International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: http://www.tandfonline.com/loi/hihc20

Lost at the Edge of Uncertainty: Measuring Player Uncertainty in Digital Games

Christopher Power, Paul Cairns, Alena Denisova, Themis Papaioannou & Ruth Gultrom

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2018.1507161

Lost at the Edge of Uncertainty: Measuring Player Uncertainty in Digital Games

Christopher Power [a], Paul Cairns[a], Alena Denisova[b], Themis Papaioannou[a], and Ruth Gultrom[a]

[a]Department of Computer Science, Deramore Lane, Heslington, University of York, York, UK; [b]Computer Science, Swansea University, Swansea, UK

ABSTRACT

Uncertainty has previously been identified as an important ingredient of engaging games. Design in games can create different levels of uncertainty in players that they can recognize and describe as being either attributable to external forces, such as chance or hidden information, or internal to their own understanding of what to do in relation to their own goals. While it appears that uncertainty can contribute both positive and negative play experiences, there is little work in trying to operationalize and measure this concept as a component of player experience. Reported in this article is an analysis of data from over 700 players using modern bi-factor analysis techniques resulting in a five factor psychometric scale which captures the broad feelings of players about uncertainty in games. Three of these specific factors appear to point toward a single generic factor of uncertainty that is internal to the players, one captures experiences relating to external uncertainty, with the final factor relating to player's experience of exploring the game to resolve uncertainty. In order to further validate the scale, we conducted an experiment with a commercial puzzle game manipulating the duration of play with predicted outcomes on the different specific factors of the scale. Overall the scale shows promise with good statistical reliability and construct validity of the separate factors and so will be a useful tool for further investigating player experiences in digital games.

1 Introduction

In digital games, players seek a wide range of different experiences. At times, it is to reach out and be part of somewhere else either physically or socially (Sherry, Lucas, Greenberg, & Lachlan, 2006; Yee, 2006). In other cases, it is to lose themselves into a digital world where the everyday troubles fadeaway (Collins & Cox, 2014). Sometimes, it is just about extracting pleasure one crushed candy at a time in 10-minute bites of joy (Juul, 2010). Similarly, developers setout to design games that evoke these experiences in players. They create diverse levels, invest in powerful narratives and hone game mechanics, often through hours of play testing, to ensure that their players keep playing.

Callois (Caillois & Barash, 1961) argues that the outcome of every type of play should be uncertain in order to be enjoyed, else if one is well trained or has the skills to defeat it with certainty it is no longer pleasing. Similarly, Malone (1982) discusses the relation between challenge and uncertainty in games: he argues that challenge is an essential element of an enjoyable game play and for an activity to be challenging, it needs to have an uncertain outcome relative to a goal, either globally regarding the outcome of the entire game, or more localized goals within the gameplay.

In fact, uncertainty and the mastering of uncertainty may actually be central to the appeal of games because uncertainty also lies in other game elements, including the journey the player follows through the game or the problem-solving skills that they might require to make progress (Costikyan, 2013). What is appealing in games is that players have the choice to engage in an unknown experience with no certain outcome, or at least unknown outcome, in order to test their abilities or skills without worrying about the consequences that might occur in real-life situations (Juul, 2011). However, uncertainty is not an experience that should be viewed as an end in itself. When well balanced, Costikyan proposes, uncertainty can lead to enjoyable positive experiences. However, when players are tipped into having uncertainty that is too high, this could lead to negative experiences.

Despite the apparent importance of uncertainty for games to actually be games, there is relatively little research specifically into the nature of uncertainty in games. This may be because it is viewed as a feature of games themselves, for example through randomness in the delivery of upgrades or buffs, and it is such sources of uncertainty that Costikyan (2013) describes. However, we argue that uncertainty might be better viewed as a player experience rather than an objective measure of the game itself. This has intuitive appeal. What may seem like unmanageable uncertainty to a novice player is well understood by a more experienced player. For example, in the turn-based strategy game *Polytopia* (Midjiwan, 2016), the player starts in an essentially random location with unknown proximity to the other tribes and to useful resources. This is the same for all players but more experienced players know something of what might happen and what risks are present. Despite the equivalence of game

CONTACT Christopher Power christopher.power@york.ac.uk
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

states for all players, the felt uncertainty is considerably less for the experienced player.

Also, there are clearly different levels of felt uncertainty for the same player at different times. For some parts of a game, like *Skyrim* (Bethesda Game Studios, 2011), part of the map may be well understood in terms of resources it offers, enemies that might be encountered and so on, whereas another part could be uncharted for that player and essentially a journey into the unknown. The experience of uncertainty would be very different even though the essential aspects of the game and its gameplay are well understood.

Existing types of player experience have been extensively studied, for instance, engagement (Brockmyer et al., 2009; Wiebe, Lamb, Hardy, & Sharek, 2014), immersion (Cairns, Cox, & Nordin, 2014), flow (Chen, 2007) and fun (Lazzaro, 2009). In some sense these are *goal experiences*. These are the experiences that players are seeking to have when they play games (Cairns, 2016). Felt uncertainty, however, is not necessarily a goal of play but is more like a *foundational experience* on which these other experiences are built. Players do not play in order to feel uncertain but it would not be much of a game if they did not feel uncertain at some point.

In order to better understand the felt experience of uncertainty, there is a need to be able to capture what constitutes this experience. Drawing on the limited literature in games research, this article describes the work of developing a questionnaire for examining uncertainty in games. Validation and analysis of the questionnaire, and a subsequent experiment establishing construct validity, suggests that felt uncertainty in games has five constituent factors but that three of these are strongly related to the sense of knowing what is going on in the game. The other two are exploration (EXP) and external uncertainty (EXU): reducing what is uncertain and the limitations of what can be made certain. As well as providing an operational definition of felt uncertainty, the questionnaire offers a tool for examining uncertainty in games more deeply in future research and opens up the opportunity to see how such foundational experiences lead to the emergence of goal experiences.

2 Uncertainty in Players

Despite the recognition that uncertainty is essential for games to be worthwhile, there is not extensive examination of uncertainty in relation to digital games and more particularly to player experiences. Within psychology, it has been recognized that uncertainty and uncertainty in relation to rewards in particular is an important factor in influencing judgments and decision-making (Kahneman, 2011). Most often uncertainty is equated with degrees of probability, the actual chance of one thing or another thing happens. However, this is only one way of conceptualizing uncertainty. People distinguish between EXU, related to how chance events in the world might turn out, and internal uncertainty reflecting their own knowledge or more rightly their ignorance (Kahneman & Tversky, 1982). For example, whether a coin comes up heads is viewed as 50–50 ($p = 0.5$) but it is EXU because until the coin is tossed, it is in principle unknowable. By contrast, if asked which of two rivers is longer, someone may hold that whether the Nile is longer than the Amazon as 50–50 because they simply don't know, but clearly it is in principle knowable and so uncertainty is internal to the person.

However, an alternative analysis of uncertainty suggests that people actually break it down by whether the uncertainty is epistemic or aleatoric (Fox & Ülkümen, 2011). Aleatoric uncertainty corresponds well to EXU but epistemic uncertainty can vary depending on the framing that people take on a subject or problem. For example, with a movie release coming, moviegoers may say: "I am 90% sure that the next Marvel movie will be good" where the framing is epistemic about their knowledge of the movie itself, such as seen in trailers or posters. On the other hand, the statement: "The chance that the next Marvel movie is any good is 90%" has the framing as aleatoric in relation to the long running perception of the quality of Marvel movies. This framing is reflected in people's language toward uncertainty even when stated quantitatively (Ülkümen, Fox, & Malle, 2016).

Turning toward games, Salen & Zimmerman (2004, ch. 15) consider how uncertainty arises in games and attribute this primarily to random or chance events in the game. In this sense, their view of uncertainty in outcome is that it is aleatoric uncertainty though they recognize that players may treat epistemic uncertainty as aleatoric simply because a game situation is too complex to be able to know what is going on. From their perspective, uncertainty functions as chance events and games exploit that sort uncertainty to make decisions in the game more interesting. LeBlanc (2006) however notes the uncertainty may actually arise due to hidden knowledge, for example, the fog of war, and so in this sense the uncertainty of outcome driving the player through the game would be better thought of as epistemic uncertainty.

It is perhaps from these different views that Costikyan (2013) develops his analysis of uncertainty based on his extensive knowledge and experience of games. He suggests 11 sources of uncertainty within in games, that is, elements within a game, that can lead to uncertainty for players. Briefly these are the following:

- Performative uncertainty: referring to the uncertainty associated with physical performance, (e.g. creating combos in *Street Fighter 2* (CAPCOM, 1991)).
- • Solver's uncertainty: concerned with weighing a group of options against potential outcomes, (e.g. cryogenically freezing a hamster in *Day of the Tentacle* (LucasArts, 1991)).
- Player unpredictability: knowledge, or lack of knowledge, of what players will do as individuals or groups (e.g. opponent strategies in *DOTA2* (Valve Corporation, 2013)).
- Randomness: uncertainty emanating from random game elements (e.g. the Alien searching and finding you in *Alien: Isolation* (Creative Assembly, 2014)).
- Analytic complexity: referring to uncertainty that comes from complex decision trees (e.g. choosing the right *Lemmings* (DMA Design, 1991)).
- Hidden information: uncertainty that emanates from hidden game elements (e.g. hidden treasure until

synchronization in *Assassin's Creed IV: Black Flag* (Ubisoft Montreal, 2013)).

- Narrative anticipation: uncertainty related to the path or the sequence of events (e.g. choosing if Hawke dies in *Dragon Age: Inquisition* (Bioware Edmonton, 2014)).
- Uncertainty of perception: players being able to get an overview or prospect on the overall situation in the game (e.g. tracking hundreds of units in *Starcraft II* (Blizzard Entertainment, 2010)).
- Malaby's semiotic contingency (Malaby, 2007): refers to the unpredictability of a meaning that accompanies attempts to interpret a game's outcome, (e.g. what does the end of *Mass Effect 3* (Bioware, 2012) mean if Shepard was indoctrinated?).

Costikyan also discusses *development anticipation*, which relates to knowing when content will be updated, and *schedule uncertainty* in games that have a mandatory wait period (e.g. *Farmville* Zynga (2009)). Both of these aspects of games are important but they influence the felt uncertainty outside of the gameplay itself. We therefore consider them to be beyond the scope of our current focus on uncertainty.

It is clear that some of these sources of uncertainty are external and aleatoric, randomness in a game in particular, and some are epistemic and internal such as Solver's Uncertainty. However, some could be viewed as either. For example, is the Performative Uncertainty of crossing now in *Crossy Road* (Hipster Whale, 2014) an internal, epistemic uncertainty of whether the player has the skill or the external, aleatoric uncertainty of what will come down the road just then. However, what Costikyan is not attempting to do is to map the different sources of uncertainty onto specific feelings and so it may be that players experience these different sorts of uncertainty in the same way or that depending on how it is viewed, the uncertainty may lead to different experiences. Part of our goal is to develop the measurement tools that make it possible to distinguish the impact of different sources of uncertainty on the player experiences.

There have been some attempts to map uncertainty into player experiences. The appeal of games has been theorized to be because games offer uncertainty but in a staged and hence manageable way (Rautzenberg, 2015). More concretely, To, Safinah, Kaufman, and Hammer (2016) consider uncertainty to provoke different types of curiosity and it is curiosity that motivates players to play and to resolve the uncertainty. This fits with the formulation of curiosity as an information gap between what a person knows and what they want to know and this drives action (Loewenstein, 1994). It seems that in this context, information itself can be a reward to humans in a way that we (and other animals) are rewarded by other resources such as food (Marvin & Shohamy, 2016). Moreover, in this formulation, the information can be intrinsically rewarding in that the information does not have to have specific utility or lead to further more fundamental resources.

From all of the above perspectives on how uncertainty may influence users, there is good reason to believe that the feeling of uncertainty in games is something that is worth investigating as an experience. However, without explicit operationalization of the feeling of uncertainty, it is only possible to theorize between how elements in a game might lead to uncertainty and therefore lead to successful player experiences. With our work, we are proposing an explicit measure of uncertainty against which theories can be evaluated and linked back to the design of particular games.

3 Development of an Uncertainty Instrument

The goal of this work is to develop a questionnaire that is able to measure the felt uncertainty of people that arise when they play digital games. We follow a process typical of many instrument developments in HCI and drawing on the best practice in the field of psychometrics. Our process therefore primarily follows that of Kline (1998). The basic steps are to iteratively generate and refine questionnaire items and use statistical analysis to first determine and then validate the structure of the questionnaire data. We were fortunate to gather a large number of responses so we exploited this to provide deeper validation of the questionnaire by conducting the initial exploratory factor analysis (EFA) on half of the data and validating it with confirmatory factor analysis (CFA) on the other half.

In outline, the steps we took were to:

(1) Generate a large item pool based on the literature and knowledge of the domain
(2) Refine the items for duplication, wording and relevance
(3) Trial the items with target audience, in our case, gamers
(4) Refine further to produce the first version of the questionnaire
(5) Administer the questionnaire to a large number of participants
(6) Conduct EFA on one half of the data and to select a subset of items to form the second version of the questionnaire
(7) Validate the second version of the questionnaire against the second half of the data using CFA.

We have previously published a non-archival report on an analysis of the questionnaire data leading to the player uncertainty in games questionnaire Power, Denisova, Papaioannou, and Cairns (2017). However, the work reported here differs in two key ways. First, the analysis there was done using traditional methods of EFA (Kline, 1994). Since doing that analysis, we have learned about newer techniques in factor analysis including item response theory (IRT) (Embretson & Reise, 2013) and bi-factor analysis (Reise, 2012). These techniques more faithfully represent the underlying structure of questionnaire data and have resulted in a more nuanced analysis and a consequently different final form of the questionnaire. Second, the previous analysis was done on the whole dataset but we have conducted a split-half analysis here exploiting the size of the dataset but also to move further down the process of validating the questionnaire. Thus, the previously published work should be regarded as a naïve analysis that has been superseded by the more sophisticated analysis reported here.

Consequently, to avoid future confusion, we refer to the new questionnaire as Player Uncertainty in Games Scale (PUGS).

3.1 Item Pool Generation

The literature was reviewed for a collection of statements that related to the different types of uncertainty that players experience in games. Starting with Costiykan's work, we used nine of the sources of uncertainty that related actually playing games, as opposed to things relating to external factors. As such, Development Anticipation and Schedule Uncertainty did not contribute to the overall set of items. Added to this were items derived from the information seeking context work, in particular Kuhlthau, Heinstro¨M, and Todd (2008) and Onwuegbuzie, Jiao, and Bostick (2004).

These initial items were evaluated in interviews with two game players. These interviews revealed inconsistencies in the narrative anticipation elements derived from Costiykan, which resulted in the rewording of several items. Further, players discussed the consequence of the outcomes of play which resonated with Juul's notion of 'negotiable consequences' (Juul, 2011). Additional items were then generated from Juul's (Juul, 2013) work to ensure that no aspect of related concepts would be overlooked. These items were then evaluated in interviews with two more players, leading to small refinements.

When first constructed, the item pool had 146 items, each consisting of a statement that a player could indicate their level of agreement (e.g. "I could not choose which actions were better"). These items were first examined through several rounds of expert reviews with researchers and gamers. This process removed duplicates, while retaining those that had variations of description for the same concept (e.g. "I knew exactly was was required of me" versus "I knew what I had to do"). This review also removed items that were clearly unrelated to the concept of uncertainty such as those relating to challenge in the game. Finally, a check was done to ensure that any remaining items covered the nine sources of uncertainty discussed by Costiykan. At the end of this review process, a broad pool of 66 items were retained for the purposes of data collection.

3.2 Data Collection

An online questionnaire was prepared in Google Forms with questionnaire items rated on a 5-point Likert scale with end points of Strongly Disagree (1) to Strongly Agree (5). This questionnaire was piloted with a small set of users to ensure the agreement scales were understandable.

This questionnaire was distributed through the Steam Community Forums. Due to the potential of limiting the pool to only "hardcore" gamers on Steam, and to head off self-selection problems within that community, the questionnaire was also distributed broadly on Reddit to reach broader player audiences. Players were asked to play any game they liked for a typical play session, and then complete the questionnaire.

3.3 Participants

Seven hundred and eight responses were received from the questionnaire with 600 men, 39 women and 69 people not identifying with a particular gender or not responding. After screening responses for anomalies (e.g. straight lining of scales), there were 674 valid responses retained. The breakdown of game genre played is presented in Figure 1, with the most represented genres of role-playing games, first person shooter, strategy and simulation games. Fortunately, each of these is rich with opportunities for different sources of uncertainty for users. The other category comprised collectible card games, adventure, sport, sandbox and puzzle.

The majority of respondents were between 18 and 21 years of age (292), followed by the ages: 22–25 (148), 26–30 (117), 31–40 (103), above 41 (27), with 13 empty responses. The majority of the participants were frequent players, playing several times a week (656) and more than an hour on each game session (576).

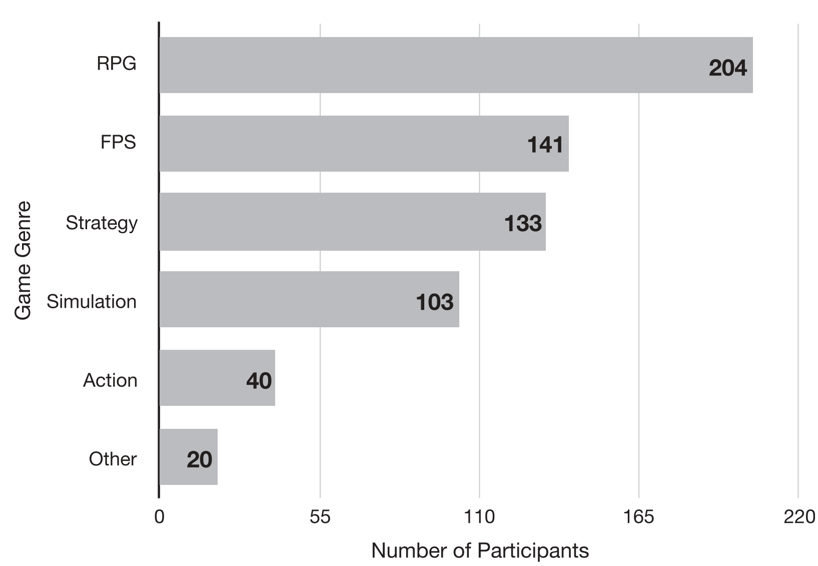

Figure 1. Distribution of genres of games played by respondents.

3.4 Introduction IRT

IRT is emerging as a successor to more traditional ways of using questionnaires in the assessment of people's traits and states, particularly in areas of health and education. It is set in contrast to classical test theory (CTT) which, as the name suggests, has been the traditional and dominant way to analyze and develop questionnaires such as presented in texts like Kline (1994) and Hair, Anderson, Tatham, and Black (1998). In both approaches, a questionnaire is intended to capture and measure the value of some underlying latent construct or set of constructs, like the feeling of uncertainty. The main difference is that, in CTT, items in the questionnaire are treated as essentially equally good in terms of capturing the value of the underlying trait across the full range of the trait whereas in IRT, items may capture some values of the trait better than others and with different sensitivity. For example, one item might on the whole be answered as strongly agree for only low levels of uncertainty but another would require a great deal of uncertainty before people would agree with it. In addition, two items might capture the same level of uncertainty but one item is more sensitive in that when that threshold is reached people immediately feel they strongly agree whereas in another, less sensitive item, people at the same threshold only feel they agree and it is at higher still levels do they begin to feel they strongly agree. There are of course important implications of this difference both for the development and interpretation of questionnaires as discussed more fully in Embretson and Reise (2000).

Originally, IRT was limited to considering only questionnaires where all items were dichotomous (yes/no answers) and the underlying construct was unidimensional, that is, a person's degree of embodying the construct was captured by a single numerical value. However, more recent developments in both mathematics and algorithms mean that it is now possible to do multidimensional item response theory (MIRT) and moreover for polytomous items, such as 5-point Likert scales. It is polytomous MIRT that is used here and in the form captured by the MIRT package in R (Chalmers, 2012).

The advantage of using MIRT over CTT is that it employs a more plausible model of the structure of questionnaire data and therefore can provide a more robust description of the questionnaire in terms of factor structures (Borsboom, 2006). Unlike CTT, because IRT treats items differently, a MIRT model is not strongly influenced by the inclusion of inappropriate items or the omission of highly relevant items. Traditional measures of suitability for factor analysis, such as the Kaiser–Meyer–Olkin Measure of Sampling Adequacy, are therefore not needed.

The disadvantage however is that there are not clear procedures that setout how MIRT changes the traditional analysis and development of questionnaires. It is still early days. In particular, in IRT, the usefulness of items is often assessed by looking at the information the item provides (in a strictly mathematical sense) across the range of values of the underlying latent concept. However, such tools are not available for higher dimensional models such as MIRT typically would produce.

Thus, in terms of selecting items, we use MIRT to produce what looks like a traditional factor structure and use the normal criteria of good loadings of items on factors. Once we have decided on which items belong to a factor, it is then possible to use item information to help assess which are the best items to represent that factor though there is a degree of circularity in that judgment so we did not use that as a primary measure of item importance. In addition, when selecting items we also consider more typical concerns for avoiding cross-loading across factors and selecting items with good loadings but very different wordings to avoid bloated specifics around particular wordings.

3.5 Results of EFA

A random sample of half of the data was extracted and EFA was undertaken to identify the potential factor structure within the 66 items. Based on the initial scree plot presented in Figure 2, we chose to explore up to six factor models for an indication of their fit to the data. A one factor model was also used to examine whether uncertainty could be considered a

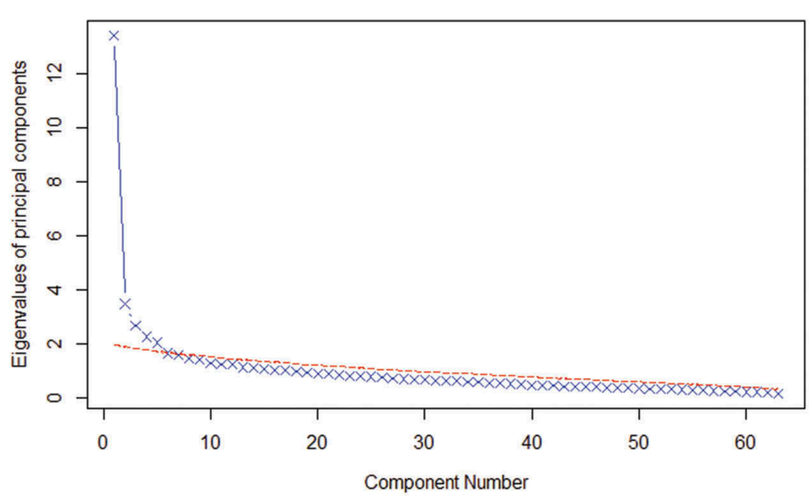

Figure 2. Scree plot for initial factor analysis. The dotted line shows the results of parallel analysis for the determination of factors.

unidimensional construct. All models were generated using the MIRT package in R (Chalmers et al., 2012). All initial solutions were rotated using direct oblimin so that factors could potentially correlate.

The one factor solution showed promise in that almost all of the items loaded on the factor with loadings above 0.3. This gives a strong indication that there is a coherent concept underpinning these items.

Looking at the best fit factors, the four, five and six factor models all showed promise in representing different aspects of this feeling of uncertainty. The four factor model had merit in that it captured four broad factors, with one dominating the others, capturing a large number of items that indicate general user disorientation. In comparison, the five factor model differentiated the single large factor into three distinct factors of tightly related statements, which is useful for working with a more nuanced set of factors. In contrast, the six factor model ended up being very fractured, with strong cross-loadings between many of the factors, which limited its explanatory power. The single factor solution and the five factor solution are presented in Table 1 with the correlation matrix for the five factors presented in Table 2.

In Table 2, we can see factors 1, 3, 4 and 5 have quite strong positive correlations, with 5 being having the weakest relationship to the other 3. Interestingly, Factor 2 is almost completely orthogonal to these other factors.

Looking at the types of statements that are contained within the factors, the following concepts seem to emerge:

Factor 1: This factor has a variety of statements related to the player evaluating their actions against potential outcomes of the game. Weighing options in relation to the outcomes, deciding their relative importance, comparing to heuristics, and the notion of weighting outcome being relatively better than another all relate to *uncertainty in decision-making* (UDM).(Sternberg & Sternberg, 2016, ch. 12).

Factor 2: This small factor of items seem to relate to players undertaking EXP of the game to discover new things to inform their decisions.

Factor 3: These statements all relate to the player *uncertainty in taking action* (UTA) in the game. Some are related to how action is actually taken within the game, whereas others are related to the players' ability to perform the actions.

Factor 4: The statements in this factor largely revolve around the player trying to understand the goals of the game and trying to decompose the different possible options available to them within the game. These statements appear to closely relate to the *uncertainty in problem solving* (UPS) processes (Sternberg & Sternberg, 2016, ch. 11).

Factor 5: These statements clearly relate to the notion of whether EXU is impacting the performance of the player. There are notions of the player perceiving the system as behaving unpredictably, discovering things through serendipity, or the situation simply being unfair.

Following typical factor analysis guidance, a loading of ±0.35 was used for an initial trim of items. When looking at the items that made this initial trim, we then removed items that largely used the same words to describe the concept of the factor. For example in Factor 4, items 11, 31 and 35 had broadly similar meaning so only item 31 was selected from the trimmed factor. This trim was supported using the notion of item information in IRT which showed that all three factors had a broadly similar information content across the range of the construct latent to Factor 4. Also, some items, though strongly loading on a factor, such as item 64 for Factor 1, because they also loaded similarly strongly on another factor, in item 64's case Factor 3.

Highlighted in Table 1 are the 25 items that have been selected for investigation in a confirmatory analysis. When looking at these 25, the majority of them appear to have strong loadings in the single factor solution, with only five dipping below the ±0.35 threshold set for the original trim.

3.6 Results of CFA

Taking the second half of the dataset, we conducted three different analyses to aid in the confirmation of the factor structure found in the first half of the data. First, a one factor structure (1F in Table 3) was examined to see the extent to which the entire structure of the questionnaire could be accounted for by a single uncertainty factor. Second, the data were fitted to the proposed five factor structure (5F in Table 3) to consider the quality of the five factor solution. Finally, a bifactor confirmatory model of a single general factor and 5F specific factors was produced (5FBF in Table 3) to see the extent to which the five factor model was independent of the general uncertainty factor.

The 5F model accounts for 47% of the variation in the data. It is clear that all items load well on their proposed five factors, with the exception of Item 1. This item, which is about strategies to achieve goals as opposed to players understanding the goals themselves, looks to be weakly related to the overall concept, possibly due to it being more metacognitive in nature.

The reliability of each factor was calculated using the more robust ω statistic rather than the usual Cronbach α (Dunn, Baguley, & Brunsden, 2014), but the interpretation of ω is the same as that of α. Traditional guidelines, for example De Vellis (2003), suggest that reliabilities of 0.7 or above are good but this is in the context of psychometric work where there is a necessary reliance on correlations between the traits of individuals. In early stages of defining and measuring concepts, lower reliabilities of 0.5–0.6 can still be useful (Nunnally, 1978), particularly where it is possible to take experimental control to help reduce measurement error, something that is not possible in psychometrics. Even so below 0.5 is considered unacceptable. Based on this, three of the five factors show good reliability, see Table 3, with EXP in the acceptable range and UPS at 0.58 is lower than would typically be preferred but still potentially useful. Overall, the indication is that the 5F model is a good description of the questionnaire data as a whole and that each separate component could be useful in capturing, at least approximately, distinct components of the experience of uncertainty.

Consideration of the 1F and the 5FBF models also show that there is a reasonable single general factor made of Factors UDM, UTA and UPS and to some extent EXU (loadings above 0.3). That is, there is good evidence for a single conceptual underpinning to the entire questionnaire. We use ω

Table 1. Exploratory analysis of a random sample of half of the data reporting the one factor solution (1F) and the five factor solution (5F).

		5F solution					
Item	1F solution	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Incl.
34. I could not predict how much time I should invest in the game.	−0.11	0.31	0.12	−0.19	−0.16	−0.05	
42. I could not say if the game had other better outcomes.	0.26	0.35	0.17	−0.03	0.03	0.17	
56. I found it difficult to form strategies in the game.	−0.12	0.32	−0.14	−0.23	0.12	0.01	
44. I was not sure what impact the game would have on me (in my life).	−0.54	0.37	0.05	−0.20	−0.14	−0.15	
53. My actions were not influencing the outcome of the game.	−0.38	0.40	−0.12	0.04	−0.06	0.11	*
49. I could not choose which actions were better.	0.32	0.42	0.12	−0.26	0.05	0.11	*
48. I could not say if the game had more than one outcomes.	0.14	0.52	0.10	0.14	0.03	0.15	*
61. I did not know how my performance influenced the outcome.	0.43	0.76	0.03	−0.02	−0.04	0.12	*
62. I did not know how the outcome(s) were connected to what I did.	0.51	0.76	−0.07	−0.09	0.00	−0.11	*
65. I knew exactly what was required from me in the game.	0.70	−0.38	0.23	−0.13	−0.29	−0.01	
60. I am well-prepared for my next game session.	0.34	−0.13	0.32	0.31	−0.23	−0.05	
32. I made a significant progress in relation to the time I spent in the game.	−0.52	0.05	0.35	0.11	−0.10	−0.15	
7. I could perform new actions with which I was not familiar.	0.49	0.01	0.38	0.16	0.13	−0.08	
4. I needed to discover things to make progress.	−0.61	−0.04	0.47	−0.03	0.26	0.00	*
22. I needed to explore in order to know what to do next.	0.61	0.21	0.49	−0.02	0.23	0.01	*
30. I was confused with what I should do next.	−0.68	0.21	0.04	−0.32	0.26	0.14	
38. I found myself going round in circles.	0.64	0.18	0.00	−0.35	−0.08	0.28	
21. I felt I was stuck during the game.	0.40	0.09	−0.08	−0.42	0.09	0.12	*
24. I thought I would fail at doing the right actions.	−0.56	−0.06	0.13	−0.44	0.10	0.04	
13. I found it difficult to keep track of all elements in the game.	−0.57	−0.02	0.03	−0.45	0.17	0.04	*
36. The game mechanics were overwhelming.	0.73	0.07	0.03	−0.48	0.11	0.02	*
43. I think what I was doing in the game was not right.	0.35	0.22	−0.02	−0.49	0.00	0.06	*
47. I was not confident that I could perform some actions in the game.	0.50	−0.04	0.04	−0.61	0.09	−0.04	*
58. The actions I had to perform were too demanding for my skills.	0.77	0.13	−0.02	−0.61	−0.02	0.06	*
46. I struggled to do the right actions.	0.62	0.02	0.01	−0.62	0.12	−0.05	*
59. I was frustrated because I knew how to achieve a goal in the game, but was unable to do so.	0.66	0.08	−0.05	−0.55	−0.20	0.18	
31. I knew how each goal could be achieved.	0.32	−0.04	−0.01	−0.04	−0.80	0.00	*
11. I knew exactly what I had to do to achieve my goals.	0.79	0.16	−0.17	0.06	−0.79	−0.09	
19. I understood the game mechanics.	−0.26	−0.16	0.10	0.09	−0.56	0.04	*
18. I felt well-prepared at the start of the game session.	−0.67	0.20	0.20	0.32	−0.46	0.00	
33. I knew how to play the game when I started.	−0.70	−0.18	−0.04	0.08	−0.45	0.20	*
1. I stuck to tried and tested strategies to complete the game tasks.	0.71	0.06	0.06	−0.17	−0.38	0.02	*
3. I was aware how my actions were influencing the outcome of the game.	0.38	−0.27	0.13	−0.04	−0.30	−0.09	
10. I found it confusing working out which actions I had to do in order to achieve a goal.	0.49	0.06	0.03	−0.30	0.36	0.07	
25. I often felt lost.	−0.53	0.17	0.13	−0.24	0.37	0.14	*
35. I knew what I had to do.	0.64	−0.12	−0.03	0.03	−0.68	0.02	
51. I could find the solutions required for achieving the goals of the game.	0.63	−0.04	0.22	0.12	−0.44	−0.05	*
37. I could not predict the outcome of my actions.	0.36	0.21	0.11	−0.19	0.05	0.35	
17. I had to repeat the same actions over and over even though I did not make any progress.	0.48	0.11	−0.05	−0.18	−0.11	0.36	
57. The game was unfair.	0.70	0.00	−0.16	−0.15	0.07	0.41	*
2. Unpredictable random elements were influencing my performance.	0.60	−0.11	0.15	−0.08	0.05	0.43	*
15. I was relying on chance in the game.	0.57	0.00	0.02	0.00	−0.05	0.73	*
28. Random elements in the game were preventing me from achieving my goal.	0.72	−0.08	−0.06	−0.13	−0.06	0.74	*
5. The outcome of my actions was mainly influenced by chance.	−0.40	0.08	−0.01	0.20	0.08	0.75	*
64. I often felt I did not know what to do next.	0.62	0.35	0.01	−0.37	0.11	−0.15	
55. I could not interpret the game mechanics.	0.45	0.30	−0.10	−0.22	0.31	0.05	
23. I did not know if my actions were the right ones to succeed.	0.72	0.00	0.36	−0.27	0.34	0.15	
40. The tasks I completed felt relevant to the general progression of the game.	−0.46	−0.16	0.43	−0.15	−0.30	−0.16	
54. My actions in the game were effective.	−0.58	−0.04	0.32	0.13	−0.40	−0.10	
20. I found it difficult to figure out what was going on in the game.	0.46	0.09	−0.01	−0.34	0.32	0.05	
6. I could think of better strategies when something went wrong.	0.36	−0.26	0.27	0.09	0.01	−0.03	
8. I did not have much information on which action was better.	−0.59	0.21	0.20	−0.10	0.17	−0.05	
9. I did not know if I had the right strategy.	0.78	0.04	0.29	−0.23	0.30	0.02	
12. I could not say if my previous actions in the game were the right ones.	0.65	0.27	0.07	−0.03	0.20	0.04	
14. I was aware of the most important aspects of the game that influenced my actions.	0.46	−0.19	0.05	0.07	−0.21	−0.10	
16. I had the required hand/eye coordination skills to achieve the goals of the game.	0.72	−0.09	0.13	0.24	−0.12	−0.19	
26. I was not sure if the outcome of the game was good.	0.55	0.15	0.08	−0.23	0.22	0.22	
27. The feedback was difficult to understand.	−0.66	0.22	0.07	−0.21	0.25	0.19	
29. I am not closer to achieving the goals of the game when I started playing.	0.72	0.09	−0.27	−0.18	0.04	0.29	
41. Things in the game I was not aware of were influencing the outcome.	0.76	0.18	0.19	−0.24	0.05	0.19	
50. I was able to complete all the goals of the game.	0.75	0.02	0.15	0.16	−0.29	−0.07	
52. I could not say what will happen at the end of the game.	−0.69	0.28	0.15	−0.03	0.15	0.07	
66. I had to think every possible way to overcome challenges.	0.08	−0.03	0.00	−0.02	−0.02	0.00	

Asterisks indicate those items included (Incl.) in the confirmatory analysis of factors.

and its variants as a measure of overall reliability of the BF model (Reise, Bonifay, & Haviland, 2013) and it shows good reliability. Indeed, the 5FBF model has an overall $\omega = 0.85$ showing that the model accounts for a very large proportion of the variance in the second half of the data. Furthermore, the general factor on its own has $\omega_H = 0.63$, and meaning that it alone accounts for 63% of the overall variance in the data and thus constitutes 72% of the variance accounted for by the 5FBF model.

Turning to the individual factors in the 5FBF though, there is a similar picture to the 1F model in that Factors EXP and EXU show substantial specific variance. This is supported by

Table 2. Correlation matrix for the five factor solution.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	1.00	−0.01	0.52	0.42	0.42
Factor 2	−0.01	1.00	−0.08	−0.07	0.05
Factor 3	0.52	−0.08	1.00	0.47	0.31
Factor 4	0.42	−0.07	0.47	1.00	0.36
Factor 5	0.42	0.05	0.31	0.36	1.00

Table 3. Confirmatory analysis on the second half of the data reporting the one factor analysis (1F), the five factor analysis (5F) and the five factor bi-factor analysis (5FBF), together with appropriate measures of reliability.

		5F					5FBF					
Item	1F Solution	UDM	EXP	UTA	UPS	EXU	General	UDM	EXP	UTA	UPS	EXU
61.	0.80	0.89					0.74	0.49				
62.	0.78	0.93					0.67	0.72				
48.	0.61	0.62					0.59	0.22				
49.	0.74	0.62					0.74	0.06				
53.	0.60	0.64					0.56	0.32				
22.	0.19		0.77				0.21		0.72			
4.	0.18		0.77				0.16		0.78			
21.	0.61			0.57			0.65			−0.08		
13.	0.58			0.56			0.61			0.04		
36.	0.64			0.65			0.64			0.13		
43.	0.67			0.65			0.65			0.10		
47.	0.65			0.74			0.63			0.58		
58.	0.75			0.78			0.74			0.25		
46.	0.78			0.80			0.78			0.24		
25.[a]	−0.75				0.59		−0.80				< 0.01	
1.	−0.15				0.19		−0.11				0.22	
51.	−0.61				0.65		−0.59				0.30	
33.	−0.47				0.45		−0.45				0.23	
19.	−0.61				0.69		−0.59				0.50	
31.	−0.47				0.74		−0.48				0.54	
5.	0.38					0.81	0.30					0.76
28.	0.46					0.71	0.40					0.58
15.	0.38					0.85	0.31					0.79
2.	0.20					0.50	0.14					0.49
57.	0.55					0.54	0.53					0.35
	ω	0.75	0.65	0.78	0.58	0.74	ω_S	0.20	0.71	0.06	0.20	0.63

[a]Item has already been reverse scored to make it consistent with the other items in its factor, as is appropriate for reliability calculations.

Table 4. PUGS: Sub-scales and items.

Sub-scales	Questionnaire items	Scoring direction
Decision-making	1. My actions were not influencing the outcome of the game.	Score
	2. I could not choose which actions were better.	Score
	3. I could not say if the game had more than one outcome.	Score
	4. I did not know how my performance influenced the outcome.	Score
	5. I did not know how the outcome(s) were connected to what I did.	Score
Exploration	6. I needed to discover things to make progress.	Score
	7. I needed to explore in order to know what to do next.	Score
Taking action	8. I felt I was stuck during the game.	Score
	9. I found it difficult to keep track of all elements in the game.	Score
	10. The game mechanics were overwhelming.	Score
	11. I think what I was doing in the game was not right.	Score
	12. I was not confident that I could perform some actions in the game.	Score
	13. The actions I had to perform were too demanding for my skills.	Score
	14. I struggled to do the right actions.	Score
Problem solving	15. I knew how each goal could be achieved.	Reverse
	16. I understood the game mechanics.	Reverse
	17. I knew how to play the game when I started.	Reverse
	18. I often felt lost.	Score
	19. I could find the solutions required for achieving the goals of the game.	Reverse
External	20. The game was unfair.	Score
	21. Unpredictable random elements were influencing my performance.	Score
	22. I was relying on chance in the game.	Score
	23. Random elements in the game were preventing me from achieving my goal.	Score
	24. The outcome of my actions was mainly influenced by chance.	Score

their ω_S values which suggest that they have substantial specific variance distinct from the general factor. They are not entirely unrelated so they do have some bearing on the general concept of uncertainty but they are likely to represent identifiably distinct aspects of uncertainty.

By contrast, Factors UDM, UTA and UPS show only small further specific variances so are mostly represented by the general factor. This is consistent with the correlations seen in the EFA on the first half of the data, Table 2.

3.7 Discussion

From the above analysis, we have produced a 24-item psychometric scale, PUGS, that captures what appears to be a broad feeling of uncertainty among players as they play games. This scale, along with the direction of scoring of participant ratings is presented in Table 4. There is good evidence from the initial exploratory analysis and subsequent validation that PUGS has five distinct factors. Three of these factors are strongly related to each other as the feeling of uncertainty in actually playing the game but with emphasis on different aspects of decision-making (UDM), taking actions (UTA) or problem solving (UPS). These three factors seem to capture important aspects of the internal uncertainty that players might experience. The other two factors are essentially distinct from these three and each other. The EXP factor concerns gathering information about the game to alleviate uncertainty. The EXU factor concerns the feeling of unpredictability, that is, aligned with uncertainty that is aleatoric and external rather than epistemic.

This gives a useful analysis of how players might experience uncertainty while playing games. In the next section, we demonstrate the construct validity of the scale with an experiment using a modern commercial game.

4 Experiment: Contraption Maker

In order to demonstrate the convergent validity of the PUGS in capturing players' feelings of uncertainty in games, the following experiment was undertaken. As players play a game and move toward completing a level in a game, they are likely to start resolving the uncertainty in the game and heading toward a known game state, that is completing or failing. Thus, the aim of the experiment was to alter the point of measurement of uncertainty so that in the early measurement, players would be further from completing a level and therefore feel more uncertainty than those measured later and more likely to be closer to completing a level. We deliberately avoided altering the game play in this study because there can be many consequences to changing gameplay that may affect experiences other than uncertainty and so alter players engagement with the game. By simply altering the time at which players were asked about their experience and without them knowing in advance when that would be, players would essentially experience the same game in both conditions. Additionally, rather than expecting to change all aspects of uncertainty, we chose a commercial puzzle game that relied heavily on problem solving and decision-making skills where all the information needed was present on the screen. Thus, uncertainty should not be due to external factors but should arise from their overall internal uncertainty, with emphasis on UPS and UDM. With more time, their internal uncertainty should reduce corresponding to them making progress on the puzzle and feeling more certain about what they should do.

4.1 Game

For purposes of this study, we chose the game *Contraption Maker* Spotkin Games (2014). In this game, inspired by the classic game *The Incredible Machine* Dynamix (1993), players must assemble a *Rube Goldberg*, Wolfe and Goldberg (2000) like machine that chains together a series of devices to solve a simple task. The game includes several different levels of increasingly difficult contraptions that players must build with an increasingly large number of options.

For purposes of this experiment, we chose an introductory level: the Bowling Blender Puzzle (Figure 3). In this level, players must activate a series of levers and conveyor belts to move bowling balls around the playfield to ultimately turn on a blender. This level was piloted with a small set of players, and we found that players with no previous experience were able to reliably finish this level in approximately 5 minutes.

In order to avoid confounding the results with issues regarding learning the game, an information sheet was prepared with an overview of the different parts of the interface, the specific controls they would use and how to activate them.

4.2 Participants

Participants were recruited using opportunistic sampling from the undergraduate and graduate students at the University of York. Students were recruited from university common areas such as the library and cafes in order to ensure students were not all specific to a particular department. Forty participants (14 male and 26 female) with ages ranging between 18 and 32 years (M = 24.33, SD = 2.89). Of the 40 participants, only 16 of them are frequent digital game players, with only eight of those playing more than 2 hours in a typical play session. There were 12 participants who reported to have played puzzle genre games, and though none of them had played the game *Contraption Maker* itself, 21 of these participants had played games like it.

4.3 Variables

The independent variable was duration of play with one group, unbeknownst to them, having 3 minutes to solve the puzzle and one group having 5 minutes. The dependent variable was player's feeling of uncertainty as measure by the PUGS.

4.4 Materials

All trials were conducted in a quiet room free of distraction on the University of York campus. The digital game *Contraption Maker* was played on 15.6-inch Windows

Figure 3. A screenshot of the initial state of *contraption maker*: the Bowling Blender Puzzle.

Laptop using a desktop mouse. The PUGS was presented on paper as the 24-Likert items in a random order.

4.5 Procedure

Participants arrived at the room and were briefed regarding the nature of the experiment and what they would be expected to do. Participants were provided with an information sheet to review, and asked to sign an informed consent form. Participants were then given the opportunity to study the information sheet that explained the components of the game interface and the controls.

After the briefing session, participants were asked to start playing the game. Each participant was randomly assigned to one of the two conditions of playing for either 3 minutes or 5 minutes. Participants were asked to play the game to solve the puzzle, with the knowledge that they could be asked to stop at any time. Participants were not told the maximum length of time they would play.

When participants had completed playing the game, they were asked to fill out the PUGS and a demographic questionnaire. Participants received a chocolate bar in thanks for their time in playing the game.

4.6 Results

After reversing items 15, 16, 17 and 19, the responses were summed within each specific factor. The scores on each specific factor of the PUGS are presented in Table 5. Due to the variation within the data, we chose to undertake a Brunner–Munzel test (also called a Generalised Wilcoxon Test) on each specific factor with an $\alpha = 0.05$, using the dominance statistic A' for effect size. The box plots for UPS and UDM are shown in Figure 4 and Figure 5 respectively.

The ratings of UPS ($U = 2.48$, $df = 36.76$, $p = 0.018$) and UDM ($U = 2.7627$, $df = 35.349$ and $p = 0.010$) were significantly higher in the 3-minute condition, while all other specific factors were found to have no significant differences. The effect size was found to be high for both with UPS having $A' = 0.71$ and UDM having $A' = 0.72$.

4.7 Discussion

The results of this experiment show that both UPS and UDM were higher when players had played for less time. This provides evidence for the construct validity of the UPS and UDM factors of the PUGS because these measures differed in agreement with the intended manipulation of uncertainty in the experiment. When individuals are given more time to solve the puzzle, their internal uncertainty decreases. Similarly, the experiment supports the identification of distinct components of internal uncertainty because the UTA factor did not differ as much as the UDM and UPS factors. This indicates that even though these three factors are related to one another, it is worth using them separately from one another to distinguish between the problem solving and decision-making cycle and realizing choices as actions in the game. This could be particularly important when we come

Table 5. Summary statistics and results of the Brunner–Menzel test for each specific factor of the PUGS comparing the two level variable of time played.

	5 minutes			3 minutes			Brunner–Menzel			
Factor	Mean	Median	SD	Mean	Median	SD	U	df	p	A'
UPS	11.95	11.0	3.58	15.25	14.5	4.66	2.48	36.76	0.018	0.71
UDM	11.25	19.9	3.99	14.25	14.0	4.19	2.76	35.35	0.009	0.72
UTA	17.45	18.5	5.61	20.20	22.50	7.40	1.50	28.03	0.144	0.64
EXP	11.40	12.0	4.43	12.60	13.0	3.30	−0.76	36.94	0.450	0.44
EXU	9.55	10.0	9.55	9.30	10.0	0.83	0.85	31.35	0.399	0.58

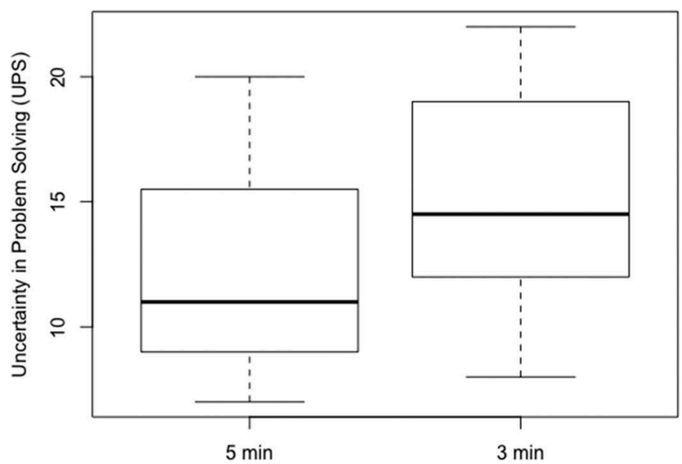

Figure 4. Boxplot of UPS in the 5- and 3-minute conditions.

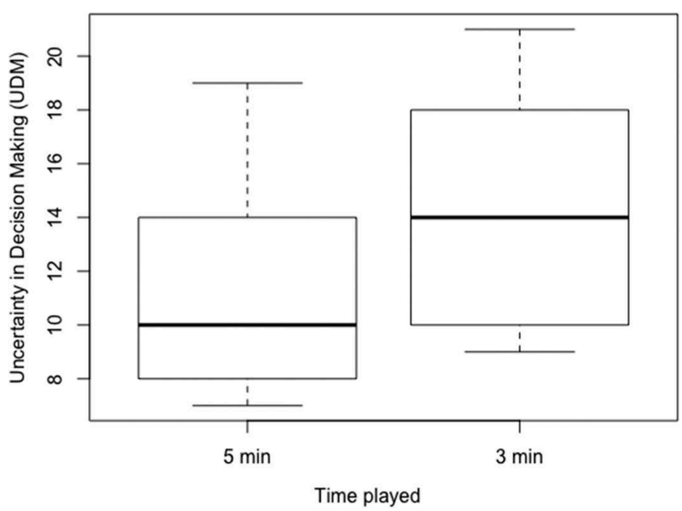

Figure 5. Boxplot of UDM in the 5- and 3-minute conditions.

to dissociate aspects of interface controls and presentation from the actual gameplay challenges.

Further, there was no detectable difference in EXU, the EXU the person feels about the game, and the EXP specific factor. This is as expected, given that all participants had a finite number of choices of devices that could be added to the contraption, and they all needed to explore a similar amount in the game regarding how different devices worked in the game playfield.

This is all very encouraging in the use of the scale. We have demonstrated that we can make specific choices about gameplay and predict the outcome of the specific factors of the scale. While we have used primarily game players from a student population, they had a range of experiences in games, with different levels of interest in playing games. It is encouraging that while most did not have experience with this specific game, they were able to play and move toward resolution of the puzzle, allowing us to detect the differences in their uncertainty. While *Contraption Maker* is but a single game in a specific genre of puzzle solving, self-paced gameplay, the different behaviors of the factor give confidence on the construct validity of the questionnaire as appropriate to measure players' feelings of uncertainty but also on the divergent validity of the components of uncertainty as capturing different aspects of the experience of play.

5 Conclusions and Further Work

Uncertainty has been consistently recognized as an important constituent of digital games from the earliest days of research in this area. However, until now, uncertainty has primarily been seen as a mechanic or attribute of the game rather than a feeling engendered in the player. With the work described here, we have taken the notion of felt uncertainty as a key, foundational experience of playing digital games and developed a measure that could be used to investigate this experience more systematically.

From the exploratory and confirmatory factor analyses, the PUGS seems to offer a statistically reliable instrument that operationalizes felt uncertainty into five distinct factors. Three of these factors are internal to the individual, representing their epistemic uncertainty in games, specifically their cognitive effort in solving problems in the goal space (UPS), making decisions around outcomes (UDM), and then meaningfully taking action to move toward the outcome(s) desired (UTA), for good or ill. Another important factor reflects the uncertainty that is external to the individual (EXU), in particular capturing both things that the player does not or cannot know about the games alongside the chance or aleatoric uncertainty about the game world. As noted by Costikyan (2013), these could arise from randomness, complexity, hidden information or even the unpredictability of co-players but are outside of the players' explicit ability to control. The EXP factor, though not strongly represented in the PUGS, suggests that players are able to think how to resolve their uncertainty and this may influence the feelings of both internal and EXU.

As well as showing generally acceptable scale reliability, the experiment shows that there is some validity in PUGS. Some aspects of felt uncertainty can be changed during the course of play. However, it is particularly notable that it was possible to dissociate changes in the internal, epistemic uncertainty components that reduced as a result of playing longer from the EXU, which was essentially unchanged with time for the period of play because there were no random or hidden aspects to the game.

The experiment with a commercial game is also informative of how this scale might be applied to other games. For example, if we take another game such as Tetris, we can generate a set of testable theories about how design choices within the game might influence uncertainty. For example, adding a preview of the next piece in Tetris might lead to people knowing more about the world and reduce EXU, but revealing three pieces increases the problem solving and decision-making load. In another scenario, we might predict that a novice at Tetris would have substantially higher uncertainty about how to make decisions and take actions in the game, and feel they need to explore much more to be successful. Conversely, the expert player may not need to explore at all, but find their internal uncertainty remains high when the game increases in difficulty. Similarly with other games, there are clear predictions that can be made around uncertainty for instance by increasing or reducing the "fog of war" to alter EXU (Kumari, Power, & Cairns, 2017), or adding complexity to games to increase the uncertainty around taking actions.

Indeed, as we understand more about this foundational experience, it will be important to connect it to other experiences that individuals have in games. The novice player above may quit the game outright if they find the uncertainty is too high, and the game becomes a hopeless task of failure, whereas if they have the uncertainty balanced to their experience the game perhaps remains challenging enough to keep advancing and not to give up. Alternatively, when the expert player becomes truly expert, the game may become boring, and they will need new uncertainties to encourage them to continue to engage and improve their skills.

Overall, the measure of felt uncertainty provided here can be easily deployed and so will help us explore the above scenarios and understand better experiences that contribute to a positive or negative, player experience.

Acknowledgments

The authors would like to thank the participants for their time and commitment in completing the studies discussed in this article.

ORCID

Christopher Power http://orcid.org/0000-0001-9486-8043

References

Bethesda Game Studios. (2011, November). *Elder Scrolls V: Skyrim. Windows, XBox 360, XBox One), Playstation 3, PlayStation 4, Nintendo Switch*. Maryland, USA: Bethesda Softworks.

Bioware. (2012, March). *Mass Effect 3. Microsoft Windows, PlayStation 3, Xbox 360, Wii U, Xbox One*. Texas, United States: Electronic Arts.

Bioware Edmonton. (2014, November). *Dragon age: Inquisition. Microsoft Windows, PlayStation 3, PlayStation 4, Xbox 360, Xbox One*. Edmonton, Canada: Electronic Arts.

Blizzard Entertainment. (2010, July). *Starcraft II. Microsoft Windows, OS X*.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425. doi:10.1007/s11336-006-1447-6

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, *45*(4), 624–634. doi:10.1016/j.jesp.2009.02.016

Caillois, R., & Barash, M. (1961). *Man, play, and games*. Illinois, USA: University of Illinois Press.

Cairns, P. (2016). Engagement in digital games. *In* Editors Heather O'Brien & Paul Cairns (Eds.),*Why engagement matters* (pp. 81–104). Springer International Publishing.

Cairns, P., Cox, A., & Nordin, A. I. (2014). Immersion in digital games: Review of gaming experience research. *Handbook of Digital Games*, *1*, 339.

CAPCOM. (1991, November). *Street Fighter 2: The world warrior*. Osaka, Japan: Arcade

Chalmers, R. P. (2012). MIRT: A multidimensional item response theory package for the r environment. *Journal of Statistical Software*, *48*(6), 1–29. doi:10.18637/jss.v048.i06

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*(4), 31–34. doi:10.1145/1232743

Collins, E., & Cox, A. L. (2014). Switch on to games: Can digital games aid post-work recovery? *International Journal of Human-Computer Studies*, *72*(8), 654–662. doi:10.1016/j.ijhcs.2013.12.006

Costikyan, G. (2013). *Uncertainty in games*. Cambridge, USA: MIT Press.

Creative Assembly. (2014, October). *Alien: Isolation. Microsoft Windows, PlayStation 4, PlayStation 3, Xbox One, Xbox 360, Linux, OS X*. Horsham, United Kingdom : Sega.

De Vellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). California, USA: Sage Publications.

DMA Design. (1991, February). *Lemmings*. Edinburgh, United Kingdom : Amiga. Psygnosis.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. doi:10.1111/bjop.12046

Dynamix. (1993). *Incredible machine. MS-DOS, 3DO, Windows, Macintosh, iOS*. Oregon, USA: Ubisoft.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey, USA: Lawrence Erlbaum Associates.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. London, United Kingdom: Psychology Press.

Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. *In* W. Brun, G. Keren, G. Kirkebøen & H. Montgomery (Eds.), *Perspectives on Thinking, Judging, and Decision Making*, 21–35, Oslo, Norway.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis. 1998*. New Jersey, USA: Prentice Hall.

Hipster Whale. (2014, November). *Crossy Road. iOS, tvOS, Android, Windows Phone, Windows 8.1, Windows 10*. Melbourne, Australia: The Arcade.

Juul, J. (2010). *A casual revolution: Reinventing video games and their players*. Massachusetts, USA: MIT Press.

Juul, J. (2011). *Half-real: Video games between real rules and fictional worlds*. Massachusetts, USA: MIT Press.

Juul, J. (2013). *The art of failure: An essay on the pain of playing video games*. Massachusetts, USA: MIT Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Basingstoke, United Kingdom: Macmillan.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*(2), 143–157. doi:10.1016/0010-0277(82)90023-3

Kline, P. (1994). *An easy guide to factor analysis*. California, USA: Routledge.

Kline, P. (1998). *The new psychometrics: Science, psychology and measurement*. Abingdon, United Kingdom: Routledge.

Kuhlthau, C. C., Heinstro¨M, J., & Todd, R. J. (2008). The 'information search process' revisited: Is the model still useful. *Information Research*, *13*(4), 13–14.

Kumari, S., Power, C., & Cairns, P. (2017). Investigating uncertainty in digital games and its impact on player immersion. In *Extended abstracts publication of the annual symposium on computer-human interaction in play* (pp. 503–509).

Lazzaro, N. (2009). Why we play: Affect and the fun of games. *Human-Computer Interaction: Designing for Diverse Users and Domains*, *1*, 155.

LeBlanc, M. (2006). Tools for creating dramatic game dynamics. *In* K. Salen & E. Zimmerman (Eds.), *The game design reader: A rules of play anthology* (pp. 438–459). Cambridge, MA: MIT Press.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75. doi:10.1037/0033-2909.116.1.75

LucasArts. (1991, November). *Day of the Tentacle. MS-DOS, Classic Mac OS, Microsoft Windows, OS X, Linux, PlayStation 4, PlayStation Vita, iOS*. California, USA.

Malaby, T. M. (2007). Beyond play a new approach to games. *Games and Culture*, *2*(2), 95–113. doi:10.1177/1555412007299434

Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on human factors in computing systems* (pp. 63–68).

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning.

Journal of Experimental Psychology: General, *145*(3), 266. doi:10.1037/xge0000140
Midjiwan. (2016, December). *The battle of polytopia. Android, iOS*.
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
Onwuegbuzie, A. J., Jiao, Q. G., & Bostick, S. L. (2004). *Library anxiety: Theory, research, and applications (No. 1)*. Maryland, USA: Scarecrow Press.
Power, C., Denisova, A., Papaioannou, T., & Cairns, P. (2017). Measuring uncertainty in games: Design and preliminary validation. In *Proceedings of the 2017 chi conference extended abstracts on human factors in computing systems* (pp. 2839–2845).
Rautzenberg, M. (2015). Navigating uncertainty: Ludic epistemology in an age of new essentialisms. *In* M. Fuchs (Ed.), *Diversity of Play*, *1*, 83.
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. doi:10.1080/00273171.2012.715555
Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. doi:10.1080/00223891.2012.725437
Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Massachusetts, USA: MIT Press.
Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. (2006). Video game uses and gratifications as predictors of use and game preference. *Playing Video Games: Motives, Responses, and Consequences*, *24*(1), 213–224.
Spotkin Games. (2014, July). *Contraption maker. Steam*. Oregon, USA: Ubisoft.
Sternberg, R. J., & Sternberg, K. (2016). *Cognitive psychology*. Massachusetts, USA: Cengage Learning.
To, A., Safinah, A., Kaufman, G. F., & Hammer, J. (2016). *Integrating curiosity and uncertainty in game design*. Dundee, United Kingdom: Digra/fdg.
Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). *Journal of Experimental Psychology: General*, APA, Washington DC, USA, *145*(10), 1280–1297.
Ubisoft Montreal. (2013, October). *Assassin's Creed IV: Black Flag. Microsoft Windows, PlayStation 3, PlayStation 4, Wii U, Xbox 360, Xbox One*. Quebec, Canada: Ubisoft.
Valve Corporation. (2013, July). *DOTA2. Microsoft Windows, OS X, Linux*. Washington, USA.
Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the user engagement scale. *Computers in Human Behavior*, *32*, 123–132. doi:10.1016/j.chb.2013.12.001
Wolfe, M. F., & Goldberg, R. (2000). *Rube Goldberg: Inventions!*. New York, NY: Simon and Schuster.
Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772–775. doi:10.1089/cpb.2006.9.772
Zynga. (2009, June). *Farmville. Android, iOS*. California, USA.

Notes on contributors

Christopher Power is an Associate Professor at the University of York and an Adjunct Research Professor at Western University. He has an active research program in accessible player experiences for people with disabilities and serves as the Vice President of The AbleGamers Charity. His favorite MCU character is Captain America.

Paul Cairns is a Reader in Human–Computer interaction at the University of York. He has long-standing research expertise in the area of players' experiences of digital games, in particular immersion and social presence, and an unusual passion for statistical methods for HCI. His favorite MCU character is Thor Odinson.

Alena Denisova is a Tutor (Associate Lecturer) in the Department of Computer Science at Swansea University, UK. She received her PhD in Computer Science (Human–Computer interaction) at the University of York, UK in 2017. Her current research is focused on the subjective experience of playing digital games.

Themis Papaioannou graduated from the University of York with an M. Sc. in Human Centred Interactive Technologies in 2016.

Ruth Gultrom graduated from the University of York with an M.Sc. in Human Centred Interactive Technologies in 2017.