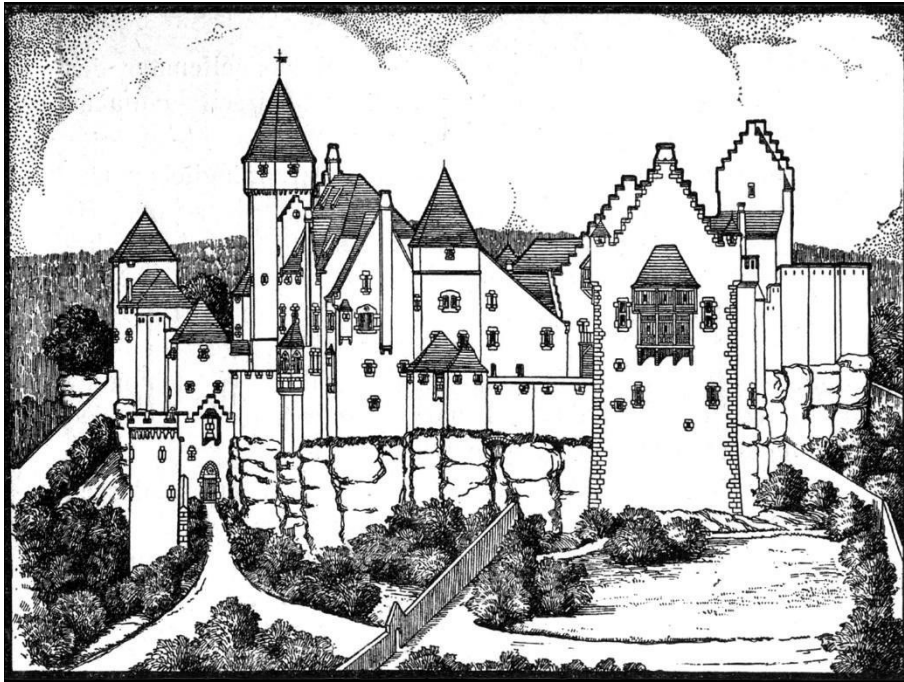
## Independent Application

The second section of the survey was designed to analyse whether respondents were capable of independently applying the reconstruction process they had been guided through in the first section. They were provided with a total of six historical images with a view of one of the castle's buildings, which shows the traces of another element that had been attached to its façade. Some guidance was provided in the form of highlights on the images to help respondents see the part under discussion, and leading questions to establish that details varied in each view and there were a number of possibilities for the reconstruction. However, they were no longer given a binary choice, but could choose from among several pre-defined options or provide their own answers. They were tasked with deciding what, if anything, had been attached to the wall in question, and which material it was made from.

The main analysis was performed not on their choices, but on the reasons they gave for those choices. Only eight out of all 360 respondents declined to engage by giving non-answers. 40.8% of replies showed that users had made their decision based on an interpretation of the data provided, while 38.9% argued using the physical properties of the materials in question, 19.2% used functional arguments, 24.7% showed reasoning relying on change over time, and 15.6% mentioned the remaining ruins.

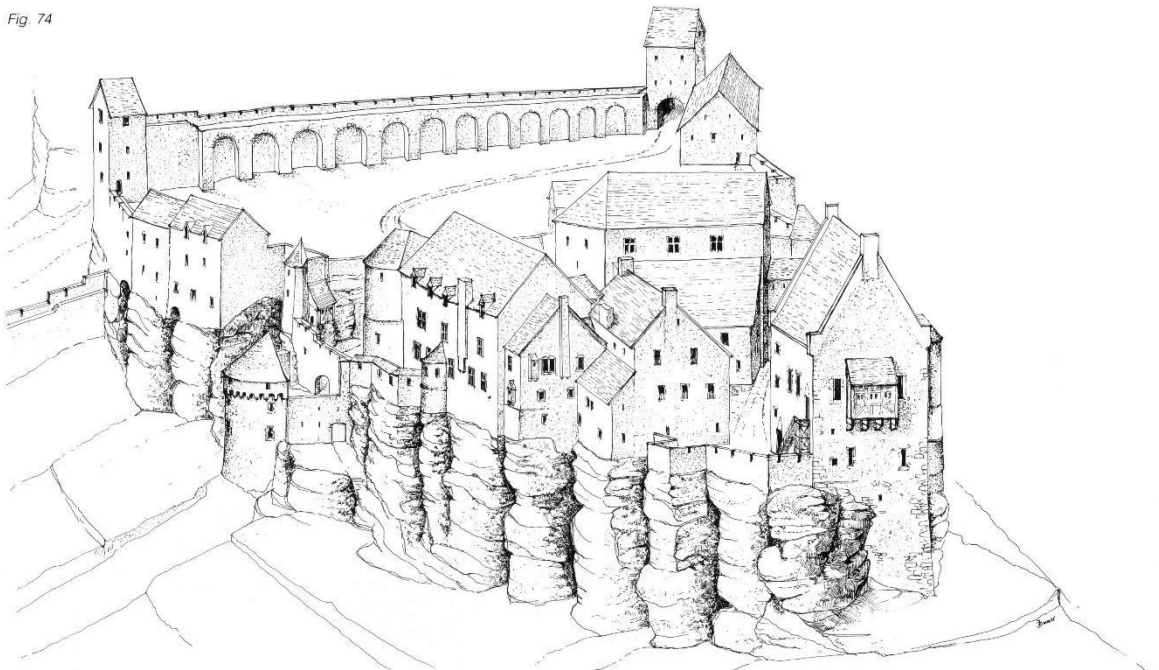
Since only 15% indicated that they were relying on previous knowledge and only 14.2% admitted that they guessed at the answer, these responses indicate significant engagement with the sources presented and a willingness to analyse and interpret them to arrive at an answer. Only one user expressed frustration and only three said they would need more metadata, showing that on the whole, they felt empowered to answer the question using the sources provided.

## What is the “best” reconstruction?



1 – JP Koenig's reconstruction

Fig. 74



2 – John Zimmer's reconstruction

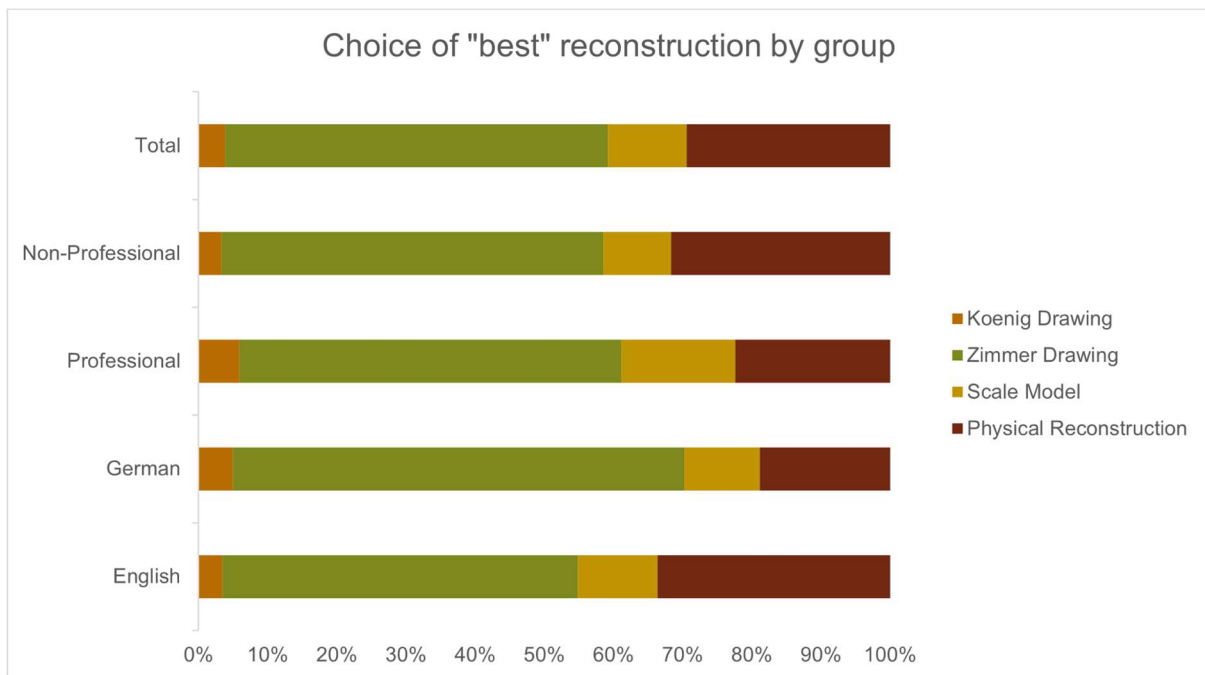


3 – The anonymous scale model



4- The physical reconstruction

The final set of questions in the analysis sought to determine how respondents applied the preceding guided reconstruction process to the analysis of four existing reconstructions. First, they were tasked with giving their first impressions and rating the perceived accuracy of each on a scale of 1-5, then asked to decide which reconstruction was “best”– a term left deliberately vague to see what respondents prioritised.



Across all groups, users clearly favoured John Zimmer’s illustration, which a combined total of 55.3% of users considered “best”. Second was the physical reconstruction at 29.4%, third the scale model at 11.4%, and the least favourite was JP Koenig’s reconstruction drawing at only 3.9%. This reconstruction has the least “scientific” aesthetic, drawn in a flat, illustrative style and embellished with fanciful details like stepped gables and numerous narrow spires. Close observers will also note that it fails to comply with the primary defensive function of a castle, as the hinges of the main gate are placed on the outside.

Therefore, it is astonishing that it was much more popular among professionals than non-professionals. While only 3.3% of non-professionals chose this representation, it was favoured by 5.9% of professional users, an increase of 69%.

While their enthusiasm for the Zimmer reconstruction corresponded to the total average, professionals were also much more sceptical of the physical reconstruction, which only 22.4% thought was “best”, as opposed to 31.6% of non-professionals.

For the analysis, the responses were encoded to compare the type of arguments used for which of the four reconstructions presented was the “best”. Percentages total over 100%, as multiple codes could be assigned to each response.

The analysis showed clear favourites in each category of responses. For knowledge, this was authority at 10.8% total. Surprisingly, professionals also favoured emotional arguments, at 10.6%, compared to 6.9% for non-professionals.

In the reasoning category, first place was tied between artefact and context, at 11.4% each, though German speakers also favoured intuition at 8.7%, while it only accounted for 2.7% of English responses.

Decision codes were rare for this question, but leaned toward bias, at 2.32% of English and a surprising 0% of German responses. Argument types were the most common codes for this question, with 1.5 codes assigned per response on average. These clearly favoured medium – how something was made – with 41.7% of respondents using this type of argument followed by realism at 30.6% and purpose at 24.2%. All three were especially favoured by professionals. Surprisingly, the argument of time – how old the reconstruction is – only accounted for 1.9% of responses and was not used at all by German speakers.

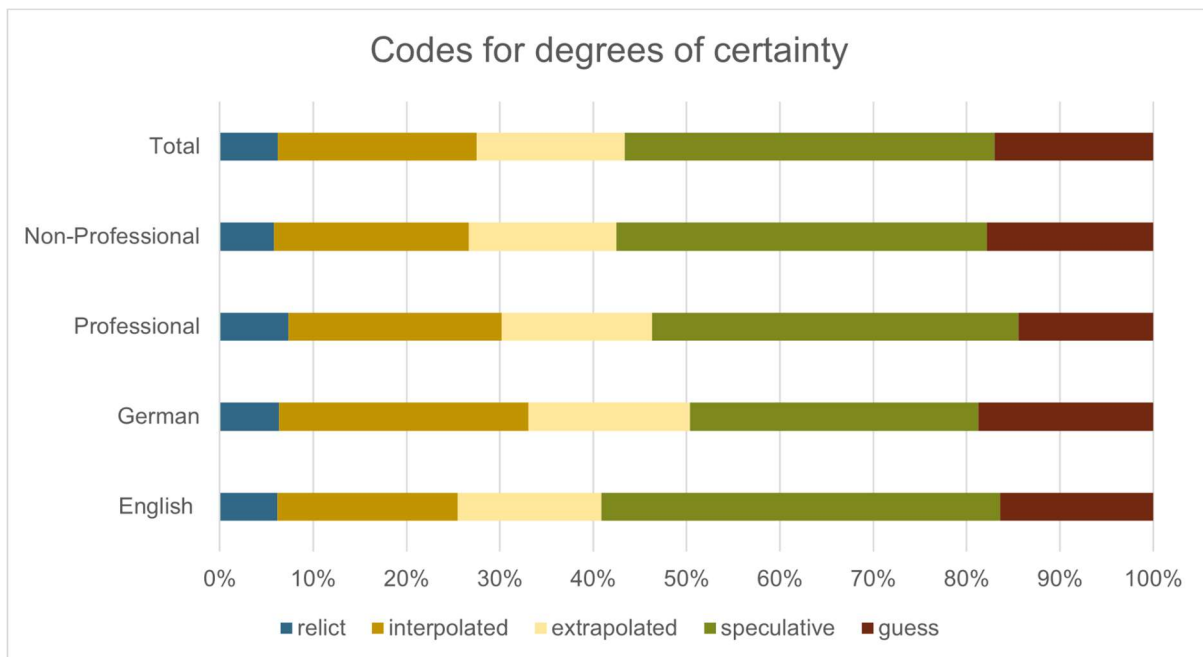
In the reflection category, 5.6% of responses pointed out the inherent ambiguity, with professionals (8.24%) 75% more likely to do so than non-professionals (4.7%). Almost as many (5%) wanted more metadata or showed self-awareness (4.7%), meaning they showed that they considered what influenced their decision-making process.



Explanation for best	Code	EN %	DE %	PRO %	NO %	Total %	
<b>Knowledge</b>	prior	3,09%	3,96%	2,35%	3,64%	3,33%	
	source	0,00%	0,00%	0,00%	0,00%	0,00%	
	authority	10,42%	11,88%	16,47%	9,09%	10,83%	
	consensus	2,70%	0,00%	0,00%	2,55%	1,94%	
	emotion	7,34%	8,91%	10,59%	6,91%	7,78%	
<b>Reasoning</b>	context	13,13%	6,93%	18,82%	9,09%	11,39%	
	function	1,93%	2,97%	2,35%	2,18%	2,22%	
	inference	0,00%	0,00%	0,00%	0,00%	0,00%	
	intuition	2,70%	8,91%	4,71%	4,36%	4,44%	
	interpretation	1,16%	1,98%	2,35%	1,09%	1,39%	
	artefact	11,97%	9,90%	9,41%	12,00%	11,39%	
	comparison	6,56%	4,95%	8,24%	5,45%	6,11%	
	evolution	2,70%	2,97%	3,53%	2,55%	2,78%	
	<b>Decision</b>	rejection	0,39%	0,99%	1,18%	0,36%	0,56%
		approval	1,93%	0,00%	1,18%	1,45%	1,39%
		sceptical	1,16%	0,99%	0,00%	1,45%	1,11%
confirmation		0,39%	0,00%	0,00%	0,36%	0,28%	
change		0,00%	0,00%	0,00%	0,00%	0,00%	
bias		2,32%	0,00%	1,18%	1,82%	1,67%	
<b>Type of argument</b>	detail	6,95%	4,95%	7,06%	6,18%	6,39%	
	time	2,70%	0,00%	1,18%	2,18%	1,94%	
	aesthetics	9,27%	8,91%	5,88%	10,18%	9,17%	
	realism	28,96%	34,65%	28,24%	31,27%	30,56%	
	personal	14,29%	25,74%	24,71%	15,27%	17,50%	
	medium	39,00%	48,51%	50,59%	38,91%	41,67%	
	purpose	18,53%	38,61%	28,24%	22,91%	24,17%	
	physicality	19,69%	8,91%	14,12%	17,45%	16,67%	
<b>Reflection</b>	self-awareness	5,79%	1,98%	3,53%	5,09%	4,72%	
	contradiction	0,00%	0,00%	0,00%	0,00%	0,00%	
	ambiguity	5,02%	6,93%	8,24%	4,73%	5,56%	
	metadata	3,86%	7,92%	4,71%	5,09%	5,00%	
	frustration	0,77%	0,99%	1,18%	0,73%	0,83%	
<b>Issues</b>	non-answer	1,54%	3,96%	0,00%	2,91%	2,22%	
	comprehension	0,00%	0,99%	0,00%	0,36%	0,28%	
	invalid	0,00%	0,00%	0,00%	0,00%	0,00%	
	technical	0,39%	0,99%	0,00%	0,73%	0,56%	

## Relict Interpolated Extrapolated Speculative

As described above, one goal of this study was to investigate the type of arguments respondents used in relation to the four categories of certainty: *relict* – *interpolated* – *extrapolated* – *speculative*, as this could indicate how receptive they might be to each type of argument or how salient those arguments seem to them. Percentages below are given out of the total of these six codes, not out of all codes assigned.



The most popular form of argument across all groups is clearly speculation – encoded as “function” or “inference”, this class covers deductions from the respondents’ own prior knowledge or understanding of the purpose, structure and logistics of a building. In part, this shows that users are attempting to fill in the obvious gaps in the sources with tertiary data; however, it could also indicate that they rely more on their own perception of history and architecture or urban planning than on the data itself.

This is underlined by the fact that the second most frequent were not arguments using the spatial or geometric context to fill in gaps in the data (extrapolation) but interpretation of the sources themselves (interpolation) – which indicates that the data themselves were not so meagre that no more information could be gleaned

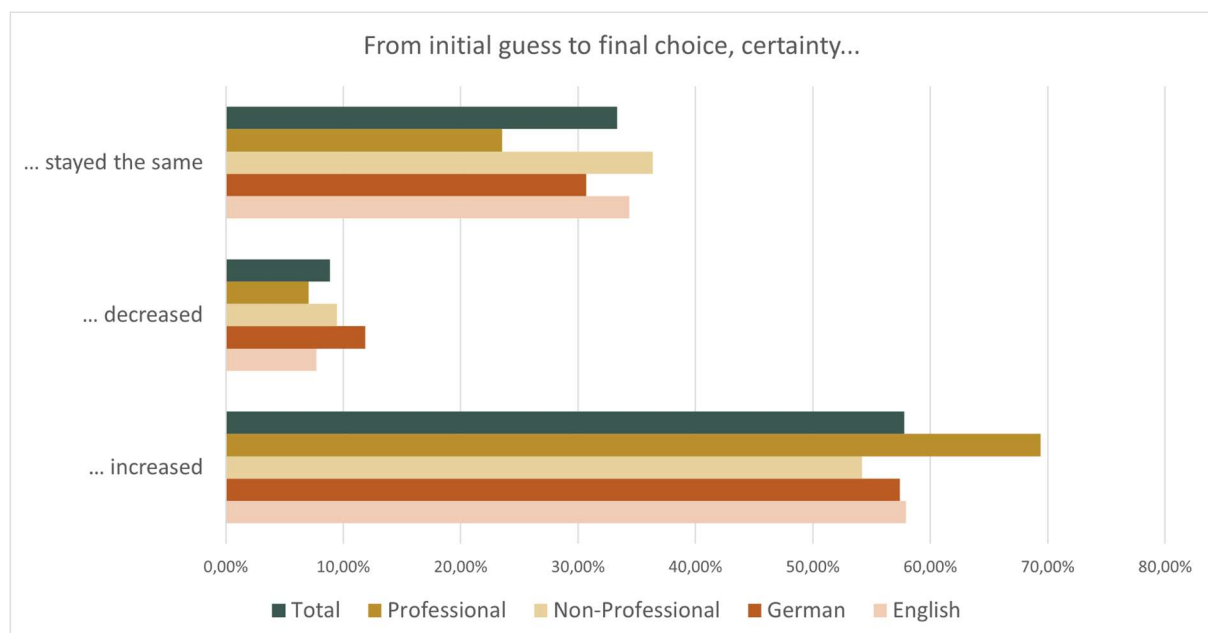
from them, but that users preferred to fall back on their own knowledge entirely rather than extrapolate from the data provided.

Arguments using relicts, or still remaining original elements, were the fewest but made up a large percentage of the arguments for the “best” reconstruction, showing that where such remains were provided as sources, they were well received.

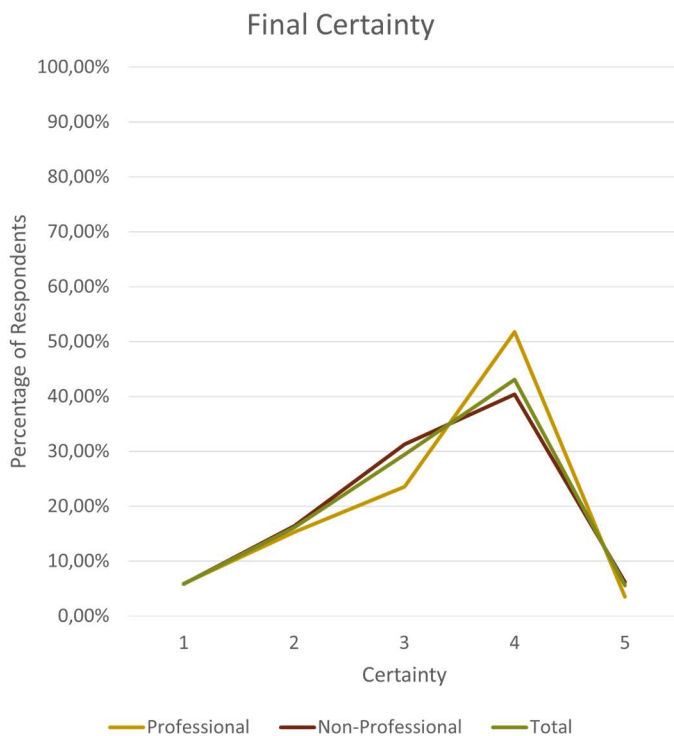
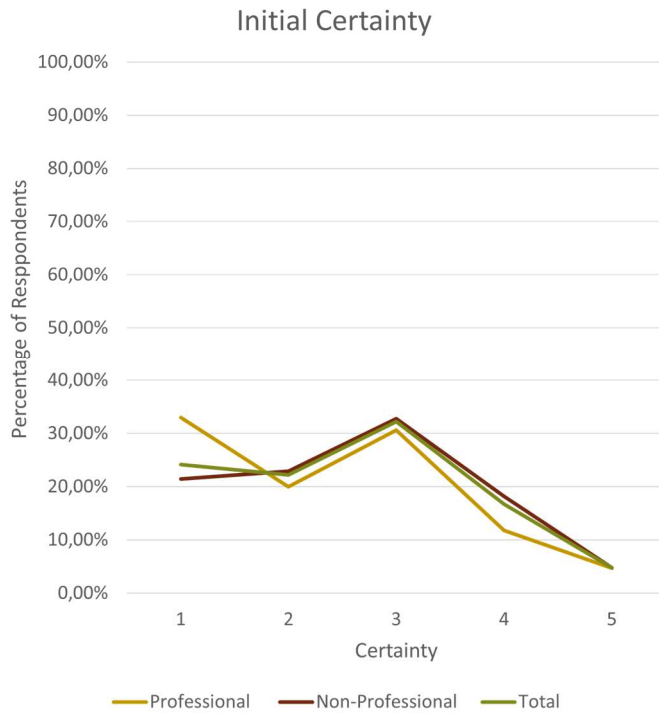
## Certainty and Accuracy

Finally, going through the questionnaire process had a clear effect on users’ certainty that they had made the correct choice in the end, with 57.8% showing an increase, 33.3% remaining the same and only 8.9% growing less certain.

Interestingly, over 70% of users indicated that they thought their new choice was more accurate than their initial guess, and only 10% thought the accuracy had stayed the same or was less accurate – which could show that they don’t equate accuracy and certainty, that they didn’t remember which certainty they had chosen earlier, or that a five-point scale was not sufficiently nuanced for their degree of change.



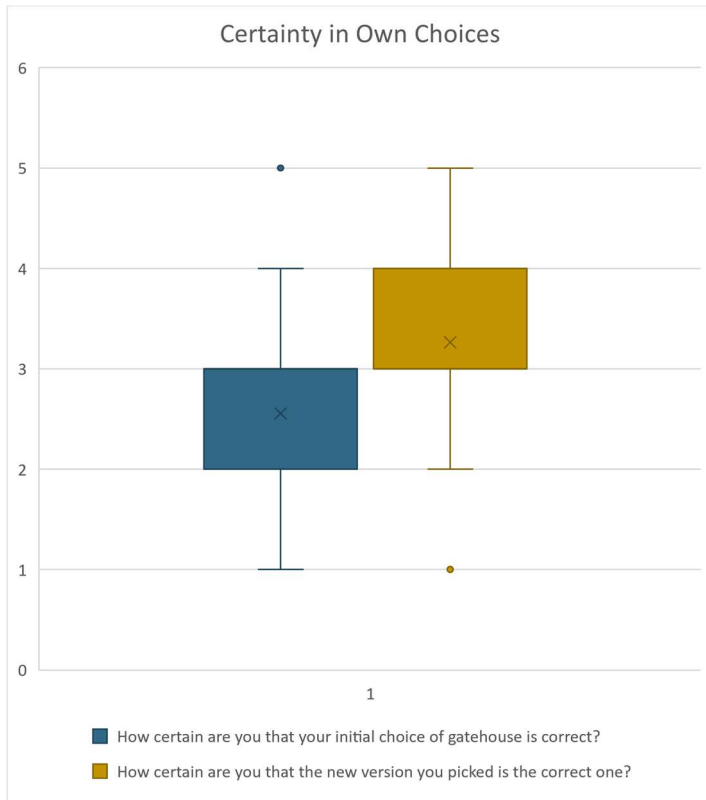




These changes are largely consistent across languages, though German speakers were 63% more likely to see a *decrease* in confidence. In general, confidence in their own choices rises by approximately 30%, with the biggest increase seen among professional respondents (41%) and the smallest among non-professional respondents (24%). This indicates that professionals are more aware of the importance of having metadata; their confidence starts out the lowest when they are just guessing, even though they have the best prerequisites for making a reasonable attempt, and rises to the highest after reviewing available data.

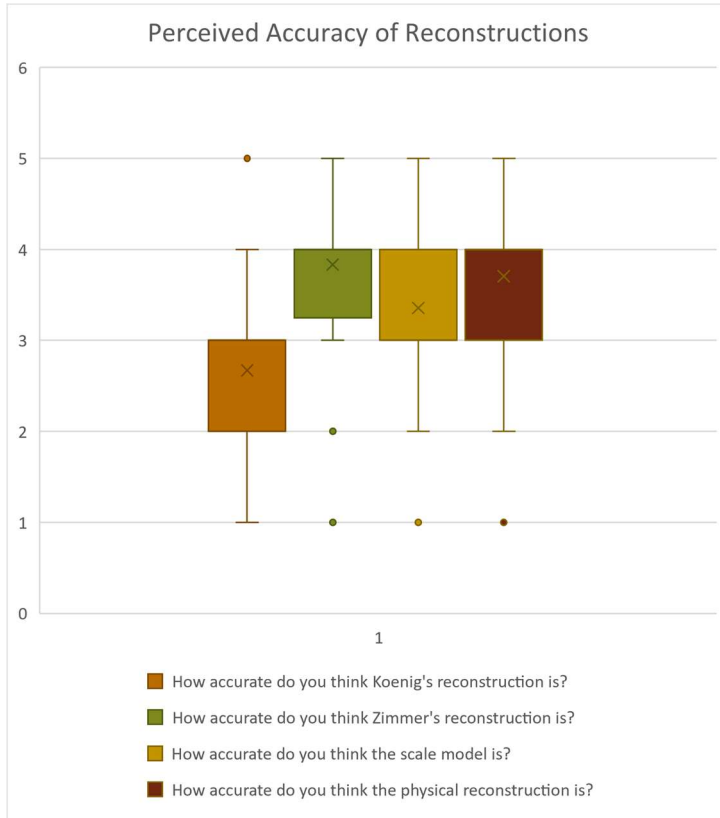
A corollary to the respondents' certainty that their own choices are correct is their perception of the accuracy of the reconstructions shown. In all cases, this perception is, on average, higher than their initial

confidence in their own guesses. For all but the Koenig reconstruction, it also exceeds their own confidence in their final choice of reconstruction, showing that they trust a presumed authority more than their own choices. Respondents have a similar level of confidence in the Koenig reconstruction, which is less technical and

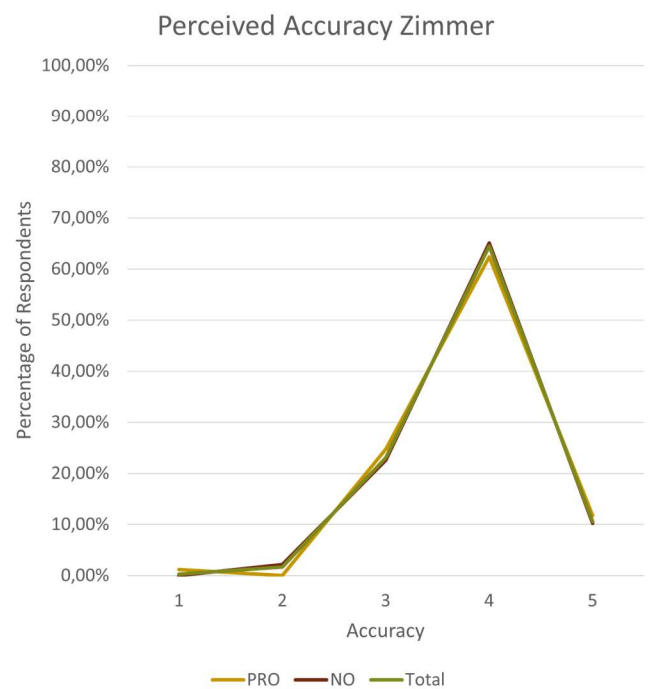
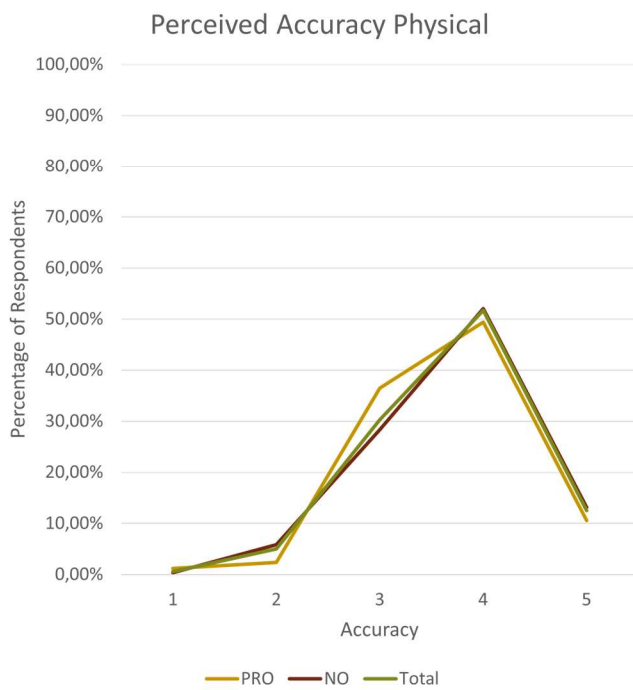
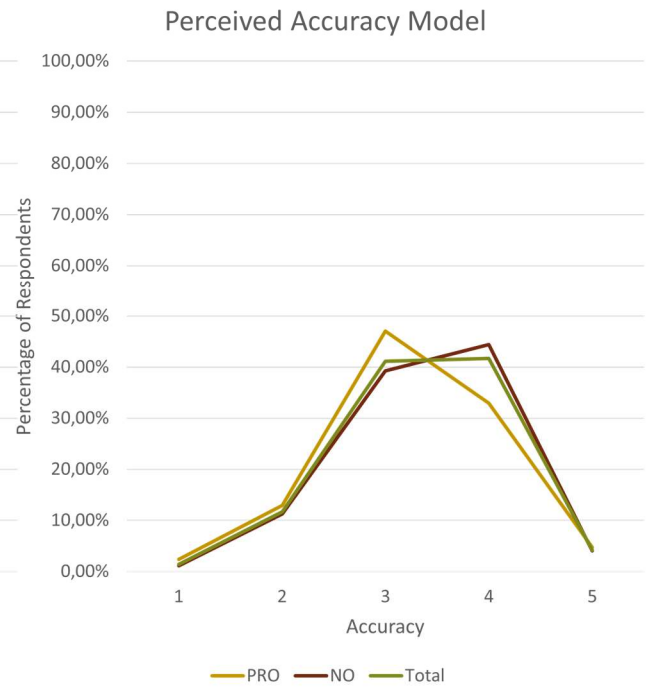
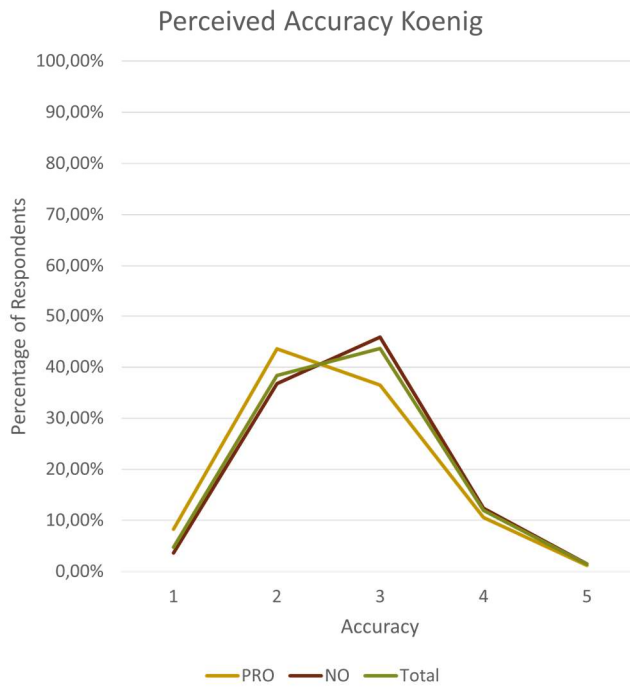


more artistic in its interpretation, as in their own initial guesses formed on the basis of very little data, perhaps implying that they believe his approach was intuitive rather than data-driven.

These perceived levels of accuracy also closely mirror the choices for which reconstruction is “best”, so many users may – consciously or not – have chosen accuracy as the deciding argument in favour of one over the other.

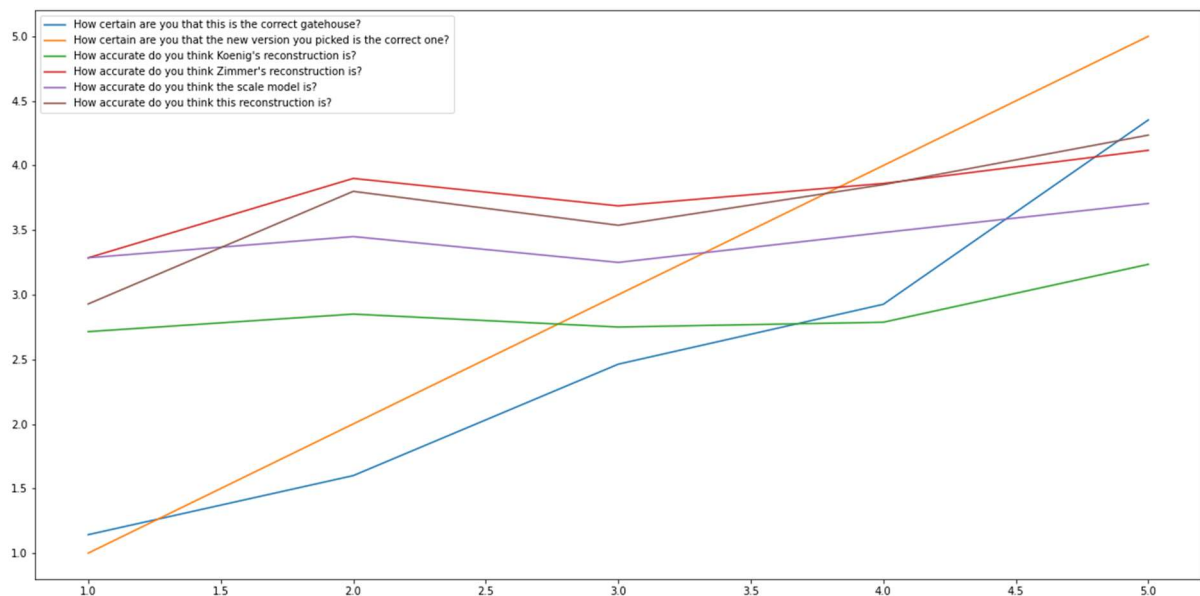


Of the four reconstructions provided, respondents had the highest confidence in John Zimmer’s illustration, averaging 3.8 on a scale of 1-5, closely followed by the physical reconstruction in situ at 3.7. The scale model of unknown origin reaches an average of 3.4, while JP Koenig’s drawing was the only reconstruction to average less than 3 at 2.7. These are similar across all groups.



To check for potential unconscious bias, the relationships between these confidence scores were also investigated, under the assumption that the values would show some correlation, i.e. that respondents might generally tend to favour higher or lower numbers. To visualise this relationship, all responses were grouped by their

confidence score for their final choice, then the other responses for those groups were averaged. This shows a correlation between the scores for their initial and final certainty, but only a very slight upward trend in the accuracy scores for the reconstructions, meaning that respondents generally made deliberate choices rather than just following a preference for high, low, or noncommittal scores.



At the opposite end of the “certainty” spectrum, users’ reaction to ambiguity was also investigated. This was consciously generated by showing conflicting sources and making decisions mandatory to progress in the survey. Of all users, 35% commented on this ambiguity at least once. How many noticed it otherwise or even left the survey unfinished because of their discomfort with it is, of course, not recorded.

Frustration with forced choices in the face of ambiguity could also correlate with the variations on “no comment” or other non-answers entered into required text fields. These fortunately only made up 0.9% of all codes, or 78 total replies. However, they were twice as likely for non-professionals (1% of codes) than for professionals (0.5%), likely because history and heritage professionals are more used to having to resolve conflicting data.

## Languages

The two different versions of the survey, English and German, were prepared by the same author, a native speaker of both languages, to minimise the influence of language or tone on the results.

Both were released across various social media channels at the same time, but seem to have travelled by different routes – while the German version was recommended to incoming architecture students by their lecturers and shared among heritage professionals and their acquaintances, many participants in the English version were apparently recruited when it was shared by re-enactment groups. However, these differences in demographics may not be enough to explain some interesting incidental observations.

The most striking of these is that the two groups had very different trends in their mistakes and mismatches – English speakers tended to swap left and right, while German speakers swapped rectangular and round. Overall, German speakers showed a lower rate of bias (19.8% vs. 33.2% of users showing bias at least once), but German speakers had 65% more final choices that confirmed their initial guess rather than being the logical result of their decisions (8.91% vs. 5.41%). English speakers, on the other hand, were almost 4x more likely (7.3% vs. 2%) to make decisions not adding up to any of the variants offered.

Finally, while all German-speaking respondents answered that they had answered the questions seriously and they could be used for research, over 6% of English speakers felt the need to qualify this statement and explain more about their background or reasoning.

Other interesting differences were found in their use of consensus as an argument, which German speakers employed only half as often, and did not use at all to justify what is “best”. They also appealed to the ideas of realism or authenticity much more often, despite having a higher proportion of professionals, who use it less than average.

## Discussion

As demonstrated above, one aspect highlighted by this study was the frequent lack of historical knowledge, or the prevalence of half-remembered facts mixed up with Fantasy. This was particularly noticeable in the answers to the question “what else could [the wooden oriel with a chimney and an internal window to the chapel] have been used for?”.

27% of respondents replied with some variation on “toilet” – which demonstrates a vague awareness of castle privies without understanding how they worked, considering the size, prominence, and location of the structure in question. Many other answers suggested that it was purpose-built for spying, lovers’ trysts, secret meetings, or hidden assassins or used as a fancy dungeon, all ideas that owe much to romanticised fantasy and little to actual life in a castle. (Sturtevant 2017)

This means that they are not starting as a blank slate, but rather, we are fighting misconceptions about history as much as providing a “correct” version. Presenting facts that oppose dearly held beliefs does not always convince people to change their minds; at worst, a “backfire effect” could be strengthening their bias and causing them to hold those beliefs even more strongly. (Lewandowsky et al, 2012) Therefore, users should be empowered and encouraged to “discover” those facts for themselves, making connections and being allowed to formulate their own theories within the framework provided, guiding them toward a more appropriate interpretation.

## Conclusions

The results of this study show that providing users with bundled, non-specific historical texts of the type often found alongside an exhibit is not enough to fulfil the London Charter’s call for allowing users to understand reconstructions. Instead, allowing them to go through the reconstruction process for themselves increases their confidence, and shows that they are aware of and interested in the data underlying such a reconstruction. For some, it was the first time even thinking about how such decisions are made; others, including built heritage professionals, were able to engage more deeply, but still found the exercise worthwhile.

The types of arguments they employ to justify their decisions, evaluate reconstructions, and explain what makes them “good” demonstrates that they prefer reasoning that gives them agency, from their own perspective concerning function and logistics than to examine the remaining ruins at the site and its surroundings – unfortunately, the latter are often the only data available to heritage site visitors.

While many respondents indicated that they were willing to listen to an established authority, the overwhelming majority preferred to be given metadata. Conversely, guide books tend to inform from a position of authority, presenting the “correct” solution rather than tracing the decision-making path that led to it being accepted as fact. (Noakes 2021) In order to increase understanding and engagement, these should be supplemented by further data.

While some users embraced the ambiguity of the “correctness” of their choices or the interpretations suggested by professionals, others were frustrated, requesting options to answer “I don’t know” or pointing out that it was impossible to be sure. This discomfort is valuable, as it illustrates how a scientific theory works: it is the best possible interpretation of the known facts, but is subject to change in the face of new data. Making this explicit is rare in the humanities<sup>3</sup> – a missed opportunity to show

---

<sup>3</sup> In fact, an older meta-study suggests that the decline in quantitative methods in published journal articles in the field of history observed from the 1970s to the 1990s is due to their being treated hypercritically by reviewers, a trend the author expected to increase. (Reynolds 1998, p 146)

that historical knowledge creation happens through a similar process, and the results are not set in stone.

However, the survey also demonstrates that the average user requires guidance – overwhelming them with uncurated data would lead to more frustration and confusion, as it was often difficult for them to do so even with this carefully chosen, progressive selection.

The high rate of mismatches between logical sum of decisions and final choice shows that future surveys of this type require fine-tuning through an iterative design process that explores whether users could not remember their decisions, decided to ignore their conclusions in favour of an intuitive choice, or did not understand the task at hand.

This is an important data point in itself – it is more difficult for the average user to successfully complete this guided reconstruction process than initially assumed. Any such material used as part of a museum exhibit or educational programme must therefore be constructed so that any frustration experienced by the user is at the inherent ambiguity of making decisions using incomplete data – good, a learning experience – and not because of comprehension difficulties or technical issues.



## Future Work

This study is an initial attempt at investigating the problem of how laypeople can be guided toward engaging meaningfully with a reconstruction and understanding its methodology, limits, and potential. Future work should focus on understanding where the barriers to understanding lie, in a larger team incorporating further disciplines. For example, users could be guided through a similar process under observation by psychologists, to determine whether the decisions they are self-reporting matches the actual arguments that can make them change their minds.

The introduction of a control group could also help determine the impact of guiding users through the process, with half completing an exercise related to demonstrating understanding of a 3D model and its data and metadata before being shown how the decision-making behind reconstructions works, the other half after.

Other stakeholders that should be involved include representatives of the site in question, curators or other experts who can identify the most significant or data-rich aspects for inclusion, (museum) educators to achieve the correct tone and scope, and finally, the researchers and graphic artists making the reconstructions, to tie the different aspects together and include recurring themes for users to discover.

Fully incorporating an exercise such as this into the visitor exhibition of a heritage site could be used to monitor engagement – how many visitors are interested – as well as increasing their understanding.

## Bibliography

Denard, Hugh, *The London Charter for the Computer-Based Visualisation of Cultural Heritage*, v 2.1, King's College London, 2009 <http://www.londoncharter.org>

Stephan Lewandowsky et al., 'Misinformation and Its Correction: Continued Influence and Successful Debiasing', *Psychological Science in the Public Interest* 13, no. 3 (1 December 2012): 106–31, <https://doi.org/10.1177/1529100612451018>.

ICOMOS, "The Venice Charter of 1964", 2nd International Congress of Architects and Technicians of Historic Monuments, Venice, 1964

Jiménez, Juan, 'Videogames and the Middle Ages', *Imago Temporis: Medium Aevum*; Núm.: 3, 1 January 2010.

de Kramer, Marleen, 'Relict-Interpolated-Extrapolated-Speculative: An Approach to Degrees of Accuracy in Virtual Heritage Reconstruction.', in *Visual Computing for Cultural Heritage*, 2020, pp. 409–22 [https://doi.org/10.1007/978-3-030-37191-3\\_21](https://doi.org/10.1007/978-3-030-37191-3_21).

Kerrigan, Susan and McIntyre, Phillip, 'The Creative Treatment of Actuality: Rationalizing and Reconceptualizing the Notion of Creativity for Documentary Practice', *Journal of Media Practice*, 11 (2010), 111–30  
[https://doi.org/10.1386/jmpr.11.2.111\\_1](https://doi.org/10.1386/jmpr.11.2.111_1).

Noakes, Louise 'Cardigan Castle: How to Create a Guidebook Your Visitors Will Want to Buy', *Museums + Heritage Advisor*, accessed 26 August 2021, <https://advisor.museumsandheritage.com/features/cardigan-castle-create-guidebook-visitors-will-want-buy/>.

Putnam, James: *Art and Artifact: The Museum as Medium*. London: Thames & Hudson, 2001

Reynolds, John F. 'Do Historians Count Anymore?: The Status of Quantitative Methods in History, 1975–1995', *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 31, no. 4 (January 1998): 141–48, <https://doi.org/10.1080/01615449809601196>.

Spring, Dawn 'Gaming History: Computer and Video Games as Historical Scholarship', *Rethinking History* 19, no. 2 (3 April 2015): 207–21, <https://doi.org/10.1080/13642529.2014.973714>.

Sturtevant, Paul, 'Medievalisms of the Mind: Undergraduate Perceptions of the "Medieval" and the "Middle Ages".' In K. Fugelso (Ed.), *Studies in Medievalism XXVI: Ecomedievalism* (pp. 213-236). Boydell & Brewer, 2017

Sturtevant, Paul "'You Don't Learn It Deliberately, But You Just Know It from What You've Seen": British Understandings of the Medieval Past Gleaned from Disney's Fairy Tales', in *The Disney Middle Ages*, ed. Tison Pugh and Susan Aronstein (New York: Palgrave Macmillan US, 2012), 77–96, [https://doi.org/10.1057/9781137066923\\_5](https://doi.org/10.1057/9781137066923_5).

Weschler, Lawrence. *Mr. Wilson's Cabinet Of Wonder: Pronged Ants, Horned Humans, Mice on Toast, and Other Marvels of Jurassic Technology*. New York: Random House, 1996

## Acknowledgements

I would like to thank the Centre for Contemporary and Digital History (C<sup>2</sup>DH), the Institute for History (IHIST), and the Luxembourg National Research Fund (FNR) (10929115).

Thanks also to my colleague Christopher Morse for his assistance with data encoding and to Jannika Lossner for data analysis support, and to Maria Christini and Francesca Doddato as well as all other reviewers for their useful comments.