

Evaluating Immersive User Experience and Audience Impact

A report produced by Nesta and i2 Media Research
for Digital Catapult, June 2018

CONTENTS

1	Executive summary
5	Introduction and research context
6	Methodology
13	Background, literature review and focus groups
21	Research findings
36	Key predictors of impact in immersive experiences
38	Toolkit development
41	Conclusion
43	Appendices
44	References & Footnotes

REPORT AUTHORS

Dr Jane Lessiter
Sam Mitchell
Eva Ferrari
Polly Borden
Hasan Bakhshi
Professor Jonathan Freeman

This report was produced by Nesta and i2 Media Research for Digital Catapult, and funded by Innovate UK.

Digital Catapult, Immerse UK and the High Value Manufacturing Catapult have been working together on a large-scale programme of business support, funded by Innovate UK, for the UK's immersive technology industries since September 2017.

This report forms part of that work along with the following complimentary reports:

- [The Immersive Economy in the UK](#)
(Innovate UK, Immerse UK and Nesta)
- [Growing Your VR/AR Business in the UK: A business and legal handbook](#)
(Digital Catapult and PwC)
- [Immersive Content Formats for Future Audiences](#)
(Digital Catapult & Limina Immersive)
- [Creative Tools and Workflows for Immersive Content Creation](#)
(Digital Catapult, Opposable Group and TechSpark)
- [Immersive in manufacturing - the adoption and use of immersive technologies in manufacturing and a report covering the feasibility of the use of immersion in a digital twin](#)
(High Value Manufacturing Catapult)

Innovate UK is part of UK Research and Innovation, a non-departmental public body funded by a grant-in-aid from the UK government. We drive productivity and economic growth by supporting businesses to develop and realise the potential of new ideas, including those from the UK's world-class research base.

Contact us

immersive@digicatapult.org.uk
www.digicatapult.org.uk

Working with
Innovate UK

FOREWORD

From fantastical virtual worlds, to augmented design tools and immersive training environments; VR and AR technology is opening doors to new opportunities every day.

Innovate UK's recent Knowledge Transfer Network report on The Immersive Economy in the UK estimates that Britain has around 1,000 immersive-specialist companies employing around 4,500 people and potentially representing as much as 9% of global market share. UK Creative Industries have a huge amount to contribute to this emerging immersive sector, not least because many of the skills involved are derived from different corners of this thriving, diverse and crucial part of the economy (such as film, TV, games, visual effects, etc).

But the challenges remain. As an early stage market with a varied set of creatives, technologists and researchers driving its development, we lack a common language to describe the way we create, define, refine and value immersive content.

Digital Catapult has commissioned a set of three reports from industry-leading companies to help demystify some of the common questions around the creation of immersive content. While there is a focus on the creative industries, much of this insight will extend across industry boundaries to other sectors implementing and experimenting with a broad and exciting range of immersive applications.

This report on Evaluating Immersive User Experience and Audience Impact, conducted by Nesta and i2 Media Research, looks at the challenges of understanding the value of creative content while the consumer market for VR and AR content is still small, and traditional quantitative measures of measuring impact are not always available or reliable. The report develops a research methodology for testing and evaluating the experimental immersive content that is being made now, in a way that will help us predict the potential audience appetite, cultural impact, and commercial opportunity in the future.

By sharing the insights from these reports, Digital Catapult hopes to consolidate key industry insights and help lower the barrier to entry to this exciting and rapidly growing market. The diversity of entrepreneurs, technologists, educators, developers and content makers working in this space is one of its greatest strengths, which is why we believe the UK will become the best place in the world to create immersive content and applications.

Jeremy Silver

CEO, Digital Catapult

Rebecca Gregory-Clarke

Lead Technologist - Immersive, Digital Catapult

EXECUTIVE SUMMARY

How can we test and evaluate experimental immersive content that is being made, in a way that will help us predict the future potential audience appetite, cultural impact, and commercial opportunity?

The UK's immersive economy, although still emerging, looks set to grow rapidly over the next few years. Creative producers, whether from the immersive specialist sector or with arts, media and film backgrounds, are pushing the boundaries of virtual reality (VR), producing an ever-expanding range of content. Within this space, Nesta and i2 Media Research explored the following research question:

“...how can we test and evaluate experimental immersive content that is being made now, in a way that will help us predict the potential audience appetite, cultural impact, and commercial opportunity in the future.”

To answer this question, Nesta and i2 Media Research developed a research protocol that was applied during a controlled lab trial with immersive experience users. This trial gave users two pieces of immersive content to experience, and asked them to complete pre-and post-test surveys to gauge their psychological, cultural and economic responses. The research also gathered data from the content creators themselves, and conducted follow-up surveys with the trial participants to elicit their longer-term reactions. The research protocol was tested with three pieces of 'room scale' VR content that allowed six degrees of freedom in relation to movement, and with varying levels of interactivity:

1. Content A: A narrative-driven, short fantasy story in VR.
2. Content B: A short animated piece that allows the user to interact with a range of situations that a single character guides them through, to showcase some of the functionality of VR.
3. Content C: A short, non-narrative conceptual VR experience which emphasises human relationships with nature.

In total, 84 trial participants experienced two pieces of content, creating a total of 168 trial returns to be analysed. This method drew on a significant body of academic and industry research into consumer psychology, cultural value and willingness to pay. It was also validated as an approach to assessing impact of VR through three focus groups with content creators, industry stakeholders and (potential) audiences.

RESEARCH FINDINGS

Sample description

- The trial participants were generally younger and more ethnically diverse than the whole UK population, were highly culturally engaged but were less familiar with virtual reality. Specifically, only a very small proportion owned VR headsets in their home; approximately one-third had no previous VR experience, whilst 42% had experienced VR once or twice, and 18% had had three or more previous VR experiences.

Key descriptive findings of individual content

- The survey findings suggest that the research protocol was sensitive to different content types. That is, the overall experience of Content C was better than Content A which, in turn, was better than Content B.
- Nonetheless, the three content pieces did share some similar global experiential and cultural value attributes. For instance, all had high scores on qualities such as whether the user found the content Memorable, Transporting, Good and Impressive, and Different and Shareable.

- When examining the ratings by individual content in more depth, factors which may have impacted the overall ratings were identified. In particular, both Content A and B left users feeling Self-Conscious, whereas Content C did not. Glitchy and Gimmicky use of technology may have negatively influenced Content B's ratings impact, whilst Content C offered a more Unusual Perceptual Experience which may have appealed to users.

Economic value of immersive experiences

- Participants in the user trials were given two hypothetical scenarios to test their willingness to pay for VR experiences. The first was a 'home-based' scenario in which they could enjoy the content alone via a headset in their home. The second scenario was to experience content in an 'arcade', in which they would go to a VR-specific entertainment centre and have the option to experience a number of different pieces of content.
- Over half of participants (58%) indicated they would be willing to pay to have immersive experiences at home, and 70% indicated they would pay to take part in the arcade scenario.
- The optimal price point for all three contents on average was £9 for the home scenario and £17 for the arcade scenario. There was some variation in the optimal pricing across contents. Of the 3 contents tested, Content A commanded highest average prices in home and arcade scenario.

Memorability of immersive experiences

- A follow-up survey with users was conducted 2-3 weeks after the research trial had taken place. The survey revealed that almost every participant was able to recall accurate details of their experience and all but two reported on them in the order they had experienced them.
- We infer that Content B was less memorable for the two participants who prioritised recall of the other content they had experienced. Consistent with this finding, Contents A and

C tended to be discussed by participants with others in a positive light, relative to Content B, for whom almost a quarter reported having discussed them in a negative light.

Comparisons between content creator intentions and user experiences (UX) of content

- The content creators were each asked to complete a short survey stating their intentions with regards to the emotional and psychological impacts on users of their experiences, to examine how closely this aligned with actual user experience.
- There was very strong general correspondence between the direction of participant experiential ratings and the content creators' intentions. Unlike for Contents A and B, the creators of Content C had a more nuanced emotional intention with some emotions and moods expected to diminish and some expected to heighten. The results show that these dual intentions were indeed experienced by users.

Key predictors of impact in immersive experiences

- We examined which of the psychological indicators within the research protocol (e.g. affect and presence measures) drove audience impact as defined by the global experiential and cultural value ratings. For all the contents combined, we found that Positive Affect (a sub-scale computed from ten mood state items such as Interested, Excited, Strong), and Engagement most consistently predicted variation in global impact and value ratings across all content and trials.
- When the contents were considered separately, each had slightly different predictors of global experiential impact: for Content A, Positive Affect most consistently predicted impact; for Content B, it was Engagement (consistently) and Sense of Presence (less consistently); and for Content C, Unusual Perceptual Experiences was the strongest predictor.

Toolkit development

- Derived from the full research protocol developed from this study, the research team has also created a toolkit which draws on the most useful measures tested in the survey. This toolkit can be utilised by content creators, immersive sector stakeholders or other researchers.
- The toolkit can be used formatively to help guide the development of impactful experiences. It can also be used to build a bank of evaluations against which any piece of content in the future can be benchmarked.

CONCLUSIONS

This research project seeks to test an experimental approach to assessing impact with regards to immersive experiences. We find evidence that the different components of the research protocol elicited nuanced responses from the user group, consistent with the variety of contents considered. The research also uncovers that, for these contents specifically, there is some consistency around the predictors of impact. A toolkit is being made available for other content creators and researchers to use.



Oscar nominated Pearl, funded by Google Spotlight Stories



'Draw Me Close'

INTRODUCTION AND RESEARCH CONTEXT

The UK's immersive economy, although still relatively new, looks set to grow rapidly over the next few years.

Research by Nesta for the Knowledge Transfer Network (KTN) and Immerse UK shows that there are currently around 1,000 immersive specialist companies working in the UK, with an estimated turnover of £660 million. Government and industry are increasingly emphasising the need for investment in this area, so that UK industry is best placed to secure an early international competitive advantage.

The immersive market for consumers is yet to consolidate through settled distribution channels and specific models for where and how immersive experiences will be consumed. YouGov, for example, estimates that only 6% of UK consumers own any form of VR headwear, including low budget options such as Google Cardboard. Although the assumptions and market indicators are that consumer VR (in particular) and immersive technology more generally will penetrate into home usage, there may be other widespread applications and use cases too – in VR notably, through cinemas, 'arcades' and one-off installations and experiences.

Creative producers, whether from the nascent immersive specialist sector or with arts, media and film backgrounds, are rapidly developing VR as a medium, and there are already a number of notable pieces of cultural content that have reached wide audiences. For example, in 2017, the National Theatre partnered with the National Film Board of Canada for a VR story entitled Draw Me Close which received widespread critical acclaim. In the same year, Google Spotlight Stories funded Pearl, an animation that became the first VR story to be nominated for an Oscar. There are now several festivals and showcases dedicated to promoting VR as an art form in a similar manner to film or music.

The context, then, is of a growing supply of VR and immersive storytelling content in the UK (and globally) and an as yet undefined market for how this content will be consumed. Nesta and i2 Media Research were commissioned by Digital Catapult to explore how immersive experiences impact upon audiences and propose a toolkit to evaluate audience reactions to this content. In particular, the focus was on uncovering:

“...how can we test and evaluate experimental immersive content that is being made now, in a way that will help us predict the potential audience appetite, cultural impact, and commercial opportunity in the future.”

The study investigates a set of questions that support this core research question through the development of a user experience trial with three pieces of VR 'story-based' content. A further goal of the study is to take the results from these three user trials and isolate the key variables that identify audience 'impact' (in its various forms) to create a short, replicable UX testing approach, or toolkit.

This report is structured in the following manner. Section 3 provides details of the methodology used in the UX trials. Section 4 outlines the background to the study, summarises the relevant literature and highlights the salient points from a series of focus groups and semi-structured interviews conducted at the outset of the project. Section 5 looks in detail at the research findings from the three user trials. Section 6 isolates the 'key predictors' of impact of the immersive experiences, and Section 7 moves on to discuss how they might feature in the UX evaluation toolkit. Finally, Section 8 draws conclusions and makes recommendations for how the findings of the study might be harnessed through the toolkit.

METHODOLOGY

The methodology used in this study centres on a lab-based UX trial with three pieces of content to elicit the different types of impact felt by individuals from immersive content. The approach combines insights from consumer psychology and cultural economics.

RESEARCH QUESTIONS

The project's main research questions are:

- How can we identify and measure suitable indicators of appreciation and impact on audiences, through qualitative or quantitative means, for example:
- How memorable is the content? (does it have a lasting impact?)
- How much does it drive social activity (do users wish to share the experience with others?)
- How repeatable it is for users (would somebody go back for more and why?)
- How might we assess the cultural value of the content?
- How can we use the above indicators to help assess the potential for future economic value of the content? (For example, this may include considerations such as whether it provides audiences with something they don't already have through other forms of media, or could it replace something they already have? i.e. does it represent time taken away from other activities).

RESEARCH APPROACH

The research comprised of five stages:

1. A review of the literature on measuring audience impact with regards to immersive experiences, along with the literature on approaches to measuring cultural experience and value more broadly. Within this, the conceptual space for the study was also outlined.

2. Focus groups and semi-structured interviews.

The research team convened three focus groups to discuss concepts of impact with regards to audience experience. These three focus groups were made up of content studios and creators, immersive economy stakeholders (such as industry bodies) and potential audiences respectively. The semi-structured interviews went into more detail on audience experience with content studios and stakeholders specifically (see Appendix B for more details).

3. The development of a research protocol, based on previous research by Nesta and i2 Media Research, along with specific questions that emerged from the literature review and interviews/focus groups.

4. A lab trial using the research protocol. The research protocol was tested on users at i2 Media Research's lab at Goldsmiths, University of London. Each participant experienced two out of three individual pieces of content. This is described in more detail below.

5. The analysis of data gathered through the research protocol. This is reported on in detail in Section 5.

Focus on VR within the trials

The project was developed to examine and evaluate 'immersive experiences', which includes both virtual and augmented reality (VR and AR) experiences. Due to the lab trial format of the study, we do not consider AR content in this paper. However, there are parts of the research protocol that directly relate to AR experiences (either on

mobile devices or otherwise) and that would relate to evaluating impact from people using augmented technology, rather than within virtual environments. Further research should look to examine how the research protocol and final toolkit differentiates AR experiences, and whether the approach is valid in those environments.

Content chosen for the UX trials

Digital Catapult provided three pieces of VR content for the trials: all broadly centred on artistic and storytelling VR content. All three were chosen to represent examples of emerging, lesser known 'formats' of content with less well understood audiences. VR contents featuring more classic 'video game' formats were not in scope for this trial; however, the toolkit we propose can be applied to immersive gaming experiences as well, and existing research on the impact of video games forms a key part of the background research for the present paper. Notwithstanding this, it should be noted that we do not hold a formal definition of VR games for this study, and in the case of one of the pieces of content, there were elements of the experience that bordered on gaming.

The three pieces of content chosen had all received some form of critical acclaim, were all under fifteen minutes in length and all were operational on the HTC Vive platform. Each piece of content was room scale, allowing six degrees of freedom within the virtual environment. All three of the content pieces also allowed for some form of interactivity, although this was varied in type, as discussed below. In each instance, the content studio was assured anonymity in participating in the trial, and so in this paper we produce a brief generic description of each piece (referred to as Content A, B and C from this point on) below.

1. Content A: A narrative-driven, short fantasy story in VR. While too structured in form to be considered a game, there are elements of gameplay within this experience and a number of points at which the user is able to make a specific choice to influence the story (not the case in either Content A or C).

2. Content B: A short animated piece that allows the user to interact with a range of situations that a single character guides you through, to showcase some of the functionality of VR. Users travel through a variety of different landscapes and conduct a few simple tasks (e.g. picking up items in VR). While there is interactivity in Contents A and B, there is no jeopardy or ability to 'lose' as in a conventional video game.
3. Content C [Treehugger]: A short, non-narrative conceptual VR experience which emphasises human relationships with nature. The user is able to interact with their surroundings, but in a non-linear format (for example, there is no plot exposition or 'levels'). In the full version of this piece of content, the experience is augmented with a physical, haptic experience: the trial only allowed the users to engage with the VR component of the full experience.

Research protocol

The research protocol for trial participants was made up of two elements. Firstly, trial participants were asked to fill out a short, pre-trial survey to capture a range of information, including socio-demographic data. Following the experience of the first piece of VR content, they completed a post-trial survey. They then experienced a second piece of content and completed another identical post-trial survey.

Following the trial, participants were invited by email to fill in a short follow-up survey between 2-3 weeks after they completed the trial, to try to understand the memorability of the content.

In order to track how well the intentions of the content creators were aligned with users, we asked the three content studios to complete a short questionnaire asking qualitative and quantitative questions about the intended impacts generated by their content. These were then analysed in relation to the same questions asked of the trial participants.

Table 1. Data gathered for user trials

DATA CATEGORY	DESCRIPTION OF DATA	DESCRIPTION OF WHAT THE DATA IS ATTEMPTING TO CAPTURE
PRE-test measures		
Socio-demographic data	Age group, Gender, UK/Region residence, Children in HH, Marital Status, Education, Work Status, Ethnicity, General Health, Household income.	Socio-demographic information on trial participants.
Interests in, and attitudes toward, arts, culture and technology	Arts and Cultural Event Visits, 10 items (e.g. a museum or gallery; some items derived from the DCMS Taking Part Survey). Lifestyle Engagement, 13 items (e.g. Eat out, Sports/ Gym/Exercise). Generating/Performing Arts and Culture Content, 14 items (e.g. written music, performed play). Attitudes to technology and social media, 9 items (some items derived from i2-Technology Experiences questionnaire, measuring 4 sub-scales: Tech Gadget; Positive Image as Technology User, TechnoAngst and Traditionalism). Previous VR experiencesexperience, 1 item. Previous Research Trial experience, 1 item.	Cultural attendance and participation, general lifestyle and attitudes towards technology and VR. With larger sample sizes in the future, audience experience can be segmented by (for example) familiarity with technology or levels of cultural engagement.
Adoption of technology	Digital technology ownership, including any VR headsets and type (8 products, e.g. Games Consolegames console). Use of Digital Services (4 services including streaming services and MMOGs).	Whether trial participants are early or late adopters of technology, and VR hardware and software specifically.

POST-test measures - General experiential and cultural value qualities (immediate reflections post test)

Global experiential quality evaluation	11 items generated by i2 to evaluate overall quality of experience, to reflect perceived sum of detailed experiential experience e.g. Good, Powerful, Worth Paying For.	Overall sense of UX. Combined, these responses provide one of the proxies for 'impact'.
Perceptions of cultural value	14 items derived from two measures: Arts Council England's Quality Metrics and Nesta's previous research with NT Live e.g. Interesting Idea, Well Presented and Produced; Engaged me on an Intellectual Level.	The underlying 'cultural value' of the experience. Combined, these responses serve as another proxy for 'impact'.

Table 1. Data gathered for user trials cont.

DATA CATEGORY	DESCRIPTION OF DATA	DESCRIPTION OF WHAT THE DATA IS ATTEMPTING TO CAPTURE
Intensity of affects/emotions	<p>29 emotions and affective (mood) states experienced during the content, items derived from:</p> <p>Positive and Negative Affect Schedule (PANAS, e.g. Excited, Watson et al., 1988, 20 items).</p> <p>Discrete Emotions Questionnaire (DEQ, e.g. Disgust, Harmon-Jones et al. 2016, 6 items),</p> <p>Profile of Mood States. (POMS, e.g. Confusion, McNair et al. 1971, 2 items).</p> <p>One additional state not covered in the above scales: 'Surprise'.</p>	<p>The emotional quality and intensity of the experience and the overall mood (affect) that people are left with. Many content creators intend from the outset to elicit particular emotions and affects/moods. We draw on a range of widely used published scales.</p>
Presence related experiences	<p>9 items derived from the ITC-SOPI (Lessiter et al., 2001) measuring 4 subscales:</p> <ul style="list-style-type: none"> • Sense of Physical Space (4 items). • Engagement (3 items). • Ecological Validity (3 items). • Negative Effects (5 items e.g. headache, eyestrain). 	<p>Users' sense of 'being there' (presence), which is typically elicited in immersive displays mediated by technology (Sense of Physical Space). It also measures related but distinct components of that 'presence' experience. 'Engagement' includes measures of enjoyment and losing track of time. 'Ecological Validity' measures the users' perceived naturalness and believability of the immersive experience. 'Negative effects' measures adverse reactions to being immersed, particularly when technology is not optimised for the UX e.g. non-optimal inter-pupillary distance for 3D displays may cause eyestrain and headaches.</p>
Unusual perceptual experiences	<p>8 items derived from Altered States of Consciousness questionnaire (Dittrich, 1998).</p>	<p>The extent to which the user experiences unusual sensations and perceptions, some of which may be associated with transcendental, almost spiritual and perhaps even magical experiences e.g. floaty, at peace, a sense of self-disappearing.</p>
Meaning/Sense Making (Conceptual Sense)	<p>1 item from i2's unpublished Salience Scale (It made sense to me).</p> <p>2 items measuring elements of Preparedness/Self explanatory content.</p>	<p>The focus group research indicates that audience experiences of immersive content are not always self explanatory or conceptually meaningful (or salient) to the user as intended by the creator, and in some instances e.g. exhibitions, audiences are prepped for their experiences.</p>
Expectations	<p>2 items measuring Expectations (met, unexpected).</p>	<p>Related to preparedness, the focus group research highlights a distinction between the experience of having expectations met, and the experience of something unexpected. Both aspects could be considered positive and/or negative, e.g. meeting expectations could be boring as well as satisfying, and experiencing something unexpected could be unpleasant and shocking or pleasing and relieving but nevertheless incredibly impactful.</p>

Table 1. Data gathered for user trials

DATA CATEGORY	DESCRIPTION OF DATA	DESCRIPTION OF WHAT THE DATA IS ATTEMPTING TO CAPTURE
Repeatability	3 items measuring Repeatability.	Impactful experiences may or may not be related to a desire to repeat the experience, perhaps depending on the content design and whether different storylines seem possible. Three inter-related items (being glad they had the experience, and desire to experience similar or indeed the same content again) were generated to evaluate the 'stickiness' of the content and style, which may hold potential for cumulative impact.
Experience of technology in this context	3 items measuring aspects of perceived quality of technology implementation (Clever Use, Glitchy, Gimmicky). 3 items derived from the System Usability Scale (Digital Equipment Corporation, 1986; e.g. "I thought this technology was easy to use").	Psychological impact of content can be 'broken' or diminished by poor technology implementation, for instance in terms of poor usability or noticeable glitches e.g. recall the days of 'snowy' pictures in the analogue TV world, or the 'blockiness' of digital signals.
Reputation	3 items measuring interest in the Production Team and Funders.	As an emerging medium, immersive content funders and producers in the focus groups are understandably keen to know whether and how their projects reflect on them. This type of impact is perhaps akin to an audience following a particular film director's work.
Shareability	2 items measuring intention to share/knowing others would like it.	Shareability of content, combined with the power of social media, have crucial implications for a Content's content's impact. These two items were developed to measure actual intent of trial users to share their experience, and the social power of being an originator of a socially appealing experience.
Self-consciousness	1 item: "I felt self-conscious".	Literature review and the focus groups highlighted the conflict between the personal immersive experience and being witnessed by others in the real social world beyond their the headset. We wondered if speculate that self-consciousness might perhaps weaken experienced impact.

Table 1. Data gathered for user trials cont.

DATA CATEGORY	DESCRIPTION OF DATA	DESCRIPTION OF WHAT THE DATA IS ATTEMPTING TO CAPTURE
Economic Value measures		
Willingness to Pay - home-use scenario for the content	Van Westendorp's Price Sensitivity Meter, 4 items comprise price points to be specified: too expensive, too cheap, not too expensive, valued bargain.	Users are given a scenario (outlined in Section 5) and then asked whether they would be willing to pay in order to experience that scenario. If they indicate that they are willing to pay, they are directed to answer the 4 questions that make up the Price Sensitivity Meter – which, according to Van Westendorp, creates a set of boundaries at which the acceptable price of a product can be found.
Willingness to Pay - arcade scenario for watching content	Van Westendorp's Price Sensitivity Meter, as above.	Same as above.
14/21-day follow-up survey		
Memorability and actual sharing	<p>Free recall: What about the contents do they remember? e.g. colours, title.</p> <p>Prompted recognition of the contents they viewed from the descriptions.</p> <p>Global experiential quality indicators (10 items, as for Post- and recalled negative effects e.g. ill, nauseous, dizzy).</p> <p>Shared (2 items: told/discussed in positive/negative/neutral way and whether shared on social media).</p> <p>Note that none of the respondents are forewarned that they will be re-contacted.</p>	The main purpose of the follow-up survey is to establish the longer term impact and 'memorability' of the immersive content experiences, and to measure behavioural sharing of those experiences following the trial. As free recall is cognitively more demanding than using prompts to recognise memories of experience, the survey begins by asking users what they remember (free recall) before providing prompts. They are asked again about the global experiential items, and some negative effects items from the SOPI to identify if there has been any changes in perceived impacts (positive or negative) over time.



BACKGROUND, LITERATURE REVIEW AND FOCUS GROUPS

The project focuses on the area of pre-testing which typically involves evaluation of an experience (in this instance, creative content in an immersive form) with a small representative sample of a given target audience population, in order to understand and predict the wider audience response.

BACKGROUND THE RESEARCH SPACE

We broadly carve the media experience space (see Figure 1.) conceptually into the following interacting elements or components:

- Media form (in this instance, the immersive technology system and what it affords).
- Media content (the creative elements, e.g. narrative, story, genre; note: creative content presented/produced with immersive form properties may or may not afford interaction as one of its formal properties of immersion).
- The media user and their characteristics (e.g. user attitudes to the form and content; their propensity to feel immersed with a mediated display, e.g. suggestibility; their mood state).
- The context of use (e.g. home/public, alone/social etc.)

This framework supports the formulation of variables and their interactions in order to develop and test, in a reliable and valid way, the constructs of interest from a UX perspective.

Immersive content can have impact and value in a variety of domains, including the psychological, social, cultural, and economic. The scope of this paper is guided by the research brief and is further explored by reviewing the extant literature and through primary research with focus group participants and qualitative interviews.

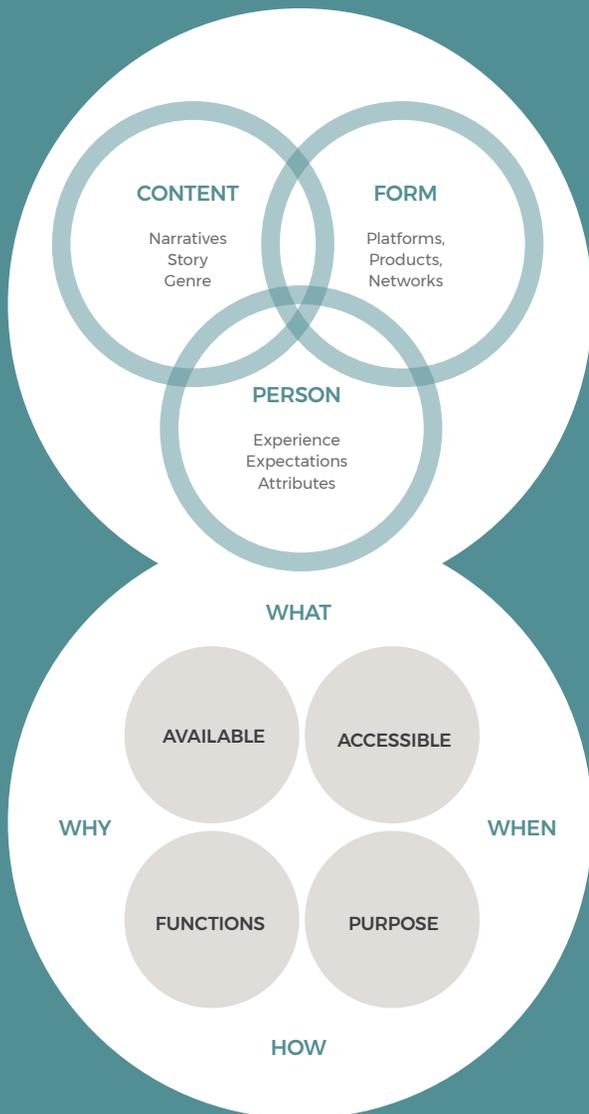
Estimating or measuring the future impact of immersive content has both short and longer term dimensions. We also consider the properties/qualities of what it might mean to have a lasting impact.

There are a number of points to disentangle around exploring the meaning of impact. In particular, what would constitute a good or bad impact? What would be considered ‘successful’ and conversely unsuccessful outcomes and whose judgment matters? Over what period is impact considered relevant?

For example, audience pre-testing may provide a useful indicator of immediate experiential impact, but these ratings may bear little relevance to either peer review of that same content, or the subsequent economic or critical success that the content may enjoy. Impact, then, could be from the perspective of the end user, industry, or relate to the similarities and differences between the end UX and the content creators’ intentions while making the work. There may not be consistency across all of these facets – for example, box office flops may later translate into home viewing successes in terms of economic performance. None of these challenges are unique to immersive experiences, of course, but navigating the landscape in an emerging media form is much more challenging.

UX IN CONTEXT

Figure 1. The conceptual media space



LITERATURE REVIEW PSYCHOLOGICAL IMPACTS OF IMMERSIVE TECHNOLOGIES

There are a wide range of psychological studies looking at immersive experiences which are relevant for this study, including those considering video games, advertising and film. We summarise some of the key concepts relevant to our study. The literature presents a number of constructs which have broad application in media, but which have particular significance for immersive experience.

The inter-related constructs of relevance, salience, attention, interest and engagement are important to psychological impact. For instance, manipulations of the media form e.g. size in visual field, brightness can increase salience (it stands out) because it makes the experience more difficult to ignore by increasing prominence within the person's sensory field.

Similarly, interactions between the media content e.g. genre and the person; genre preferences, can also impact what is considered a salient, captivating experiential impact in both short and longer-term periods e.g. memorability. People's attention is also particularly sensitive to novelty; VR technology is new and offers a different type of mediated experience to other media technologies. People's attention is often distributed (focus and passive monitoring) and they can be sensitive to mismatches in perceptual expectations, such as computer glitches.

The things that grab a person's attention are therefore evaluated against associated expectations (at conscious and unconscious levels) based on previous experiences. It is important to explore against what references users will judge, evaluate and compare immersive VR experiences to less immersive mediated ones. The relationships between attention and memory

(recall and recognition) and what they might tell us about psychological impact are also of interest. How users make meaning of experience also provides a context for understanding impact. What aspects of the experience linger for end users, what do they tend to remember and share about that experience with their peers?

In the wider context of motivation, the notion of 'appetite' (or appetitive motivation) is relevant to understanding repeatability of experience and what drives repeated experiences for the audience.

Presence – the sense of being there in an alternative (partially or fully) mediated environment as compared with being here in the fully non-technology mediated real world – is a commonly reported user experience of immersive media, and to varying degrees, other media, such as reading books. It has been documented and explored widely in the psychology and human-computer interaction literatures. In immersive settings, users commonly report an overall (holistic) feeling of being 'transported' to another place. This subjective sensation tends to increase as the level of physical immersion increases e.g. 3D tends to provide stronger experiences compared with 2D visuals, and likewise for surround sound compared with stereo. Interestingly, people tend to unconsciously 'fill in the gaps' of their experiences, so if the experience is sufficiently compelling, they may feel even more saturated by the experience than seems warranted. For instance, in VR, walking the plank (Richie's Plank) is psychologically difficult for many users, even though they know they are safe.

Researchers have developed tools for measuring presence, including via self-report questionnaires e.g., ITC-SOPI: Lessiter et al., 2001; MEC-SPQ: Vorderer et al., 2004, and measures of behavioural realism – behaviours

and physiological reactions such as sweating that would be expected in the real world e.g. Freeman et al. 2000.

Presence may be a useful predictive experiential component of a successful high-impact experience. Formal properties of the physical display and the immersive content/narrative interact with person factors to produce user sense of presence.

Related to these psychological impacts, some relevant studies are set out below. Shortened versions of a number of the research instruments outlined are used within the final protocol developed by Nesta and i2 Media Research.

VIDEO GAMES

- The evaluation of impact of immersive experiences is probably most prominent in the video games industry, and there is a wealth of published research investigating the psychology effects of immersive gaming (for example, see Boyle, E. A., Hainey, T., Connolly, et al., 2016).
- Ravaja et al. (2004) explore whether emotional response patterns to video game content characteristics could have practical pre-testing implications for optimising user experiences of entertainment; their inference being that emotional responses represent more powerful predictors of game success than asking players how much they enjoyed the game ('good game'). In a study with 37 (almost exclusively male, and young, 20-30 year olds) undergraduates, Ravaja et al. measures presence (using the SOPI) and other emotional game-related responses e.g. state measures of joy, pleasant relaxation, to four games (Tetris, Super Monkey Ball 2, Monkey Bowling 2 and James Bond 007: Night Fire).

The game content broadly varied in terms of game viewpoint, naturalness, and violence. They also measure two more stable trait characteristics of the participants: Impulsive Sensation Seeking (using the Zuckerman-Kuhlman Personality Questionnaire) and Self Forgetfulness (measured with the Self Transcendence Scale). Game characteristics are found to be associated with different emotional responses, and game engagement varies as a function of Sensation Seeking. The authors suggest that different games might be optimised for different personalities. The 'goodness' of the game is related more to emotional intensity than emotional valence, though this may depend on intended duration of use/exposure. It is worth noting their sample bias; sex differences may also be a relevant exploration.

- Coppi et al. (2014) explore the impact of embodied interaction on subjective experiences when playing a video game using either a standard joystick or a motion-sensitive Kinect device. Using post-test self-report measures of presence (ITS-SOPI) and flow (using the Flow Short Scale: Engesser & Rheinberg, 2008), they report increased Flow Absorption and SOPI Sense of Physical Space, Engagement and Negative Effects for embodied interaction relative to standard interaction, demonstrating the powerful impact of purposeful and free user movement in mediated spaces. Conversely, how appreciation of content is impacted on when user movement is implied yet not afforded, is also of relevance here. Expectations not met may increase a sense of restriction and frustration in the mediated display.

FILM

- Hu, Janse and Kong (2005) explore how a user's fun and presence experience of short film content is impacted by the distribution of media presentation. The UX of 'Deep Sea Adventure' comprises a 3D virtual underwater movie presented on a 42" plasma display with HiFi audio system and coloured lighting in the viewing room. The experience affords interaction via a portable touchscreen display (iPronto) and GamePad controller, as well as a toy submarine. The three presentation modes of the film are: automated (akin to traditional TV viewing), game (left and right navigation of the experience gamified such that the user avoids obstacles for points), and discovery (free navigation of 3D space and speed, points gained for collecting fish). They find that increased user control generally increases sense of presence (measured with the SOPI) but also increases task difficulty. Furthermore, increased control does not increase fun, which they attribute to the potentially low sensitivity of their measurement tool (Appeal questionnaire) which is designed for passive, rather than interactive media. This suggests that the choice of measurement tool is critical for identifying true effects.
- Hu, et. al., (2005) also consider the effects of the distribution of presentation, and find that only ratings of 'naturalness' (from the SOPI) are impacted. Their qualitative research suggests that presentation distribution also increases distribution of attention which is distracting, and suggests that developers should take care that visual attention is managed in more distributed immersive environments.

THERAPEUTIC, HEALTH AND WELL BEING APPLICATIONS

- The relationships between presence and emotion are complex, and others have found no significant differences between indicators of immersion e.g. 2D/3D on presence and emotion. In one such example, Baños et al. (2008) emphasise that the content of their [tested virtual environments are designed to enhance positive affect, which may mitigate the usually enhanced impact of immersion.
- These studies show how ratings of content, concept appreciation and evaluations of 'success' may relate differently to presentation factors, person factors, and content factors. The research reinforces the importance of the quality of measurement tools, and also the benefits of using a multi-method approach to making sense of data.

ALTERED STATES OF CONSCIOUSNESS

- Suzuki et al. (2017) explore manipulations of an immersive visual display and what it tells us about altered states of consciousness. In the "Hallucinogenic Machine", photorealistic content is presented panoramically on a head-mounted device and an algorithm ('Deep Dream') is applied to modify cues in the visual field and sensori-motor contingencies. Psychological effects evoked by the Hallucinogenic Machine are found to be comparable in some ways to those induced by psychedelic drugs. They use the Altered States of Consciousness questionnaire (Dittrich, 1998) to measure different aspects of experience during unaltered panoramic video compared with the same video manipulated with Deep Dream. They report significant differences in subjective ratings of intensity, patterns, imagery, ego, arousal, strange, vivid, space, muddle, spirit across the different types of video image experiences.

CULTURAL VALUE

One of the assumptions made in the development of the research protocol is that the immersive experiences evaluated through the study are in fact, cultural experiences, and therefore can be subjected to the same kinds of analysis as those which have taken place in domains such as theatre or music. There is next to no literature on cultural value in relation to immersive experience, however there is a large body of academic work looking at approaches to cultural value more generally.

Understanding the Value and Impacts of Cultural Experience, a literature review published by Arts Council England in 2014, explores the different methodologies that researchers have used to try and capture the intrinsic value of the arts. One common finding is that the value tends to depend on the individual's wider interaction with culture. That is, those with a sustained interest in culture, or equivalently a high degree of cultural capital, are likely to feel additional benefits when exposed to the arts. The report makes an important contribution in situating cultural value on a timeline that does not only occur when the exchange is made (e.g. the moment the artwork is seen), but starts from the moment a cultural event is known about, to the ticket buying and so forth, on to the person's feelings about the piece after an experience.

Building on this, Knell and Whitaker (2016) separate out the artistic quality of the cultural experience e.g. the play from the experience of the participant, whether individually or collectively – their engagement, social interaction and so forth. There is also broad agreement amongst academics using post-experience survey methodologies, that the experience must be viewed in the context of the particular goal of the cultural experience e.g. whether it is supposed to be educational or, say, transcendent.

One of the primary findings of the Arts Council review is that while commonalities across questions often exist between surveys, the interpretation of those questions can be different, leading to different categorisations of where 'value' might lie. In the absence of agreement about this though, we should be explicit about where we will identify the value of immersive experiences, recognising there is likely to be some ambiguity.

The Quality Metrics project, which began in Australia before being trialled in England, is an example of how a simple evaluation approach can be used to measure cultural value. In a trial of 150 arts and cultural organisations, 12 common indicators were used across three constituencies – the arts organisation, the audience and the organisation's peers – to determine quality. Audience responses, it turns out, tend on the whole to align with self review. However, in aggregate, peer-reviewers tend to be harsher.

WILLINGNESS TO PAY

As we noted at the outset, the market for VR and other immersive content is growing, but low current levels of market penetration means that most of the public are unaware of prices – either on platforms such as Viveport or at one-off exhibitions and installations.

Absent a defined market, therefore, this study explores the possible economic value of immersive experiences to audiences. In contingent valuation studies, users or customers are directly asked about the price point at which they would be willing to pay for a product or service.

Bakhshi et al. have in a number of studies tested the applicability of contingent valuation (CV) methodologies in arts and cultural organisations.

They conclude that:

'CV is a viable approach for measuring economic values in the context of cultural institutions. The method produces realistic values that vary in ways that are consistent with economic theory and previous findings, across different institutions, scenarios, payment vehicles and population groups.' (Bakhshi, et. al., 2015)

While the present study considers immersive experiences, the type of content might arguably be seen as equivalent with the outputs of a cultural institution.

Another similar approach to determine viable pricing, drawing from marketing research rather than economics, is the Van Westendorp Price Sensitivity Meter (PSM). This method, rather than asking for one specific price, asks four questions to elicit a range of acceptable prices for a product.

The questions are as follows:

- At what price would you say this would be too cheap?
- At what price would you say this would be a bargain?
- At what price would you say this would be getting expensive, but you would still consider it?
- At what price would you say this would be too expensive to consider?

In an examination of three forms of valuation technique (contingent valuation, Price Sensitivity Meter (PSM) and the Becker-DeGroot-Marschak mechanism), Marcus Kunter finds that the PSM produces robust predictive values in the optimal price point it derives, while not being able to rule out that these are because of biases within the survey. Against this, the PSM is seen as intuitive by respondents and is often applied to determine prices for new and innovative products (Lyon, 2002). For this study, therefore, we use the PSM to analyse the willingness to pay of users in relation to immersive experiences and VR content.

SOCIAL VALUE

The literature review uncovers a number of interesting studies in the grey literature focused on social value. How virtual reality facilitates social connection, commissioned by Facebook, examines how people interact in two identical situations – meeting a stranger on a train – with the exception that one experience is in person while one is in VR. The study finds that those interacting within VR do not have to exert undue cognitive effort, and that conversations generally go by quickly. Introverts are particularly drawn to interacting in VR, and respondents note that they find themselves able to discuss even personal subjects with greater ease.

In 2017, the BBC conducted a longitudinal piece of work to study the impact of VR on participants that had previously had little interaction with the technology. Specifically, they gave a number of individuals a headset for a three-month period, conducting interviews with the participants and asking them for self-reported reflections on the technology.

One of the key findings of the study, which has an interesting implication for the social impact of VR and immersive experiences, is that audiences find VR to be a technology that requires undivided attention. Whereas traditional media forms are now dominated by the use of multiple screens, VR provides a space where attention has to be on the content within the headset. This is identified as a unique characteristic of VR, and also a challenge:

Social interaction – for some audiences the insular / individual nature of the experience was off-putting as they preferred connecting with others either digitally or in physical space.

FOCUS GROUPS AND SEMI STRUCTURED INTERVIEWS

Within each of the three focus groups and interviews, participants were asked about both their conceptions of 'impact' in a broad sense, along with the support mechanisms and measurement tools of that impact.

CONTENT CREATORS

The qualitative work suggests that content creators primarily viewed impact in relation to achieving their artistic intentions. Much of the impact they described can be seen as applicable to cultural experiences outside of immersive content – for example, delivering powerful, compelling experiences and great storytelling, which does not necessarily relate specifically to VR. That said, there were a number of impacts that they sought to deliver that were more closely linked to the medium itself. For example, a number spoke about the idea of transporting the user, and presence - the idea that the experience was putting you into another place, person or time.

There were a number of technical aspects of the experience that content creators particularly stress in relation to VR. Giving the audience or user a sense of intuitive interaction with their virtual environment, along with a sense of agency within the environment were seen as important facets of the VR experience. Some participants of the focus groups and interviews also noted that virtual environments are particularly strong for eliciting certain emotions. For example, a number mentioned that VR is particularly good at creating empathetic experiences, allowing users to see things from the perspective of others.

The relatively novel nature of the technology (particularly with new-generation VR headsets) means that content creators have a heightened sense of how the technology itself relates to the experience. For example, a number of interviewees mentioned the quality of the rendering in the content they had developed, and how important it was to ensure a seamless experience in terms of the technical delivery in VR.

END USERS OF IMMERSIVE CONTENT

In their focus group, the end users, or audience, of the immersive experiences described impact in terms of a mix of personal variables that relate to the specific affordances of technology, and more general responses to cultural activity. For example, they described the sense of spatial presence, or being transported to another place, as a specific impact. However, they described more generic forms of impact too, such as experiencing satisfying and rewarding or surprising and novel content.

The end users were particularly clear that impact was dependent on their mood state. Allied to this, it is thought that certain personality types might be more suited to highly impactful immersive experiences, particularly those who have low cynicism, [and] high openness to experience.

The onboarding of immersive experiences was seen as particularly important to impact. Aiding the transitioning between the real and immersive worlds is one of the ways to support impact generally. Allied to this, the comfort and quality of the headset, and a glitch-free experience with visual fidelity are also valued.

End users raised a variety of concerns too, including those relating to motion sickness, self-consciousness (in social settings) and ethical concerns with content; these were all seen as potential barriers to rewarding immersive

experiences. Several end users also pointed to a paradox in some immersive experiences, in being about relaxation on the one hand and being stimulated and having novel experiences on the other.

IMMERSIVE SECTOR STAKEHOLDERS

Immersive sector stakeholders, such as funders, industry trade body members and sector champions, were a little more hesitant in defining the parameters of impact specifically. One stated that it was potentially too early to have a good feel or define impact within the sector, and that it remained to be seen in which market segment experiences would have the most penetration.

Some expressed the view that storytelling, narrative and creative expression are more primary drivers of impact than the technical affordances of immersive technology. There was also some support for the view that VR creates a paradigm shift in the production of cultural content more broadly – providing an altogether new kind of storytelling experience.

Stakeholders had a much firmer view of impact in terms of hard outcomes than the other groups, however, such as audience figures and reach of pieces of content, and whether the content has impact in the eyes of peers and critics. There was also interest in pursuing the potential of VR for enhancing accessibility to culture for people with sight loss or other disabilities.

One particular area of interest for the stakeholders who participated in the research was in the sense of agency that immersive experiences might give back to users, as opposed to other, more passive forms of media. Mixed media with or without location-based experiences, such as immersive content that relates to and extends existing content (e.g. TV series, film) were seen as a key opportunity.

RESEARCH FINDINGS

The UX trials were conducted at i2 Media Research's research lab at Goldsmiths University of London and ran from December 2017 to January 2018.

1. SAMPLE AND DESIGN

The trials involved 84 participants, each of whom evaluated two of three test contents, which are anonymised here and referred to as Contents "A", "B" and "C". This gave 168 content views (A n=54, B n=57, C n=57), split in a way so as to counter any ordering effects that might otherwise take place within the trial, and 168 corresponding participant evaluations.

Participants first completed a pre-test measure (comprising a short survey of socio-demographics and prior engagement with technology and arts and cultural attendance), before being prepared for their immersive experiences. Each participant first completed the current Vive tutorial to familiarise themselves with controls within the VR experiences in question, and then engaged with each of the two contents assigned to them. After each content experience, participants completed self-report evaluations (which took the form of a web-based survey), where participants rated the content for perceived overall quality, cultural value,

and psychological impact, including positive and negative affect/emotions, and willingness to pay (described in detail in Section 3). Additional data on intended audience impacts was acquired from the content creators (n=3), and a follow-up survey (n= 24) issued 2-3 weeks after the trial exploring content memorability.

The sample age was skewed towards younger participants and just over half the sample was female. The majority (68%) described themselves as either "White British" or "White Other". About a third of the sample had no previous VR experiences, whilst 42% had experienced VR once or twice, and 18% had experienced VR on three or more occasions. 7% of the sample reported owning a VR headset: Google Cardboard was the most commonly owned 'headset' (n=4), whilst two owned the HTC Vive headset used in the lab study (see Figures 2a-d below).

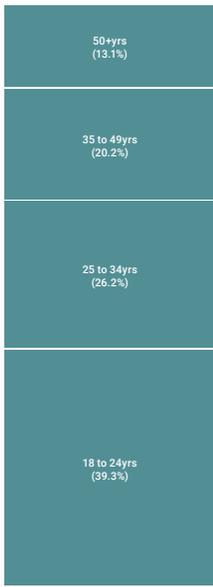


Figure 2a

Distribution of age in the sample



Figure 2b

Distribution of gender in the sample



Figure 2c

Distribution of ethnicity in the sample



Figure 2d

VR experience - Distribution in the sample

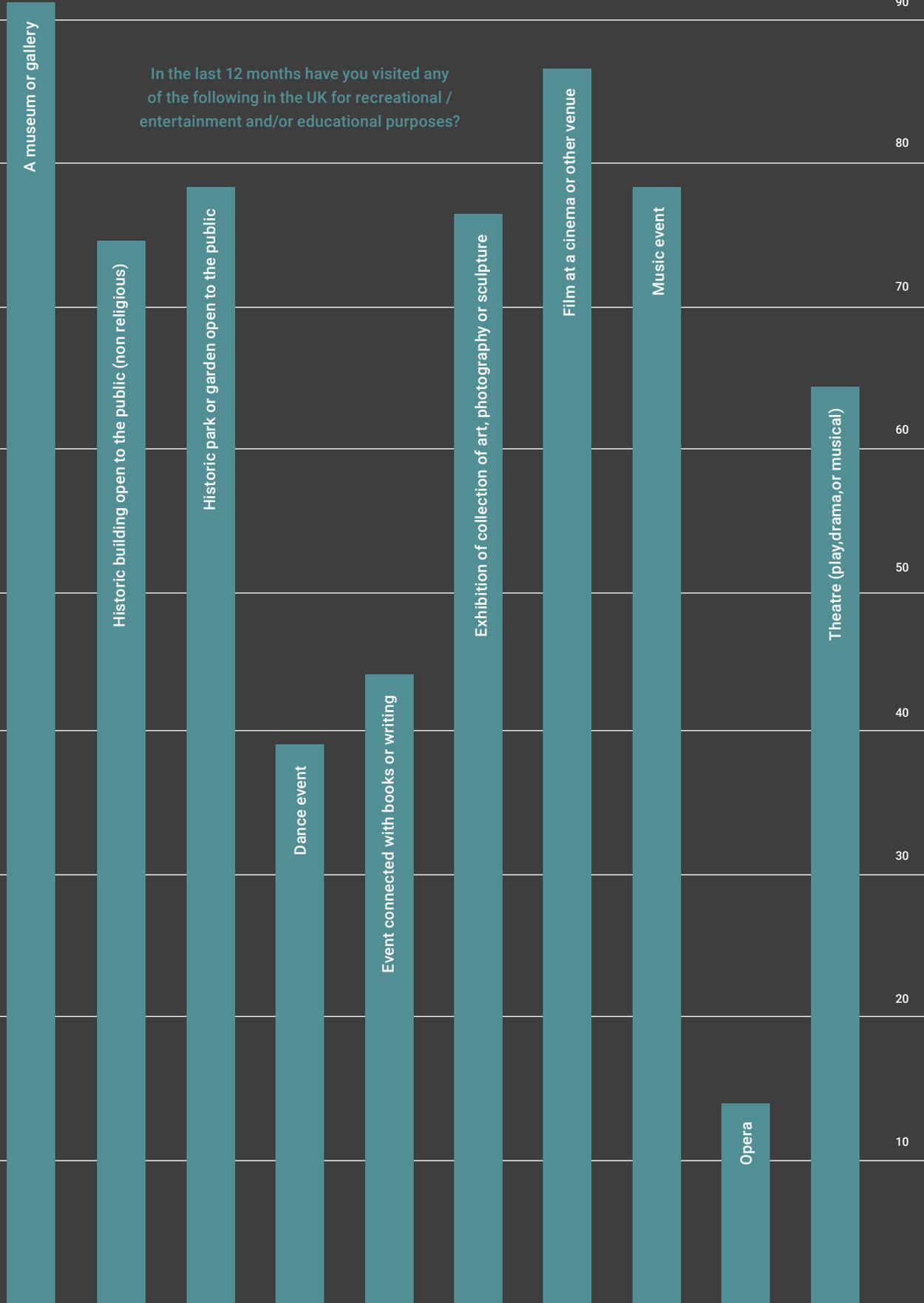
The participants reported generally high levels of attendance and engagement with arts and culture over the last 12-months (see Figure 3). The range of arts and cultural experiences varied widely. The most commonly reported were “Museum or Gallery” visit (92%), “Film at Cinema or other venue” (87%) and “Music Event”, “Historic Park or Garden open to the public” and “Exhibition or Collection of Art, Photography or Sculpture” (almost 80% for each). Least commonly reported were “Opera/Operetta” (14%) and “Dance Event” (39%) and “Event Connected with Books or Writing” (44%). While not directly comparable, we surmise that the sample was made up disproportionately of those engaged in culture. For example, Taking Part data from the DCMS suggests that only 52.3% of adults had visited at least one museum or gallery over the past 12 months.

A strength of the study design is that it enables a practical exploration of the impact of using more than one piece of content within an evaluation. The contents experienced were counterbalanced in their trial orders, such that around half of the participants for each content experienced it first, and the other half experienced it second. In the event, there turned out to be some interesting differences in trial order: for each piece of content, the ratings of that piece when viewed second in the trial order were consistently (and in most instances significantly) lower.

This could conceivably have reflected a number of different effects. For instance, the novelty of the VR experience in experiencing the first piece of content, delayed gratification from watching the preferred piece of content or trial fatigue after having been immersed in the first experience and having completed a quite long questionnaire. In terms of the patterns of results obtained – both in the descriptive findings and analysis of relations between the measured quality of experience measures (described below) – the results are generally consistent. Instances where this is not the case are indicated.

Figure 2a-d
Sample distributions of demographic measures

Figure 3
Cultural participation over the last 12 months of the research participants



2. KEY DESCRIPTIVE FINDINGS OF INDIVIDUAL CONTENT

The post-test evaluations provided descriptive data on different dimensions of the quality of experience and impact elicited by each content, in terms of the various forms of value examined through the study (global, cultural, psychological, economic). We report these descriptive data here, followed by an analysis of the relations between the different quality of experience dimensions measured. Where there are statistically significant differences, they are indicated at the 5% level.

Following the structure in Table 1, this section outlines the measures used within the research protocol. First, it discusses the general or global experiential quality indicators, before moving onto examining specific aspects of cultural value. It then goes on to look at the detailed experiential qualities, and finally the economic value of the pieces of content.

A. Global experiential quality indicators

In the post-test evaluation, participants first rated the content using 11 global experiential quality dimensions on a scale of 0-100 to represent their immersive experience: the higher the score, the greater the positive perception, e.g. Good vs. Not Good. These data were analysed to explore

descriptive qualities considered most and least inherently characteristic of each individual content. Whilst we highlight the highest and lowest scoring dimensions for each content, any score reported that is above 50 is towards the more positive end of the relevant dimension's scale.

The highest rated qualities differed quite considerably across the contents. While Memorable, Impressive and Good were in the top three descriptors for at least two of the contents (see Table 2), there was no common highest scoring descriptor across all three pieces. This leads us to conclude that there is a degree of sensitivity across experience to the dimensions used here.

When examining the dimensions that the users felt described the content least well, there is consistency in that Worth Paying For and Emotionally Moving were common across all three contents. Again, it is worth noting that there was only one negative (sub-50) score in the whole test – Content B scored only 42 out of 100 for being Emotionally Moving. Given this content was an animated, task-based experience it seems entirely consistent that user responses would form a negative judgement on its ability to be Emotionally Moving. We discuss the figures on whether the experience was Worth Paying For later in the paper when looking at potential economic value in more detail.

Table 2. Average (across trials) global experiential ratings characteristic of each content

	CONTENT A	CONTENT B	CONTENT C
Highest Scoring descriptors			
[1]	Memorable [80/100]	*Good [79/100]	Transporting [81/100]
[2]	*Impressive [79/100]	Transporting [77/100]	Memorable [80/100]
[3]	*Recommendable [79/100]	*Impressive [76/100]	Good [79/100]
Lowest scoring descriptors			
[1]	Worth Paying For [60/100]	*Emotionally Moving [42/100]	Worth Paying For [60/100]
[2]	**Emotionally Moving [61/100]	*Worth Paying For [60/100]	Emotionally Moving [65/100]
[3]	Rewarding [63/100]	*Rewarding [62/100]	Satisfying [72/100]

B. Cultural value indicators

Alongside the global experiential quality ratings, the study also used questions related to the perceived cultural value of the content. There were 14 indicators, rated on the same 0-100 scale as the global quality ratings above.

It is of note that there were overlaps across contents in the highest and lowest scoring qualities (see Table 3). For instance, having high scores on being Different as well as Shareable, were common to all three contents. The shareable

nature of the content was also common across all three experiences. In a similar manner to other cultural experiences, VR lends itself to group or social engagement, even though the experiences are primarily individual.

The lowest scoring cultural value indicators were also similar across content (Intellectually Engaging received some of the lowest scores across all three pieces). Note however that these 'lowest' cultural value ratings exceeded 50, still indicating positive impact.

	CONTENT A	CONTENT B	CONTENT C
Highest Scoring descriptors			
[1]	Different [84/100]	Able to Hold Attention [79/100]	Different [86/100]
[2]	Shareable [79/100]	Shareable [79/100]	Shareable [81/100]
[3]	Well Produced/Presented [77]	Different [78/100]	Interesting Idea [79/100]
Lowest scoring descriptors			
[1]	Intellectually Engaging [53/100]	Intellectually Engaging [50/100]	Better as Social [59/100]
[2]	Offers New Perspective [56/100]	Says Something about World [54]	Intellectually Engaging [59/100]
[3]	Says Something about World [59]	Thought Provoking [56/100]	Offers New Perspective [66/100]

Figure 3

Means totals for
(a) global experiential quality
(b) cultural value indicators
(c) Experiential and cultural totals combined

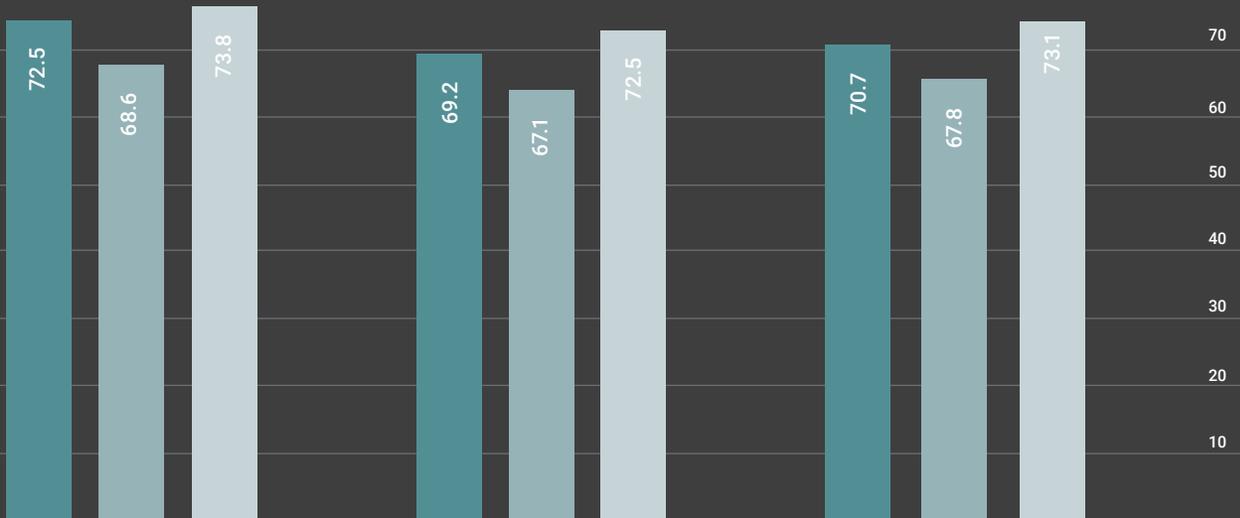
A comparison across the contents (totals for experiential, cultural value, then combined total, see Figure 3), indicates that Content C was on the whole most positively evaluated, followed by Content A, and then Content B indicating positive impact.

Global Evaluations: Mean Subtotal and Total Quality ratings

(A) Global Experiential Quality

(B) Cultural Value

(C) Global Quality Total



	CONTENT A	CONTENT B	CONTENT C
Experiences associated tendency towards Agree/Disagree or neutral [3]			
Presence	Sense of Physical Space [4.0] Engagement [3.9] Ecological Validity [2.9] Negative Effects [2.0]	Sense of Physical Space [4.0] Engagement [4.0] Ecological Validity [2.7] Negative Effects [1.0]	Sense of Physical Space [3.8] Engagement [3.9] Ecological Validity [3.0] Negative Effects [2.0]
Self-conscious	Felt self-conscious [3.4]	Felt self-conscious [3.3]	Felt self-conscious [2.9]
Unusual Perceptual Experience	Had unusual experience [3.0]	Had unusual experience [2.9]	Had unusual experience [3.3]
Meaning/ Sense Making	Made Sense [3.5] Self Explanatory [3.4] Wish Better Prepared [2.7]	Made Sense [3.3] Self Explanatory [2.9] Wish Better Prepared [2.6] Met Expectations [3.3] Unexpected [3.5]	Made Sense [3.2] Self Explanatory [2.7] Wish Better Prepared [2.8]
Expectations	Met Expectations [3.2] Unexpected [3.5]	Glad Experienced [4.3] Imagine More Experiences like this one [4.5] Like Repeat of Content [3.4] Plan to Tell Friends [4.1] Know People Would Like [4.1]	Met Expectations [3.1] Unexpected [3.7]
Repeatability/ Share	Glad Experienced [4.1] Imagine More Experiences like this one [4.1] Like Repeat of Content [3.1] Plan to Tell Friends [4.0] Know People Would Like [3.9]	Glad Experienced [4.3] Imagine More Experiences like this one [4.5] Like Repeat of Content [3.4] Plan to Tell Friends [4.1] Know People Would Like [4.1]	Glad Experienced [4.2] Imagine More Experiences like this one [3.9] Like Repeat of Content [3.4] Plan to Tell Friends [4.1] Know People Would Like [4.0]
Technology Quality	Clever Tech Use [4.1] Gimmicky [2.5] Glitchy quality [2.5]	Clever Tech Use [3.8] Gimmicky [3.1] Glitchy quality [3.9]	Clever Tech Use [4.0] Gimmicky [2.6] Glitchy quality [2.8]
Ease of Use	Easy to Use [3.8] Confident with Tech [3.7] Needed to Learn [2.4]	Easy to Use [3.8] Confident with Tech [3.7] Needed to Learn [2.3]	Easy to Use [3.7] Confident with Tech [3.5] Needed to Learn [1.9]
Reputation	See Another Team Production [3.9] Interest in Creative Team [3.7] Interest in Funders [3.9]	See Another Team Production [4.0] Interest in Creative Team [3.6] Interest in Funders [3.6]	See Another Team Production [4.1] Interest in Creative Team [3.6] Interest in Funders [3.7]

C. Detailed experiential quality indicators

In addition to the more generic impacts and cultural responses to immersive experiences, we can expect them also to invoke a range of more specific psychological and experiential impacts. In order to test these, we examined impact indicators including measures of: Presence (Sense of Physical Space, Engagement, Ecological Validity/Naturalness and Negative Effects), Unusual Perceptual Experiences, Meaning/Sense Making, Expectations, Repeatability, Technical Quality, Ease of Use, Reputation, Shareability and Self Consciousness. These are discussed in more detail in Section 3. Each item for each of these dimensions was rated on a 5-point scale from [1] Strongly Disagree, through [3] Neither Agree nor Disagree, to [5] Strongly Agree. Higher scores therefore indicate increased agreement with each statement.

Users of all three contents shared dispositions toward agreement with a Sense of Physical Space, Engagement; Meeting Expectations whilst also being Unexpected, Repeatable and Shareable; for all contents, Reputation fared positively. There were more nuanced experiences within and across content for other indicators, including Meaning/Sense Making and Technology Quality. A summary of the results for each content is provided in Table 4 below. We highlight the responses for each that scored over 4.0 out of 5, indicating a high level of agreement with the psychological and experiential impacts.

D. Positive and Negative Affects

Participants gave post-test ratings from 0-100 to a range of adjectives relating to mood states and emotions e.g. Interested, Distressed elicited during the experience, which we refer to here as 'affects'. These adjective descriptors were derived from a range of existing measures where possible. We used the 20 items in the Positive and Negative Affect Schedule (PANAS) to compute total Positive Affect (this is the mean of Interested,

Excited, Strong, Enthusiastic, Proud, Alert, Inspired, Determined, Attentive, Active) and Negative Affect (Distress, Upset, Guilty, Scared, Hostile, Irritable, Ashamed, Nervous Jittery, Afraid).

There was a similar pattern for each content whereby Positive Affect was more prominent (higher than 50) than was Negative Affect (lower than 50).

- Content A was considered to elicit some Positive Affect (62.9/100), and less but some Negative Affect (48.1/100).
- Content B elicited some Positive Affect (62.0/100) and a lower score for Negative Affect (45.1/100).
- Content C also elicited some Positive Affect (62.2/100) and lower but some Negative Affect (41.9/100).

E. Economic Value

The research questions for the study set out also to explore the potential economic value of immersive experiences, at a point in which market penetration of devices is still low and content is distributed in different ways. We therefore employed a methodology which reveals plausible ranges of pricing for immersive experiences, namely Van Westendorp's price sensitivity meter (1976) in relation to two hypothetical use-case scenarios for the content used within the trial. These scenarios are based on what is currently known about the VR market in the UK and globally; and draw on both the background literature and the focus groups. Note however that several other scenarios could have been used (for example, free platforms like YouTube for VR, or subscription-based services like Netflix for VR). The questionnaire asked participants to indicate price points at which the experiences would be considered: too expensive; too cheap; quite expensive but not out of the question; and a bargain. The scenarios were described as follows:

Virtual reality home use scenario

Think forward to 2-3 years in the future, when virtual reality headsets are more mainstream with consumers. You own a virtual reality headset such as the one you used in the trial today. Within your home, along with activities such as watching television or listening to music, there is the option to have virtual reality experiences by renting them through an online store for a few days. Once you have rented the content, you can use it again at your discretion for the rental period.

Virtual reality arcade use scenario

In future, virtual reality might be showcased in a setting where you and a few friends visit a virtual reality Arcade. This arcade would be similar to a multi-room cinema, in which there are a variety of similar-length (10-15 minute) experiences that you can pick and choose to partake in, shown in different rooms. In this way, you watch four pieces of content of a similar standard to the one you just had. The total time within the Arcade, with breaks in between content, would be around two hours.’

The price sensitivity data provided by participants across all 3 contents are given below (see Table 5 and Figures 4 and 5). The home scenario generated optimal prices that were on average £7 lower than for arcade scenarios, which was given a mean optimal price of £16.

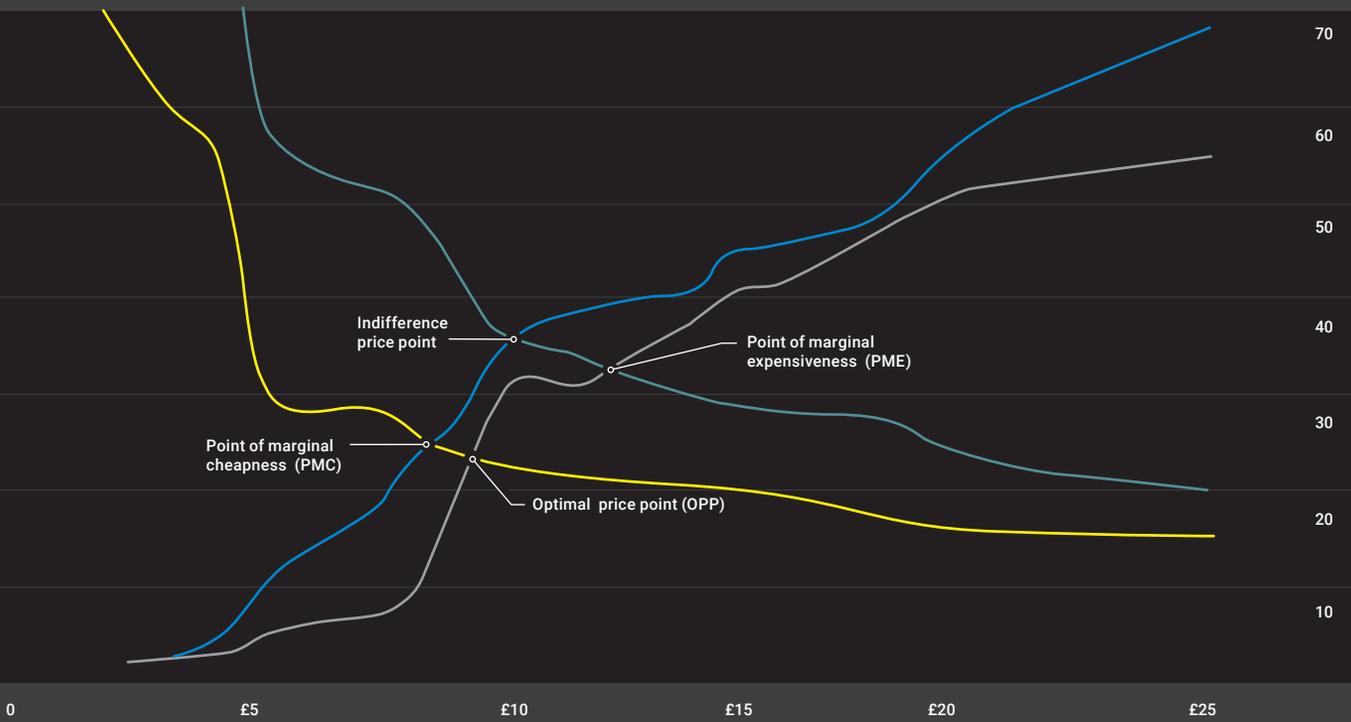
For the home use case, Content A commanded the highest price (£12) followed by Content C (£8.50), closely followed by Content B (£7). For the Arcade use case, whilst Content A still received the highest price (£17.50), Content B commanded a higher price (£16.00) than Content C (£11.50). Across all Content, the maximum price range did not exceed £18 for the Arcade scenario. The consistency of current pricing of VR content available to the public, in home and out of home settings, demonstrates high face validity of these estimates of users’ willingness to pay for immersive experiences.

Price-sensitive data are also shown separately for each Content (A: Figures 6 and 7, B: Figures 8 and 9, and C: Figures 10-11). The graph uses the standard presentation format for Van Westendorp’s price sensitivity meter, in which cumulative frequencies for each of the price categories e.g. ‘too expensive’, ‘too cheap’ are plotted, with the data for ‘too cheap’ and ‘bargain’ inverted to produce a graph with intersecting lines. The intersections are interpreted as providing an indication of different price points and a range of acceptable costs.

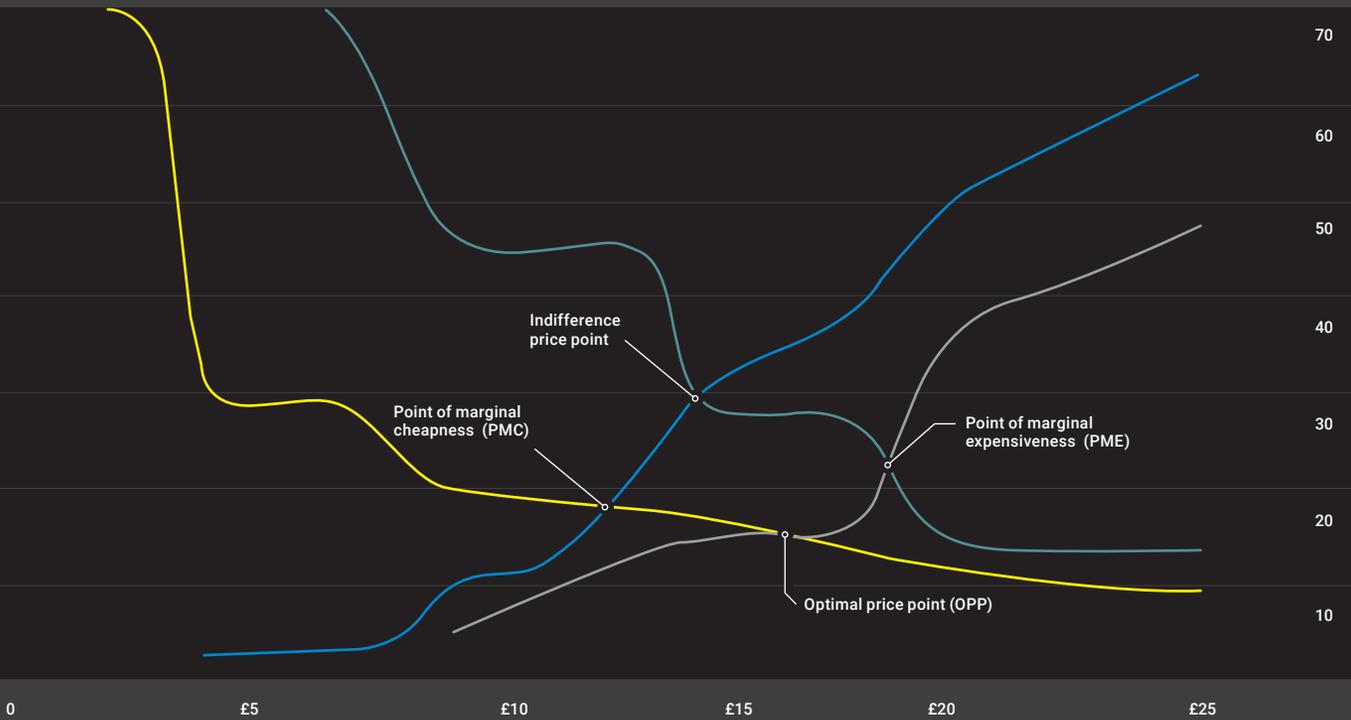
A prior question included in the survey showed that over half of participants (58%) indicated they would be willing to pay to have immersive experiences at home, and a higher proportion (70%) indicated they would be willing to pay to have immersive experiences in the Arcade scenario.

CONTENT	HOME USE	ARCADE USE
Content A	£12 (range £9-£15)	£17.50 (range £15-£18)
Content B	£7 (range £6-£8)	£16.00 (range £12-£18)
Content C	£8.50 (range £8-£11)	£11.50 (range £11-£18)
Across all Contents	£9 (range £8-£12)	£16 (range, £12-£18)

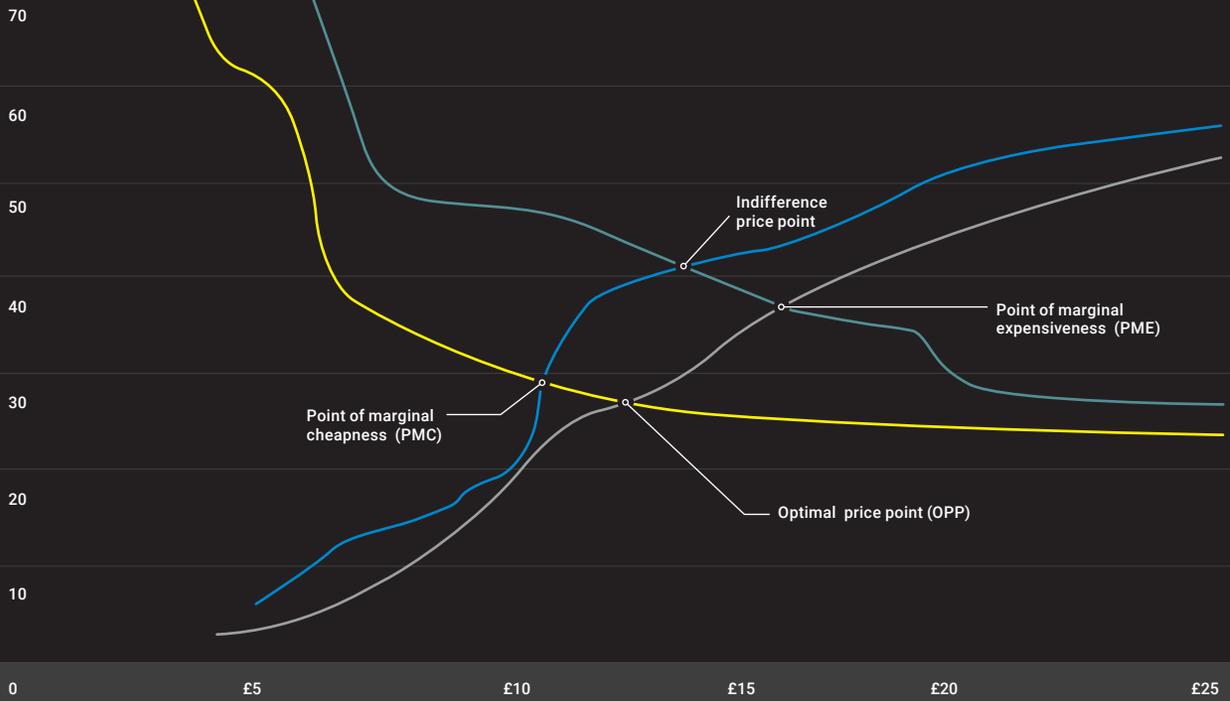
All content (WTP home)



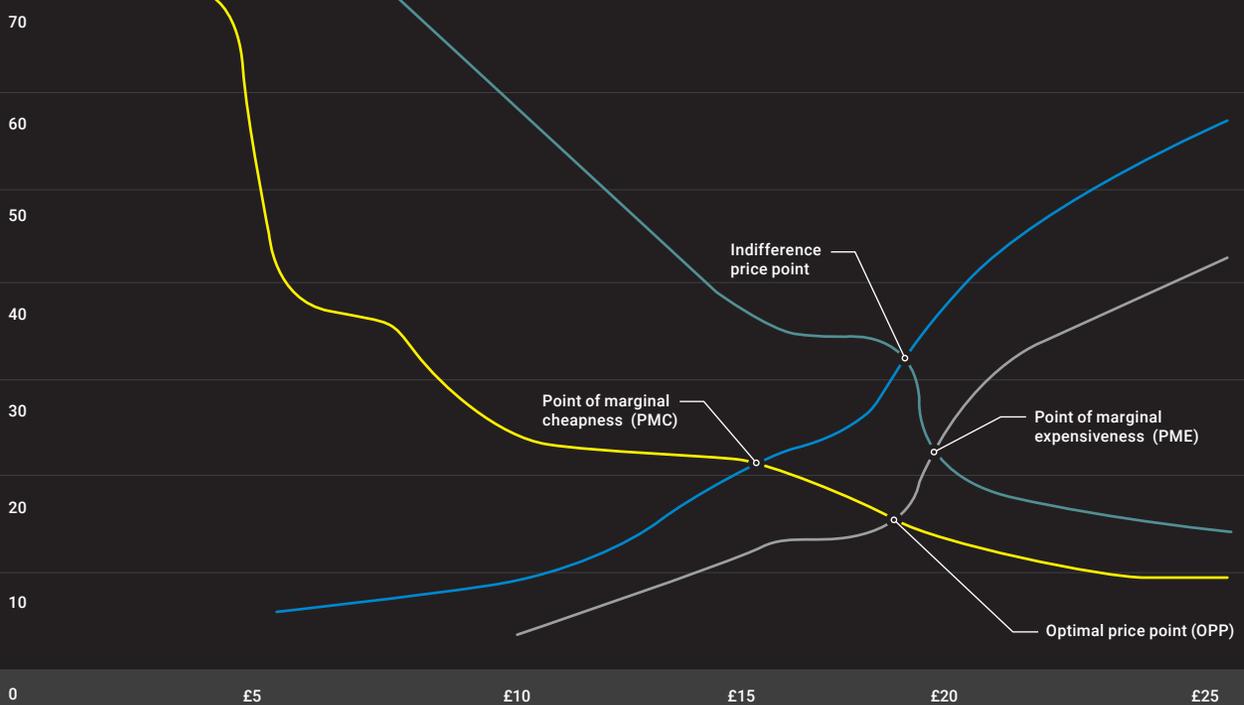
All content (WTP arcade)



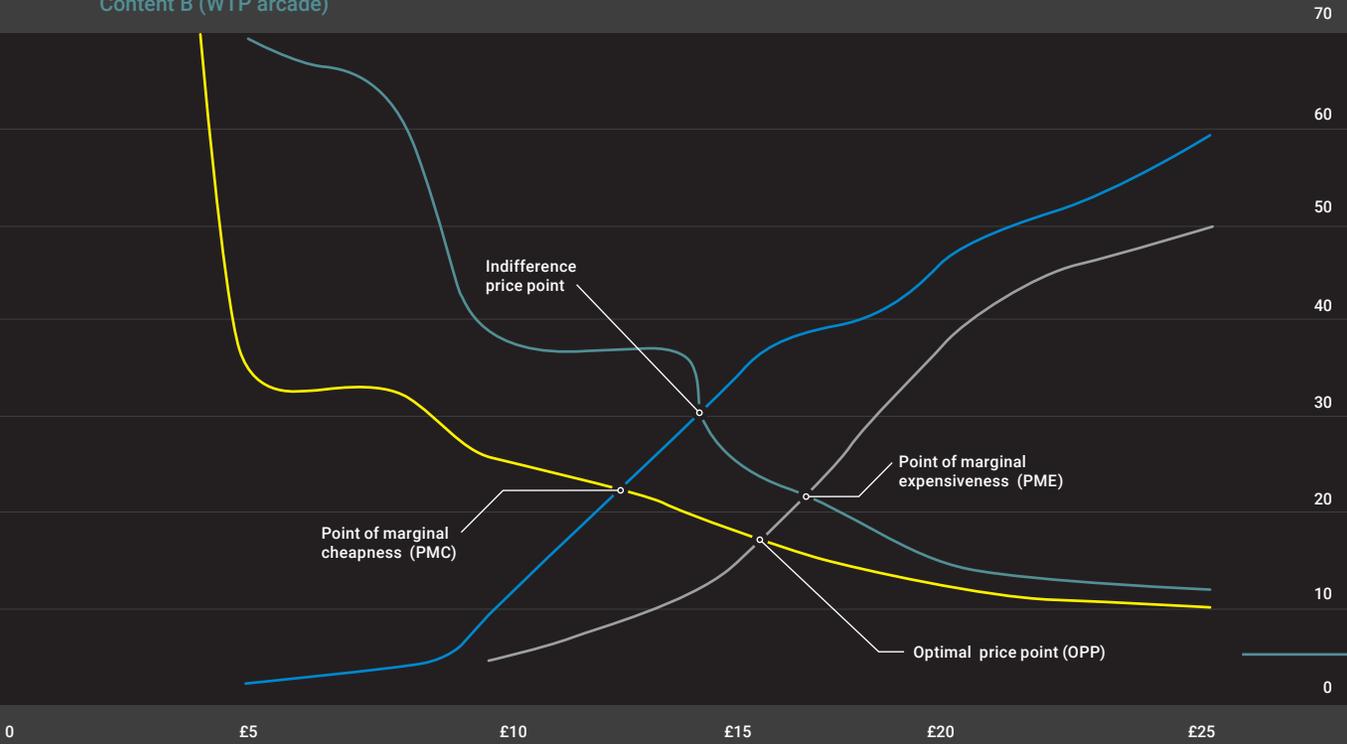
Content A (WTP arcade)



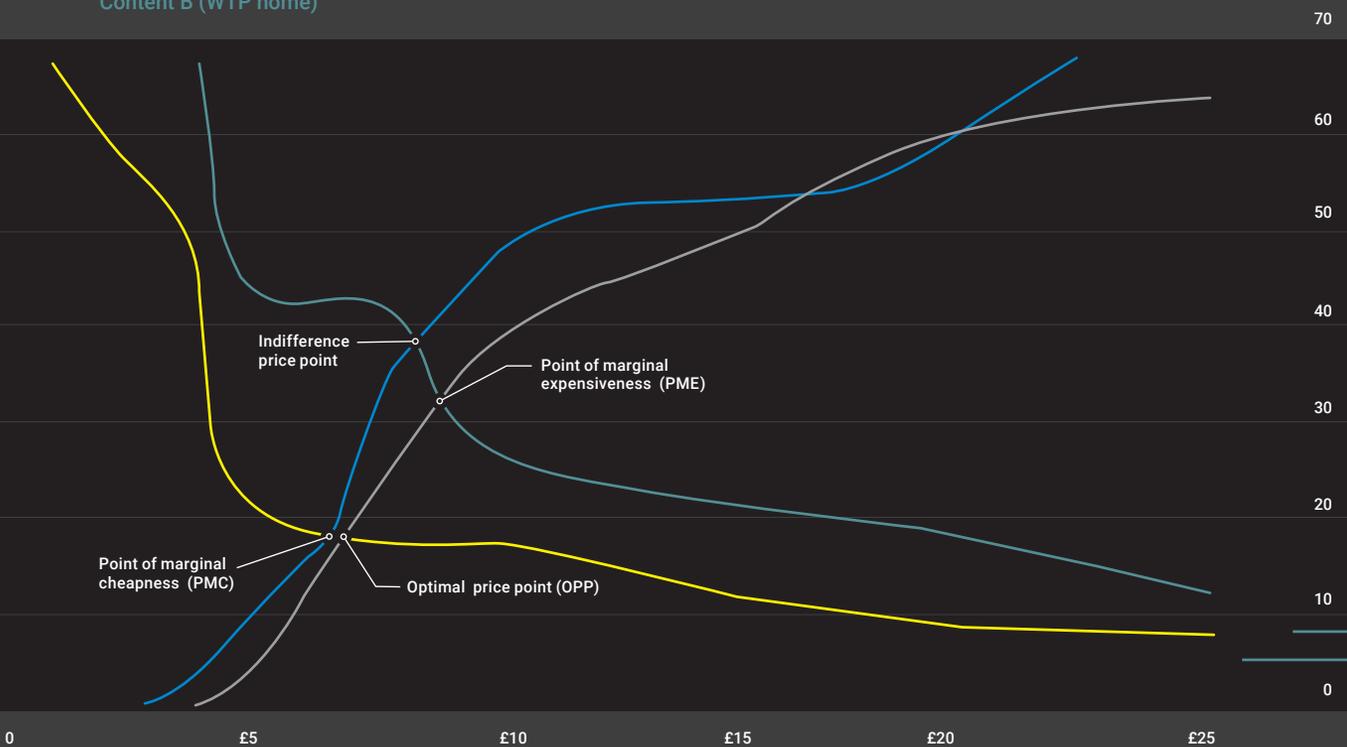
Content A (WTP home)



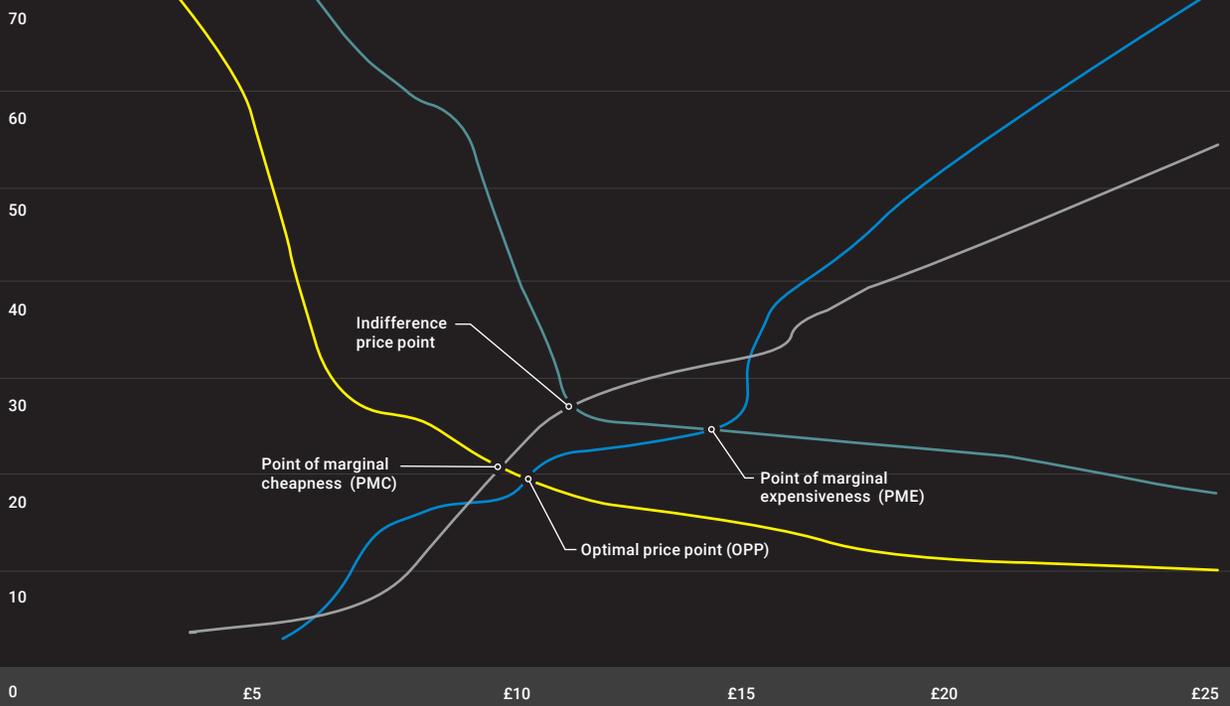
Content B (WTP arcade)



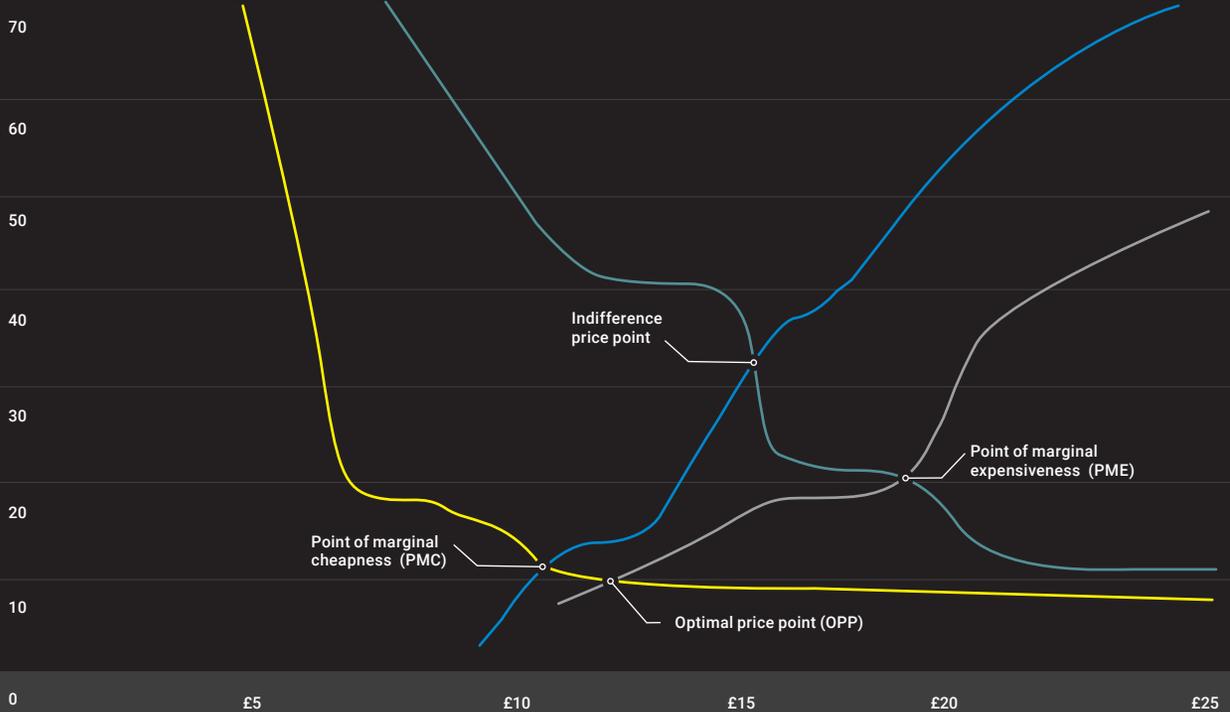
Content B (WTP home)



Content A (WTP arcade)



Content A (WTP home)



3. WHAT MADE THE IMMERSIVE EXPERIENCES MEMORABLE?

All participants were invited to fill in a follow-up survey, between two and three weeks after they had taken part in the lab trials. The goal of the survey was to assess whether memory matched to the immediate experience of the content. 24 participants responded to the questionnaire, just under a third of the original sample.

In order to acquire unprompted data on recall of the two contents, participants were asked at the beginning of the survey to freely describe what they remembered experiencing. An analysis of the qualitative data obtained showed that nearly the whole sample (22 out of 24) reported an accurate description of their experiences, which were also described in the order in which they were presented; in only 2 cases the order was reversed (in favour of Content A or C over Content B).

A disparity was found in terms of sample size for each of the content experienced: 75% indicated having watched Content A (n=17) (as one of the two content pieces experienced), 59% Content B (n=13), and 64% Content C (n=14). Of those who experienced it, 81% discussed/talked in a positive way about Content A, slightly less so in the case of Content C (77%), and 59% in the case of Content B. This finding should be treated with caution, however, as it is based on a low number of survey returns.

A comparison on the global quality indicators from the original sample with data from the follow up survey, showed an analogous pattern in terms of

means for the three contents: Content C held the highest global quality of experience rating (63.8/100), followed by Content A (62.8/100) and Content C (62.6/100). However, the actual scores exhibited a significant reduction (p. <0.02) in all three contents (a drop of 9.7 for Content A, 6 for Content B and 10 for Content C).

4. COMPARISONS BETWEEN CONTENT CREATOR INTENTIONS AND UX/IMPACT OF CONTENT.

Content creators were asked, via a separate survey, to indicate the intention for their content to increase (100), decrease (0) or not alter (50) each of the 29 specific audience's emotions and mood states (outlined in Table 6) using a scale of 0-100. We compared their ratings with those given by participants in the same section of the questionnaires administered in the lab trials ("During the experience I felt...: Much more (100), much less (0), about the same (50)").

The participants' ratings of these specific emotions scored consistently with content creators' expectations (above 50 when expected to increase, below 50 when to decrease), therefore matching the creators' intentions. In particular, the audience emotions that content creators intended to impact accounted for the majority of the top-scoring emotions elicited in user trials (4 out of 5 in the case of Content A, 9 out of 10 for B, and 10 out of 14 for Content C).

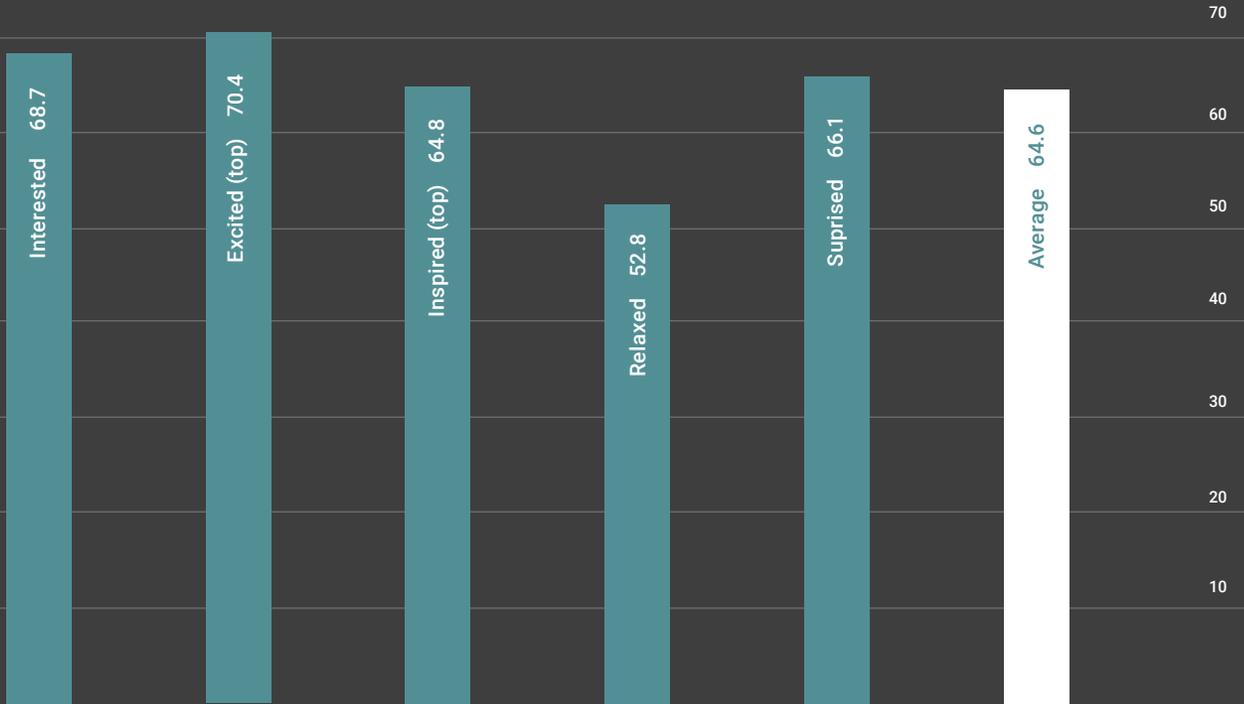
These participant results for each intended emotional change are represented graphically below (see Figures 12-15).

	EMOTIONS EXPECTED TO INCREASE	EMOTIONS EXPECTED TO DECREASE	SUCCESSFULLY ELICITED	TOP SCORING
Content A	Excited, Interested, Surprised, Inspired, Relaxed (5)	None	5 out of 5	4 out of 5
Content B	Interested, Excited, Attentive, Enthusiastic, Happy, Inspired, Surprised, Active, Determined, Relaxed (10)	None	10 out of 10	9 out of 10
Content C	Inspired, Interested, Excited, Attentive, Happy, Enthusiastic, Relaxed, Active (8)	Upset, Angry, Distressed, Anxious, Fatigued, Afraid (5)	14 out of 14	10 out of 14

Content Creator A: ratings obtained for emotions intended to increase

Content A ●

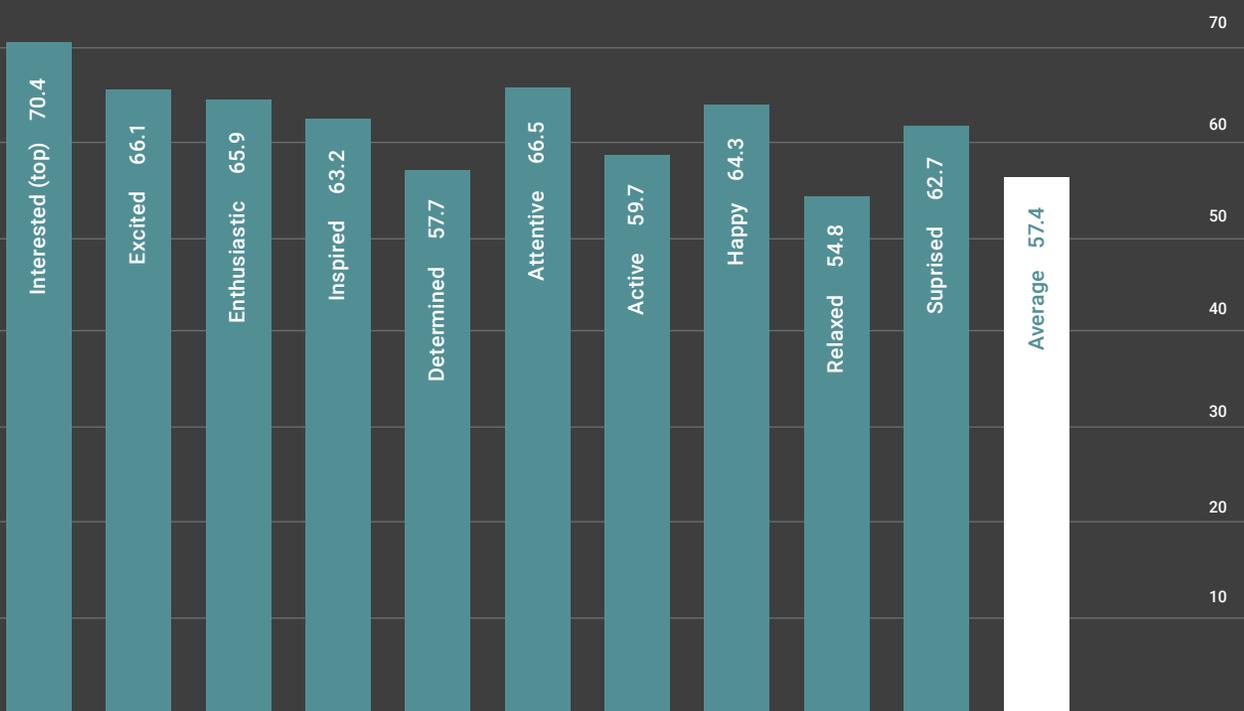
(During the experience I felt... Much more (100), About the same (50), Much less (0) than before)



Content B: Participant ratings of the moods and emotions intended by content creators to increase

Content B ●

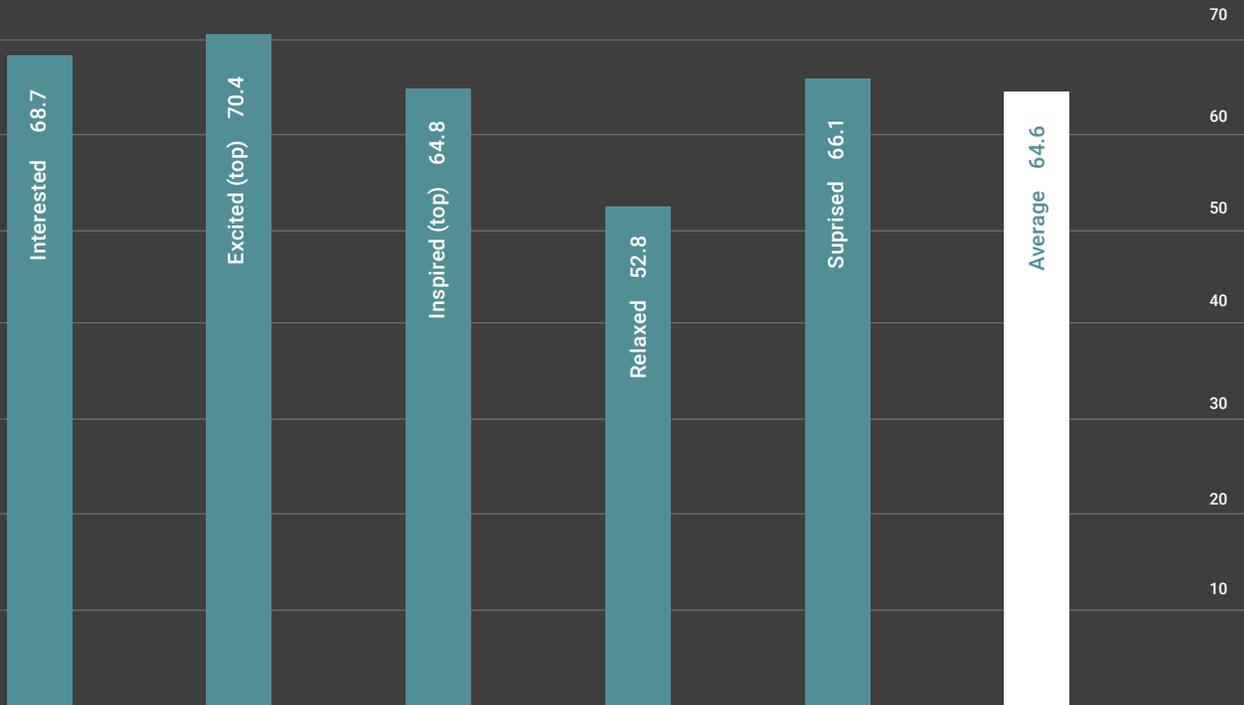
(During the experience I felt... Much more (100), About the same (50), Much less (0) than before)



Content Creator A: ratings obtained for emotions intended to increase

(During the experience I felt... Much more (100), About the same (50), Much less (0) than before)

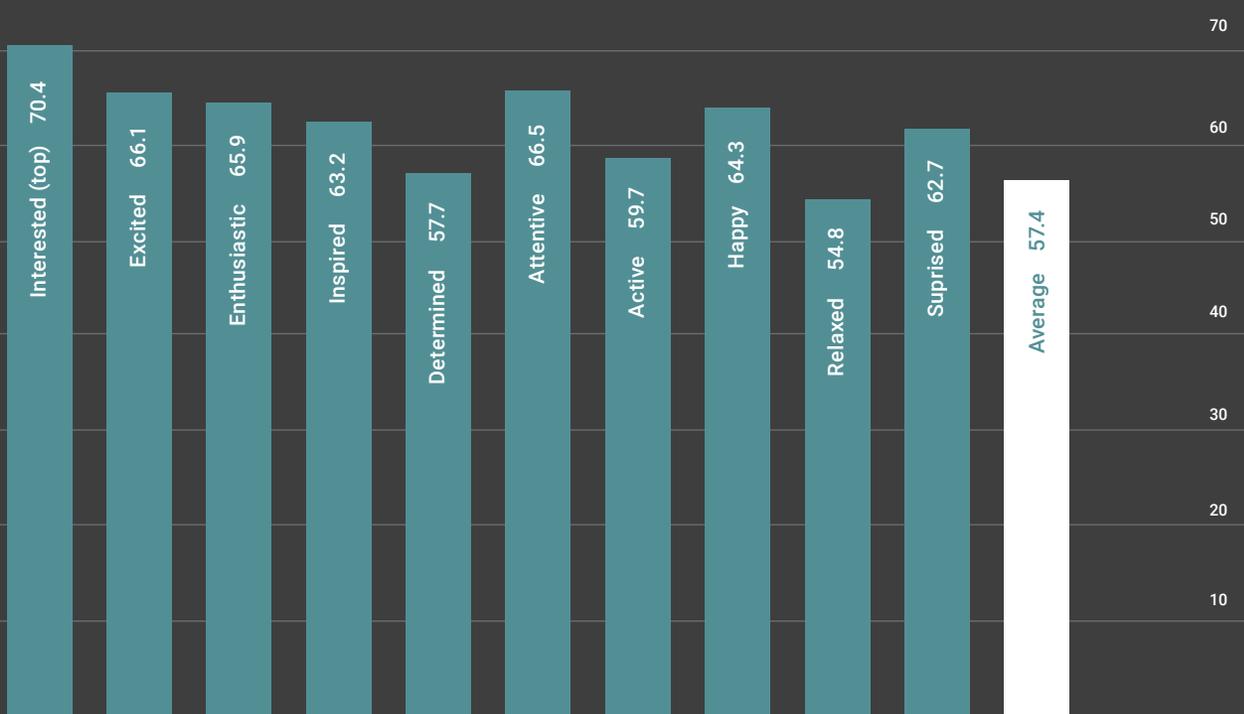
Content A



Content B: Participant ratings of the moods and emotions intended by content creators to increase

(During the experience I felt... Much more (100), About the same (50), Much less (0) than before)

Content B



KEY PREDICTORS OF IMPACT IN IMMERSIVE EXPERIENCES

The data were further explored to address the questions: Which of the more detailed experiential components influenced people's overall global impact judgements (which we consider a proxy for impactful throughout this study)?

Can we predict what experiential qualities makes content Good, for instance? Of course, people will have different tastes and genre and quality preferences, but are there consistent generalities across pieces of immersive content that more strongly or significantly predict impact?

In Section 4, two dimensions of overall impact were derived from averaging the global experiential and cultural value indicators, respectively (see Figure 5.1 'Mean totals for (a) global experiential quality (b) cultural value indicators and (c) global experiential and cultural value totals combined'). As shown in Figure 3, the scores for these two dimensions (global experiential and cultural value) were combined to provide an overall impact total.

- Global Experiential Total +
- Cultural Value Total
- = Total Impact
(Global Experience + Cultural Value).

A. Overall: relationships between global impact (experiential quality, culture value) and components of experience

In this section, we use regression analysis to understand what the relationship is between the individual components of the experience that users identified and the global impact, as defined by the mean of the experiential quality and cultural value scores.

Correlations and multiple regressions were conducted to identify the most significant predictors of global experiential and cultural value. The following potential predictors were entered into these analyses: Sense of Physical Space, Engagement, Ecological Validity (i.e. perceived naturalness and believability), Negative Effects, Positive Affect, Negative Affect, and Unusual Perceptual Experiences.

There were several significant correlations between the target impact outcome variables and the experiential qualities. Experiential qualities consistently associated with significant ($p < 0.001$) increases in impact ratings were: Sense of Physical Space, Engagement, Ecological Validity, Positive Affect, and Unusual Perceptual Experiences.

Negative Affect (comprising adjectives such as Distress, Upset, Guilty) was consistently associated with significant decreases ($p < 0.01$) in impact measures.

Positive Affect and Engagement were found to be the strongest correlates of global experiential total with Unusual Perceptual Experiences close behind (see Table 7 below).

	POSITIVE AFFECT	NEGATIVE AFFECT	SENSE OF PHYSICAL SPACE	ENGAGEMENT	ECOLOGICAL VALIDITY	NEGATIVE EFFECTS	UNUSUAL PERCEPTUAL EXPERIENCES
Global Experiential Quality Total	.743**	-.317**	.558**	.764**	.531**	-.09	.655**
Cultural Value Total	.724**	-.335**	.495**	.722**	.498**	.062	.618**
Total Impact	.748**	-.333**	.535*	.756**	.524**	-.076	.648**

The regression analysis reveals that all predictors considered together explained 68.5% of variation in global experiential - ($p < .001$). The significant predictors were Positive Affect (.353, $p < .001$), Engagement (.337, $p < .001$), and Unusual Perceptual Experiences (.201, $p = .002$).

A separate regression analysis using Worth Paying For (as a proxy for economic value), rather than experiential value as an outcome variable revealed that all predictors explained 48.8% of the variation in Worth Paying For. The significant predictor was Positive Affect (.325, $p < .001$).

We checked for the consistency of these regression findings by considering each trial (first trial, second trial) separately as well as combined, which provides more nuanced results.

B. Per Content: relationships between global impact (experiential quality, culture value) and components of Experience

In this section, we conduct the same analysis as in sub-section a., however we look at the findings for the individual pieces of content, rather than the three pieces of content combined. This analysis is conducted to establish whether there are different relationships between impact and the specific components of experience at the level of individual pieces of content.

Multiple regressions were run for each content (A, B, C) separately, using the same experiential predictors as described on previous page in (a). As with the overall results reported in (a), for each analysis the combined trial 1 and 2 data were used first, then explored independently to check for

consistency in the regression results. The findings show there is some variation by content in the significant predictors of impact by individual piece of content.

For Content A, Positive Affect was the only significant predictor of impact when the two trials were combined, although for trial 1 data Engagement was also found to be significant. For Content B Positive Affect, Engagement and Sense of Physical Space all emerged as significant indicators of impact. However, again there was some variation in the individual trials – for trial 1 only Engagement and Sense of Physical Space were significant, and for trial 2 only Engagement was found to be a significant predictor of impact.

Content C was found to have two significant predictors when the two trials were combined – Positive Affect and Unusual Perceptual Experiences. For trial 1, no significant predictors were found, however for the trial 2 data, Unusual Perceptual Experiences was found to be a significant predictor of impact.

More detail on the individual predictors by content trial and level of variance predicted by each indicator is available in appendix A. In terms of economic value, Positive Affect emerged as the significant predictor of Worth Paying For in the regression models for Content A and B (and across all contents considered together, as reported above). Engagement significantly predicted Worth Paying For in Content B only. Finally, Unusual Perceptual Experiences significantly predicted Worth Paying For in Content C.

TOOLKIT DEVELOPMENT

Alongside testing the research protocol formally on three pieces of content, a toolkit has been developed from the research, to allow content creators and immersive sector stakeholders to proactively develop their own evaluations of immersive experiences.

The toolkit can be used formatively to help guide the development of impactful experiences. It can also be used to build a bank of evaluations against which any piece of content can be benchmarked (compared with any other content previously evaluated, and compared more specifically within genre, or with different affordances). This will create a potentially significant resource in predicting the likely experiential impact and cultural value of the content as it is developed. Despite the use cases for the toolkit, there remain several limitations which may be explored in a further study. The research has been unable to test more fully and rigorously the influences of personality (trait) and other person variables on their user evaluations, such as their openness to experience; the impact of novelty; the impact of particular content affordances, such as the level of interactivity within the individual piece of content.

New scoring methods (see Potential ideas below) and the development of more bespoke forms of the toolkit will build on the findings that Engagement and Positive Affect are important determinants of at least experiential and cultural value. With a wider range of content and users, these research questions can be investigated. In particular, we see a real opportunity to test

the toolkit and research protocol across a wider set of immersive formats and genres. Applying the toolkit to the more developed VR gaming content available is another possible avenue of exploration. As mentioned earlier in the report, other immersive technologies such as augmented reality could also be examined.

The components of the final toolkit specify the most significant aspects of experience to measure impact of a given piece of immersive content, based on our testing of the contents in the present study. The corresponding questions and scoring of these subjective aspects are also provided in the toolkit. This is not to say that the other items are not valuable – with a wider range of content, different determinants may be found. The full set of items tested in the study will continue to be further explored for a more comprehensive and accurate evaluation of a content's impact at any given point in time (and indeed, its impact 'trace' over time, as new content is produced for this market).

We have been able to categorise the data into 9 sections that make up the UX toolkit. These are outlined below. The toolkit allows the content creator to apply the same questions as outlined within this study to review their own immersive content.

Audience quality of experience	Audience attitudes (content & tech)	Creator's intended impacts
Audience characteristics	Interaction affordances	Economic Impacts
Audience behaviours	Audience traits (innersive tendencies)	Stakeholder target impacts

Potential ideas for development

New Scoring Methods: With further data exploration and testing, the toolkit could begin to bank and store the data from user trials with those tested contents. Contents could be evaluated not only in relation to their internal validity i.e. against themselves, but in terms of comparative qualities with other content for which test data has been banked in the database. Individual pieces of content could be tagged for their affordances and/or genre to cluster them.

- (a) over-index total score (i.e., summing the breadth and depth of the increase in scores from average),
- (b) under-index total score (i.e., summing the breadth and depth of the decrease in scores from average); and
- (c) the trade-off between over and under indexing: the differential, and whether it's swings towards being over or under index, overall.

Some examples are given below of the over-index scoring described above, using current data.

When one piece of content is compared with others and more importantly combined contents (from which we can also classify characteristics such as their affordances and genres), cross-content averages (an index) can be more reliably computed. The different ways our current data are explored below show how the results can be viewed under different lenses for implementation of the toolkit in the future. This type of analysis could yield the following useful information for content creators (see Figure 16 for an example illustrating this scoring method compared with the one presented in this report):

In Figure 16 we provide an example of this index using dummy data. The grey columns on the right represent the average score of the total number of contents tested within the toolkit, while the left hand blue column represents the piece of content being indexed against the average. As can be seen, for the question of how 'powerful' is the content, the content is 7 points over-index, while for 'worth paying for' it is 4 points below the index. The total index differential for this dummy piece of content would be +3 (sum of the strengths minus sum of the weaknesses). In this way, we could consider both the overall score of a piece of content and the individual facets in comparison to a group benchmark.

Figure 16

Example of how (over/under) indexing (average score) could be used as a toolkit metric to compare impact of immersive content

Example of over/under indexing compared with reported standard scoring



Index of the global qualities of content

Based on the index method explained above, we now consider Content A, B and C using this approach. Compared with Contents B and C, Content A is over-index and scores highest of all the contents for Impressive, Recommendable, and Memorable. Content B is only very marginally most Worth Paying For, whilst Content C is associated with the greatest number of highest global ratings. Of all the contents, Content C is considered the most Good, Powerful, Satisfying, Rewarding, Pleasurable, Emotionally Moving (Content C > Content B, $p < 0.05$), and Transporting. Content C consequently has the highest total global quality score.

To consider these data another way, if the value by which any given quality is over-index, we find that Content A is characterised by ratings of being Emotionally Moving (around 5 points above all content average), Memorable (by around 3 points above average), and Recommendable, Powerful and Impressive with slightly smaller over-index margins. When these differences from the mean are summed, it gives a total over index of around 10. For Content B, with only one quality (Worth Paying For) rated just marginally higher than average (0.1), it's total over-index score is just 0.1. In stark contrast, Content C has the highest total over-index of around 25, marked by over-indexes on nine different qualities, particularly Emotionally Moving (by around 9 points), Powerful (almost 5 points above average), Rewarding (nearly 3 points above average), and Transporting (just over 2 points higher).

Index of the cultural value qualities of the content

When looking at content profiles relative to each other and the index for the whole sample, interesting patterns emerge in the cultural value data.

Content A is over-index on five cultural value qualities and scores higher than either of the other contents on only being Well Produced and

Presented. It also shares prime position with Content C for wanting to Seek Similar Content. Content B is rated more highly compared with both other contents in being Able to Hold Attention, and is considered Better as Social, providing content creators with useful feedback about desired affordances. Finally, compared with either Content A or B, Content C scores highest for 11 of the 14 cultural value qualities: Interesting Idea, Different, Thought Provoking (Content C > Content B, $p < 0.05$), Like Repeat Experience, Says Something About the World, Well Constructed, Intellectually Engaging, offering New Perspectives, Sharable, Seek Similar Content (=Content A), and Stimulated my Creativity.

The over-index difference scores reflect this pattern, showing that Content C has the highest over-index difference of nearly 40 points, followed by Content A (over-index difference total of around 7) and finally Content B with a total over-index score of 1.2 (Better as Social was not included as it suggests the Content could be better designed, and should more correctly be reverse scored). Index of the psychological indicators of the content.

There are marginal, not statistically significant, differences in ratings of presence across the three pieces of content: Content A receives the highest scores on Sense of Physical Space and Negative Effects (=with Content C); Content B scores highest on Engagement, and Content C scores highest on Ecological Validity and joint highest on Negative Effects (=Content A). The index differentials concur with these findings, with only small and identical increments (0.2) for each content compared with index. This suggests that the presence stimulated by the formal properties of the medium (technological immersion) is similar across contents, such that quality and impact ratings go beyond presence.

CONCLUSION

This paper reports the findings of an experiment to assess UX impacts with regards to immersive experiences. We aim to examine impact related to whether immersive experiences provide social, cultural, and psychological impacts.

We also explore the likely future economic impact of immersive content through valuation techniques related to individual pieces of content. As the market for immersive content diversifies in all forms, it seems likely that forms of audience evaluation will proliferate, as they have done across other forms of media. While many content developers are still in the early stages of development with these technology forms, there has been little prior research looking at formal audience evaluation in this field.

Despite the lack of specific studies about immersive experiences of VR on audiences, there are adjacent areas of both academic and industry research that are significant for the field. Notably, there is a widespread literature on cultural value which relates across to virtual or digital experiences, and there is a range of psychological literature on the effects of video games, film and other media forms that consider the 'immersion' of the participant. This study has shown that the practical application of various measures drawn from these fields are relevant and applicable to immersive content.

In terms of the specific findings, we find encouraging evidence that the different components of the research protocol elicited nuanced responses from the user group. Similarly, the data collected on users' willingness to pay for immersive experiences in different settings has a strong level of validity with what we currently know about markets for VR consumption in the wider economy.

Content creators are also closely aligned to users in terms of the impact they intend to generate from their experiences. In detailing which measures they expected to increase (and in one instance, decrease), the content creators were tended, in the main, to agree with what users experienced.

In examining what drives the overall perception of positive 'impact' on audiences, we report early evidence that two sets of questions are particularly strong predictors of quality for the user. These are the set of metrics on Positive Affects, and the set of metrics related to Engagement. Interestingly, even within quite a small range of VR content, there were several other predictors, such as Unusual Perceptual Experiences, that also drove impact. It is worth noting however that the trial was limited in size, meaning that the findings here should be read as tentative, rather than conclusive, with regards to what drives the overall impact for the user.

Conclusion *cont.*

The set of measures used in this study is lengthy for a user trial, however our findings give encouragement to the fact that the set of questions and scales used in the study were sensitive to different content types, albeit through a relatively small study of three pieces of content. A shortened version of the toolkit (including only main items for each of the scales we included in the initial toolkit) should be similarly sensitive. There are several items within the research protocol that did not change between the three content types tested, however we would recommend testing a larger range of content before removing them. This is because very different content types – for example genres such as horror, or shocking content – might be expected to score very differently on these measures.

The immersive UX toolkit and its future usages

The shorter toolkit for evaluating immersive user experiences will be made available through Digital Catapult and i2 Media Research. It contains a condensed version of the research protocol for content creators, along with a set of guidelines on how to implement a UX trial.

We see two primary uses for the toolkit, moving forward from this study. Firstly, there is an opportunity for content developers to

use it formatively during the development of immersive experiences, using the tool with users or audiences during beta phases of their projects. This would allow creators to fine-tune some of the effects they wished to elicit on their primary audiences in a more formal way than might currently be taking place in the market.

Secondly, there is scope to build a bank of evaluations against which any piece of content can be benchmarked (compared with any other content previously evaluated, and compared more specifically within genre, with different affordances and other variables). This would allow for content creators to test both the internal validity against their own assumptions about the content, and externally compare their content against what is currently available on the market. The goal of this is not to determine the best or most audience-friendly piece of content, but to allow those creating immersive experiences to examine specific facets of the audience experience for their work in comparison to previously released content. The toolkit approach could also prove valuable in identifying impactful audience experiences across a wider bank of content, leading to trend-spotting with regards to promising emerging genres, styles and types of content.

APPENDICES

Table 10.1

Relationships between global impact (experiential quality, culture value) and components of Experience In the table below the significant predictors of impact are analysed by individual content type

SIGNIFICANT PREDICTORS OF IMPACT BY INDIVIDUAL CONTENT			
Content	Combined results	Trial 1 results	Trial 2 results
A	Positive Affect (.464, $p > .001$) 75.8% of the variance ($p > .001$)	Positive Affect (.404, $p = .022$) Engagement (.572, $p = .037$) 57.3% of the variance ($p = .002$)	Positive Affect (.499, $p = .003$) 82.9% of the variance ($p > .001$)
B	Positive Affect (.285, $p = .01$) Engagement (.493, $p > .001$) Sense of Physical Space (.238, $p = .008$) 73% of the variance ($p > .001$)	Engagement (.557, $p = .001$) Sense of Physical Space (.339, $p = .025$) 63.6% of the variance ($p > .001$)	Engagement remained significant (.522, $p = .006$) 68.4% of the variance ($p > .001$)
C	Positive Affect (.377, $p = .006$) Unusual Perceptual Experiences (.362, $p = .007$) 59.8% of the variance ($p > .001$)	No individual predictors were found to be statistically significant. 43.3% of the variance ($p = .009$)	Unusual Perceptual Experiences (.706, $p > .001$) 84.1% variation in impact ($p > .001$)

10: REFERENCES

Bailey, P., Garnham, R. (2017). Shakespeare still has power to shock. Ipsos MORI.

Bakshi, H. (2012). Measuring Cultural Value: Keynote speech delivered at Culture Count: Measuring Cultural Value Forum, Customs House, Sydney, Australia, Tuesday 20th March 2012, Nesta.

Bakshi, H., Dolan, P., Fujiwara, D., Lawton, R., Mourato, S. (2015). Measuring Economic Value in Cultural Institutions: A report commissioned by the Arts and Humanities Research Council's Cultural Value Project. AHRC.

Baños, R. M., Botella, C., Rubió, I., Quero, S., García-Palacios, A., & Alcañiz, M. (2008). Presence and emotions in virtual environments: The influence of stereoscopy. *CyberPsychology & Behavior*, 11(1), 1-8.

Barger, V., Barger, V., Peltier, J. W., Peltier, J. W., Schultz, D. E., & Schultz, D. E. (2016). Social media and consumer engagement: a review and research agenda. *Journal of Research in Interactive Marketing*, 10(4), 268-287.

Blachnio, A., Przepiorka, A., & Rudnicka, P. (2016). Narcissism and self-esteem as predictors of dimensions of Facebook use. *Personality and Individual Differences*, 90, 296-301.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., ... & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178-192.

Brown, A., Carnwath, J.D. (2014). Understanding the Value and Impacts of Cultural Experience. Arts Council England.

Crossick, G., Kaszynska, P. (2016). Understanding the value of arts & culture: The AHRC Cultural Value Project. AHRC.

Danielson, K., Jenkins, J., Phillips, M., Jensen, E. (2015). Cheltenham Festivals: Real-time Event Feedback. Digital R&D Fund for the Arts Final Report. Nesta.

Dittrich, A. (1998). The Standardized Psychometric Assessment of Altered States of Consciousness (ASCs) in Humans. *Pharmacopsychiatry*, 31, 80-84.

Edwards, J., Rae, J. (2015). Virtual reality at the British Museum: What is the value of virtual reality environments for learning by children and young people, schools, and families?. Museums and the Web Conference 2016.

S. Engeser and F. Rheinberg. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32, 158-172.

Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & Jsselsstein, W. (2000). Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators and virtual environments*, 9(2), 149-164

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PloS one*, 11(8), e0159915.

Hu, J., Janse, M., & Kong, H. J. (2005). User experience evaluation of a distributed interactive movie. In *HCI International*. (http://www.idemployee.ie.tue.nl/hu/publications/HCI2005_jeccream.pdf)

Fiennes, T. (2017). Putting audiences at the heart of VR. BBC internet Blog. Friday 21 July 2017. <http://www.bbc.co.uk/blogs/internet/entries/c438a2cd-fcd4-42f3-ab69-244d3c579011>

Knell, J., Whitaker, A. (2016). Quality Metrics Final Report: Quality Metrics National Test. Arts Council England.

Kunter, M., (2016). The Van Westendorp Price-Sensitivity Meter As A Direct Measure Of Willingness-To-Pay.

Lyon, D. W. (2002). The Price is Right (or is it?). *Marketing Research*, 14(4), 8-15.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and virtual environments*, 10(3), 282-297.

McNair, D. M. (1971). Manual profile of mood states. Educational & Industrial testing service.

Ravaja, N., Salminen, M., Holopainen, J., Saari, T., Laarni, J., & Järvinen, A. (2004, October). Emotional response patterns and sense of presence during video games: Potential criterion variables for game design. In *Proceedings of the third Nordic conference on Human-computer interaction (339-347)*. ACM.

Suzuki, K., Roseboom, W., Schwartzman, D. J., & Seth, A. K. (2017). The Hallucination Machine: A Deep-Dream VR platform for Studying the Phenomenology of Visual Hallucinations. *bioRxiv*, 213751.

Unknown Author (2017). How virtual reality facilitates social connection. Facebook and Neurons Inc.

Vorderer, P., Wirth, W., Saari, T., Gouveia, F. R., Biocca, F., Jäncke, L., ... & Klimmt, C. (2004). Development of the MEC spatial presence questionnaire (MEC-SPQ). Report to the European Community, Project Presence, MEC.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063.

11. FOOTNOTES

1. Digital Catapult (2017), Request for Proposal: The Supply of Research and Consultancy to the Digital Catapult. Contract Reference: CON-BEI-WP10/2
2. Nesta, Mapping the Immersive Economy (Forthcoming 2018).
3. Innovate UK, (2017) Industrial Strategy Challenge Fund – more challenges, more opportunities. See, <https://innovateuk.blog.gov.uk/2017/11/30/industrial-strategy-challenge-fund-more-challenges-more-opportunities/>
4. YouGov (2017), VR headsets more popular than tablets and wearables were at same stage. See, <https://yougov.co.uk/news/2017/05/19/vr-headsets-more-popular-tablets-and-wearables-wer/>
5. CNN Travel, (2017) How VR Theme Parks are Changing Entertainment in Japan. See, <https://edition.cnn.com/travel/article/vr-parks-on-japan/index.html>
6. Digital Catapult have commissioned research into emerging content forms by the consultancy Limina, which will be published in Spring 2018.
7. Mashable,(2017), Google's 'Pearl' is first VR film to be nominated for an Oscar. See, <https://mashable.com/2017/01/24/google-pearl-vr-oscar-nomination/#UyPmRsCqaqK>
8. See, for instance, Venice Virtual Reality at Lazzaretto Vecchio Island: <http://www.labiennale.org/en/cinema/2017/venice-vr>
9. Digital Catapult (2017), Request for Proposal: The Supply of Research and Consultancy to the Digital Catapult. Contract Reference: CON-BEI-WP10/2
10. Digital Catapult (2017), Request for Proposal: The Supply of Research and Consultancy to the Digital Catapult. Contract Reference: CON-BEI-WP10/2
11. Note: An 'item' here refers to a specific question, often drawn from a larger bank of questions related to a subject matter.
12. DCMS, (2016) Taking Part Adult Questionnaire 2015/16.
13. Knell, J., Whitaker, A., (2016), Quality Metrics Final Report: Quality Metrics National Test.
14. Nesta (2010), Beyond Live: Digital Innovation in the Performing Arts.
15. Brown, A., Carnwath, J.D. (2014). Understanding the Value and Impacts of Cultural Experience. Arts Council England.
16. Knell, J., Whitaker, A. (2016). Quality Metrics Final Report: Quality Metrics National Test. Arts Council England.
18. Kunter, M., (2016) The Van Westendorp Price-Sensitivity Meter As A Direct Measure Of Willingness-To-Pay.
19. Fiennes, T., (2017) Putting Audiences at the Heart of VR. See, <http://www.bbc.co.uk/blogs/internet/entries/c438a2cd-fcd4-42f3-ab69-244d3c579011>
20. DCMS, Taking Part Survey: England Adult Report 2016/17 (2017).
21. As discussed in table 1, Global, in this context, refers to a simple and potentially universal set of indicators that could be considered in relation to whether a piece of content is impactful or not. An example indicator is whether content is 'recommendable' or 'not recommendable' for an individual.
22. The global experiential quality ratings were: Good, Powerful, Recommendable, Memorable, Satisfying, Rewarding, Useful, Pleasurable, Emotionally Moving Transported me elsewhere and Worth paying for. For each of these, their inverse was also presented to the trial participant.
23. Statistical analysis (T-test, comparing means to value of 50) confirms significant differences for all emotions (p < 0.05 or lower) with the only exceptions of Relaxed for Content A and B. In line with this finding is the fact that the top scoring emotion (Excited for Content A, Interested for Content B and Inspired for C) is always present among content creators' intentions.
25. The trials were analysed separately (e.g. trial 1 responses, then trial 2 only responses). Trial 1 only results showed a decrease this time in explained variance: 49.3% variance (p > .001). This variance was significantly predicted by Positive Affect (.323, p = .002) and Engagement (.392, p = .002). When trial 2 data was considered separately, 75.3% in impact variation across all Content was significantly explained (p > .001). Positive Affect (.361, p > .001), Engagement (.262, p = .016) and Unusual Perceptual Experiences (.332, p = .001) were demonstrated to be significant predictors.

ACKNOWLEDGEMENTS

Many thanks go to all of the focus group participants and interviewees at the outset of the project. Our thanks also go to all of the participants that took part in the lab trials at i2 media research's facilities at Goldsmiths University between December 2017-January 2018, and the companies who kindly provided their content for testing.



Nesta is a global innovation foundation. We back new ideas to tackle the big challenges of our time. We use our knowledge, networks, funding and skills - working in partnership with others, including governments, businesses and charities. We are a UK charity but work all over the world, supported by a financial endowment.



i2 Media Research is the expert consumer insight and user experience research and strategy consultancy based at Goldsmiths University of London, delivering rigorous research and innovation to business - since 2002. The research base on which i2 was founded was on user experience of immersive media, starting in the 1990s. Fully self-funded through commissioned projects and through collaborative R&D, i2 media has successfully completed more than 250 commissions for clients spanning the commercial, public and third sectors.



Digital Catapult is the UK's leading advanced digital technology innovation Centre, driving early adoption of technologies to make UK businesses more competitive and productive to grow the country's economy.

We connect large established companies, startup and scaleup businesses and researchers to discover new ways to solve big challenges in the manufacturing and creative industries. Through this collaboration businesses are supported to develop the right technologies to solve problems, increase productivity and open up new markets faster.

Digital Catapult provides physical and digital facilities for experimentation and testing that would otherwise not be accessible for smaller companies.

As well as breaking down barriers to technology adoption for startups and scaleups, our work de-risks innovation for large enterprises and uncovers new commercial applications in immersive, future networks, and artificial intelligence technologies.

CATAPULT
Digital

Digital Catapult
101 Euston Road
London
NW1 2RA
0300 1233 101
www.digicatapult.org.uk

