5.

Modeling and Designing for Key Elements of Curiosity

Risking Failure, Valuing Questions

Alexandra To, Jarrek Holmes, Elaine Fath, Eda Zhang, Geoff Kaufman, & Jessica Hammer

Transactions of the Digital Games Research Association

November 2018, Vol 4 No 2, pp 138-171

ISSN 2328-9422

ABSTRACT

In this paper, we present a design model of curiosity that articulates the relationship between uncertainty and curiosity, and defines the role of failure and question-asking within that relationship. We explore ways to instantiate failure and question-asking within a cooperative tabletop game, share data from multiple playtests both in the field and lab, and investigate the impact of design decisions on players' affective experiences of failure and their ability to use questions to close information gaps. In designing for *comfort with failure* we find that helping players manage the aversiveness of potential failure can help prevent it from stifling curiosity, and that affective responses to failure can be modified by aesthetic decisions, as well as by group norms. In designing for *comfort with questions* we find that empowering quieter players supports the entire group's efforts to express curiosity, flexibility in enforcing rules fosters curiosity, and questions can serve multiple simultaneous roles in supporting and expressing curiosity. We discuss how these findings can be used in other games to support curiosity in play.

Keywords

curiosity, uncertainty, game design, failure, question-asking;

INTRODUCTION

Fostering curiosity – a mindset that relishes uncertainty and motivates its reduction through inquiry and exploration – is a common goal in game design, but is nonetheless an undertaking that presents considerable challenges to designers. Whether player curiosity is viewed as a means of triggering and sustaining engagement during play, or as a transformational aim of game play itself (e.g. to trigger players' curiosity about a particular topic or context featured in the game), designers must contend with the fact that curiosity involves acknowledging gaps in one's own knowledge and taking steps, often without any guarantee of success, to reduce them (Loewenstein 1994). Thus, curiosity requires individuals to frame *uncertainty* and *the risk of failure* in a positive light, to be motivated and energized by unknowns, and to accept that one is bound to make mistakes in the pursuit of discovering new knowledge. A key factor in facilitating this positive framing, we argue, is an individual's affective (i.e., emotional) experience of uncertainty and failure. In the face of uncertainty, will individuals feel capable, well-equipped, and secure in their ability to reduce a gap in knowledge, or will the anxiety of the unknown, a lack of self-efficacy, or insufficient agency prevail?

Within a game, designers can construct contexts and situations that influence individuals' curiosity-relevant affective states. Games are rife with moments of uncertainty and failure and, if designed with an understanding of the role of player affect, can offer players a safe environment in which to experience these potentially aversive states as motivating rather than threatening (Gee 2003). For example, most games are repeatable experiences, giving players the opportunity to learn from and correct previous mistakes – and to view past or present failures as challenges, not threats. Presenting players with the *right* amount of safety to confront uncertainty and failure, however, requires a delicate balance – if repeatability completely removes uncertainty and the potential for failure, then curiosity itself is thwarted. Thus, shifting the safety balance too far in one direction can result in either disinterest if excessive familiarity or predictability breeds habituation and boredom, or disengagement if excessive uncertainty or unmitigable randomness becomes overwhelming rather than energizing.

At the same time, curiosity-focused design requires more than simply igniting and sustaining the motivation to inquire and explore – it also means providing the support and the tools to do so effectively. We focus here on *questions* as a specific tool that can enable players to express and potentially satisfy their curiosity. By asking questions, game players can confirm knowledge gaps, voice their uncertainty (thereby creating social norms of uncertainty in multiplayer settings), and ultimately reduce uncertainty through developing and deploying "good" questions.

In this paper, we aim to articulate the complex relationships between curiosity, uncertainty, failure, and questions through a design model of curiosity. We demonstrate this design model by describing the design work on our curiosity tabletop game, *Outbreak*. *Outbreak* is an asymmetric, cooperative board game for two to five players. Together, players must explore a rogue scientist's laboratory to find the antidote to a dangerous disease. One player takes the role of a robot, who can explore dangerous spaces within the laboratory. The rest of the players, in their role as scientific investigators, must question the robot to discover what challenges stand between them and the antidote, collaboratively develop hypotheses about overcoming those challenges, and manage limited resources in executing their plans.

In *Outbreak*, we operationalize curiosity through two specific curiosity elements: (1) *comfort with uncertainty,* which relates to players' perceptions of failure, their comfort and willingness to take risks, and their search for unanswered questions, and (2) *comfort with questions*, which relates to players' perceived abilities to fill a knowledge gap and cope with uncertainty, their persistence towards understanding, and their assessment of their own knowledge states. We detail a three-month period of playtesting in both lab and field settings, discerning player responses to these curiosity goals through both observational and self-report measures deployed during these sessions. In our analysis of this data, we centered on two key themes: (1) shifting players' orientation toward failure as a challenge rather than a threat, and (2) developing effective question formulation skills in curiosity-driven exploration. We then link these emotional and behavioral outcomes to specific design decisions and game mechanics related to curiosity, and detail our iterative game design process. We close by presenting a set of implications and general considerations for curiosity-oriented design.

LITERATURE REVIEW

Our survey of the literature on curiosity provided insights about the affective and behavioral experiences of and responses to curiosity, in particular the emotional consequences of uncertainty and the risk of failure and the key mediating role played by exploratory responses, such as question-asking, in managing those emotional consequences. These insights directly informed the development of a working design model of curiosity, and, as we foreshadow in each of the following subsections, produced concrete game design goals that directed the development of *Outbreak*.

Curiosity and Uncertainty

Curiosity can be understood as an appetite for information, or the desire to fill an information gap (Loewenstein 1994). This gap, a violation of what is known or expected, can motivate a range of responses depending on the affective state that the newly salient uncertainty triggers. Among the factors that affect whether this discomfort is felt as a curiosity "itch" rather than an aversive "irritant," an individual must see themselves as able to close that information gap and resolve the uncertainty (Proulx & Inzlicht 2012). If the gap in knowledge is too wide to be perceived as surmountable – for example, if a student believes they are not capable of learning a new subject – it can result in frustration, disengagement, or trivialization (Proulx & Inzlicht 2012). If the gap is too narrow – as in the case of a student who gets the answers to the test ahead of time – it can inspire indifference, as the gap is not seen as challenging, surprising, or compelling enough to merit further investigation (Engel 2013).

In designing for curiosity, we need to create compelling information gaps that game players can become aware of and feel challenged by, but that they also feel capable of resolving. Presenting players with elements or experiences of uncertainty is a key component of existing models of game engagement (Costikyan 2013), and our own work has begun to further elucidate the links between curiosity and uncertainty from a game design perspective (To et al. 2016a). At the same time, if uncertainty becomes unmanageable or uninteresting to players, it has the potential to disrupt the experience of flow by creating an imbalance between perceived challenges and perceived skills (Csikszentmihalyi 2014). As game designers, we can seek to create games that encourage an instance-specific curiosity known as state curiosity (Carlin 1999). In addition to presenting moments of uncertainty to players, ensuring that the uncertainty presents the appropriate level of challenge, and equipping them with the skills to navigate and resolve that uncertainty, supporting uncertainty means triggering positive affect. Challenge is known to be one of the core pleasures of gameplay (Hunicke et al. 2004). In moments when players have both the ability and the desire to answer questions, a “virtuous cycle” of curiosity can therefore occur, in which players cyclically uncover information gaps, become immersed in the search for answers, and become more deeply engaged in the play experience (Engel 2013; Jirout & Khlar 2012). That is the primary focus of this paper. As discussed in more detail below, the design of *Outbreak* specifically aimed to provide social and instrumental support for confronting and overcoming uncertainty -– for example, by making the confrontation of uncertainty a shared, collective experience, and equipping players with resources to scaffold the question-asking process. Of course, game design may also aim to have a lasting impact on player’s trait-level curiosity (i.e., their individual preferences for uncertainty). While the concepts discussed here may be extended towards long-term changes in trait curiosity, this is beyond the scope of the present work.

Curiosity and the Risk of Failure

Designing for curiosity means supporting positive affective experiences in the face of uncertainty, particularly when risking failure. However, positive affect is by no means a given when it comes to confronting uncertainty. Acknowledging a lack of information or a gap in knowledge can be an aversive state. Leading theories of curiosity posit that self-efficacy, the perceived ability to fill an information gap, plays a key role in determining whether uncertainty triggers affective states that are more positive or negative (Loewenstein 1994). If the level of uncertainty is too high, if the information gap is not obvious, or players do not perceive themselves as being capable of surmounting the challenge, curiosity may be stifled by the threat of failure (Berlyne 1966; Litman & Jimerson 2004; Loewenstein 1994; Proulx & Inzlicht 2012; Engel 2013; Rinkevich 2014). In contrast, when individuals experience the risk of failure as energizing, knowledge gaps can be framed and experienced as a challenge to overcome (Litman & Jimerson 2004; Loewenstein 1994; Berlyne 1966). Finally, in group settings, attitudes toward failure are often socially constructed – groups develop norms about expressing uncertainty and enforce social consequences for disclosing ignorance (Feldman 1984). These norms affect how much a person is willing to disclose their own knowledge, or lack thereof, to the group.

In games, the affective and social consequences of failure may be reduced compared to non-game contexts. Klopfer, Osterweil, and Salen (2009) identified failure as one of the five “freedoms” of play -– while we cannot truly “fail” at play, we can do things during play that look like failure in other contexts, but with lower risk and a more explicit opportunity for learning and growth. Similarly, Gee (2003) writes that in games, the risk of failure is lowered and, in fact, that failure is a good thing -– players can feel empowered to take more risks, get feedback when they fail, explore more, and ultimately learn from the experience. Juul (2013) argues that failure may be the central aesthetic experience of play. By confronting players with their limitations, games can provide players the opportunity to emerge victorious over their past failures. According to Juul’s analysis, becoming a better player means becoming a better fail-er. In short, games are already suited to pose potential failures as learning opportunities. However, game designers must still take into account players’ varying emotional relationships with failure, and imbue their games with safeguards to help players maintain a positive affective state (i.e., one that is motivated and energized rather than discouraged or disinterested). Below, we detail how we identified such safeguards in the iterative design of *Outbreak*, including the reduction of game elements that heightened players' anxiety about the consequences of failure (such as the potential loss of a character) and the importance of replayability in helping players realize opportunities to learn from and rectify their previous failures.

Curiosity and Questions

One safeguard against disengagement is the provision of tools that allow players to mitigate uncertainty and build self-efficacy around their ability to close information gaps (Proulx & Inzlicht 2012). The tool that we focus on here is the use of *questions*. When players encounter uncertainty, they can ask questions in order to express their curiosity, and they can use the information they receive to resolve information gaps. Questions are particularly useful for games utilizing hidden information or unsolved puzzles to build uncertainty (Costikyan 2013). Players can pose inquiries (e.g., to the game itself, to one another in social deception games, etc.) to reduce the information gap. Furthermore, in collaborative games like *Outbreak*, in which players have unique resources, questions may also aid in collective knowledge assessment. When players discover new information through their questions, question-asking can invoke the pleasures of discovery and exploration (Hunicke et al. 2004). Even the feeling of anticipation as the player waits to see what they will discover can be a source of pleasure in gameplay (Schell 2014).

While questions are a valuable tool for reducing uncertainty, and guiding players toward greater comfort, asking questions can be challenging. People's relationship with questions influences their likelihood to entertain, and willingness to voice those questions when facing uncertainty. First, individual personality factors such as assertiveness, self-esteem, and social anxiety determine one's general likelihood of asking questions (Mahdikhani et al. 2015). Second, social and situational cues indicate the cultural norms of question-asking in a given environment (Rocca 2010). For example, voicing uncertainty through question-asking can pose a social risk, but can also serve as a valuable means of assessing the relative or collective knowledge of the group (Mohammed & Dumville 2001).Finally, a person's perception of an authority figure can alter their relationship with questions. In the classroom, students' perceptions of a teacher as supportive versus condescending can dramatically alter their likelihood of asking questions (Mahdikhani et al. 2015). In game contexts, this might include player relationships with a gamemaster or with fellow players who have more information. In addition to comfort asking questions, we acknowledge that the content of those questions is of great importance, but falls beyond the scope of this work. While developing better question formulation skills can increase the odds of getting information that reduces information gaps, good questions can also reveal new gaps through the knowledge they yield.

BUILDING A DESIGN MODEL OF CURIOSITY

When creating games, game designers have limited control over player experience. They can produce rules, game systems, resources, narrative elements, and audio-visual assets. However, they cannot directly control player experience, and have limited control over player behavior. Game design theories, such as the MDA model (Hunicke et al. 2004), acknowledge this limitation. Designers can create systems of game mechanics, but they must predict both the dynamic behaviors that emerge from those mechanics when players interact with them, and the aesthetic experiences that players will have as a result. This model suggests a design challenge in creating games for curiosity. Curiosity is a player experience that can be *provoked* by game elements and *expressed* during play, but not directly manipulated by game designers. Creating games for curiosity therefore means developing a design model of the relationship between curiosity and uncertainty, and exploring how that relationship is mediated by specific elements that can be instantiated in gameplay.

Building on the literature reviewed above, we understand curiosity and uncertainty as existing in a dynamic system (Thelen & Smith 1996) with their interaction mediated by players' comfort with the risk of failure, as well as their comfort and proficiency with questions. Figure 1 illustrates the working model of the cyclical interrelationships between these elements that guided the present work.

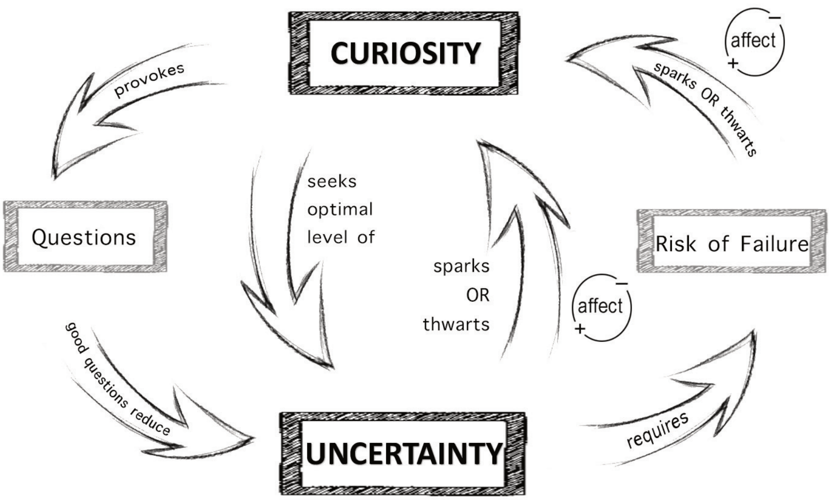

Figure 1:Uncertainty and curiosity have a cyclical relationship that is mediated by the risk of failure as well as by questions.

This model proposes that in order to spark and sustain players' curiosity and increase engagement and exploration, designers should strive to:

1. Present players with a level of uncertainty that is "optimal" – that is, a level that is experienced as challenging rather than overwhelming
2. Provide players with opportunities, in facing uncertainty, to fail in their attempts to reduce information gaps, and to perceive failures as energizing rather than threatening
3. Equip players with the ability to ask questions, and to

increase their proficiency with question-asking, in the pursuit of resolving uncertainty

In this way, the right-hand side of the model can be thought of as a "growth" cycle between curiosity, uncertainty, and failure. If curiosity is triggered by a manageable level of uncertainty, and players construe failure as a challenge, both uncertainty and failure are more likely to elicit positive affective responses and spark higher levels of curiosity. The left-hand side of the model represents a "reduction" cycle between curiosity, uncertainty, and questions. Curiosity motivates inquiry, and good questions ideally (but not inevitably) reduce levels of uncertainty. In both of these cycles, designers must help ensure player comfort (e.g., comfort with the expression of uncertainty, the possibility of failure, and the process of formulating and posing questions) to sustain engagement and, at the same time, prevent player complacency (e.g., by helping players to manage but not fully remove the risk of failure, and reduce but not fully resolve uncertainty).

This model provided us with a set of guidelines and goals for our design of the game *Outbreak*: creating an overall level of uncertainty that would be experienced as challenging rather than overwhelming, helping players experience failure as energizing, and increase player proficiency with question-asking. The following sections describe how the iterative design and testing of the game were informed by this model, and reveal the design lessons and implications that emerged in the process.

GAME DESIGN AND DEVELOPMENT

The "Sensing Curiosity in Play and Responding" (SCIPR) project aims to design and study game-based interventions for encouraging curiosity through play, particularly for marginalized students who may benefit from increased comfort with curiosity (e.g., female science students, racial minorities). These games are targeted toward middle school (9-14 year old) students. As a part of the SCIPR project, we have iteratively designed and prototyped several games. This paper focuses on one of those games, *Outbreak* (Figure 2). We use tandem transformational game design, which emphasizes iterating game designs alongside theoretical understanding of transformational goals – in our case, our design model of curiosity (To et al. 2016b).

Outbreak is a cooperative question-asking game for two to five players, in which the group must save a town from a rogue scientist by searching their laboratory for antidotes to a disease. Most players assume the role of scientific investigators, while one player takes the role of their robot assistant. Each investigator player receives a set of resource cards (e.g. characters or pieces of equipment) that include different skills (Figure 2D), such as strength, computer hacking, and friendliness (Figure 2C). Each time they enter a new room in the mad scientist's lair, the robot player can enter first and safely investigate the room. However, the robot cannot describe what they see. They can only respond to questions put forward in the *question-asking phase* by the investigator players, who then select the resource cards that will neutralize the threats inside and unlock the antidotes for that room.

On a given round, the robot player reads the back of a room card, which includes a description of the room and lists the skills needed to survive (Figure 2A). Because the robot player portrays a "sensing" robot, they cannot read aloud the card description. They can only answer questions posed by the other players. Investigator players have limited time during the question-asking phase to ask questions, following which they enter the *discussion phase* where they collaboratively either choose which cards to risk in that room or they can choose to pass the room. If they choose a successful combination of cards, they keep their cards and roll to receive antidote tokens. If they fail, they must discard their cards. If they choose to pass on the room, they keep their cards, but the countdown to the end of the game continues.

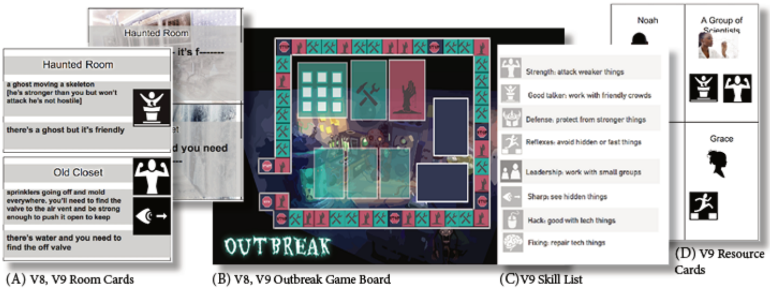

Figure 2:Outbreak game with components from V9 including (A) room cards, (B) the game board, (C) the list of skills, and (D) resource cards.

Outbreak, to date, has gone through 12 iterations. In this paper we discuss versions five, eight, and nine (V5, V8, V9) of *Outbreak*, all of which were studied with players from our target demographic, and which reflect major shifts in both our playtesting and design. Between V5 and V8, we moved from playtesting in the lab to playtesting in the field, and adjusted affective elements of the game; between V8 and V9, we changed the question-asking system and added new data collection measures. We discuss these choices further in the next section of this paper.

METHODS

This paper reports on the iterative design and playtesting process for *Outbreak*. Over the span of four months we playtested V5, V8, and V9 with participants in our target age demographic, 9-14 years old. Other versions of the game were playtested with players outside our target audience (e.g. for game balance) and are not reported in this analysis. We conducted two playtests of V5 in a controlled lab setting, referred to as the lab playtests (“Lab”). We conducted ten field playtests with versions eight (V8) and nine (V9) at two local summer programs in Pittsburgh, PA, referred to as the field playtests. Site one was a local science center (“SC”), and site two was a YMCA in a primarily black, low-SES neighborhood (“YMCA”). See Table 1 for playtest details and codes.

Our playtesting process included 1) development of tools to measure players’ responses, 2) deployment of those measures, and 3) analyzing their responses. We focused our analysis on understanding players’ affective responses, particularly around uncertainty and failure, and on their ability to ask questions.

Group ID	Site	Game Version	Group ID	Site	Game Version
L1, L2	Lab	V5	Y2a, Y2b	YMCA	V9
Y1a, Y1b	YMCA	V8	Y3a, Y3b	YMCA	V9
S1a, S1b	SC	V8	S3a, S3b	SC	V9

Table 1. Group IDs for theOutbreak playtest groups. Each ID represents a single group of 3-4 players. With the exception of the lab studies, groups with the same number were played on the same date.

Measure Development

In addition to regular playtesting practices (e.g., observing player behavior, and focus group interviews about player experience) we set out to measure player experiences related to *Outbreak*’s transformational goals. We adapted best-practice methods from related fields when a validated measure did not yet exist, and then iterated those measures based on usability observations in the field.

Measuring Affect

In lab playtests of V5 and field playtests of V8, we collected player affective data using the Feelings Wheel (Kelley 2016). The Feelings Wheel includes six core emotions in the center of the diagram, and expands each outward into more specific emotions for a total of 77 feelings (see Figure 3A). To adapt this measure to our audience, we removed the emotion “sexy” as it was deemed inappropriate and uninformative. By circling emotions, players could capture how they felt during the game even if they did not have the language to generate emotion words on their own.

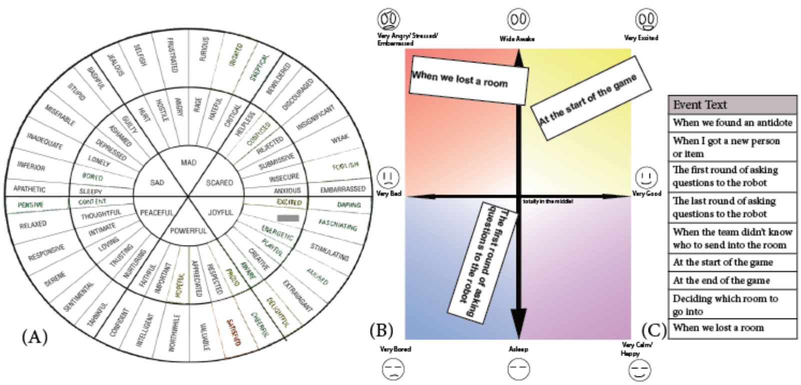

Figure 3:(A) The Feelings Wheel where participants circle distinct emotions felt (B) The valence-arousal map with sample event slips that participants place as a marker for emotions felt (C) List of game events used for Outbreak

For V9, we developed a version of a valence-arousal map for children’s emotion self-report. Our goal was to connect player emotional reactions to specific elements of gameplay. To accomplish this, we combined emotion valence mapping diagrams (Barrett 2004) and design-based post-it clustering activities (Hanington & Martin 2012). These cross-disciplinary tools both seek to capture and describe the user's self-reported spectrum of emotion with as much granularity and detail as possible. The map asks players to place prompts related to game moments (see Figure 3C) on a quadrant (see Figure 3B). The instrument was validated through multiple rounds of expert heuristic evaluation by cognitive psychologists and designers, and tested for usability in the field with children.

Game events were selected for their relationship to curiosity, uncertainty, failure, and question-asking. We coded each event for different types of curiosity (e.g., conceptual curiosity), different types of uncertainty (e.g., hidden information), game outcomes (e.g., failure/negative events), and when in the game we expected events to occur (e.g., early in the game).

Valence-arousal results were coded based on the x,y coordinate of the top left corner of each slip and the quadrant or quadrant boundary where it was placed. We also captured the relative horizontal and vertical placement on the graph in comparison to the other game events, using a ranking of 1-9. Slips that were placed on top of one another were given the same ranking.

Capturing Questions

We developed a field notes template for our playtest observations, both to standardize data capture across members of the research team and to ensure we captured relevant data. In our field playtests, we were unable to record video due to the limitations of the spaces available, in which children who had not consented to being videotaped were regularly present. We therefore manually captured the questions that investigators asked the robot player during the question-asking phase. Researchers were also directed to capture visible emotional responses to the game, unusual player behavior, and the gist of side conversations between players. When possible, researchers noted the game outcome, whether players succeeded in a particular room, and other observations related to playability and balance.

We coded the questions based on their form and content. A codebook was developed through a bottom-up analytic process led by researchers who had not participated in the design of the game. For example, questions were coded “skill word” if players directly asked about a word from the skill sheet (e.g. “Is it strong?”), “discovery” if they asked about the existence or something in the room (e.g. “Are there any computers?”), and “building off” if they ask a question that builds on information received within the round (e.g. “Are there zombies?”, “Are the zombies friendly?”). Questions could have multiple codes and every question was coded as “concrete” or “abstract”. Questions coded as concrete were ones that cited specific concepts or seemed to represent a specific hypothesis (e.g. “Is there a zombie?”, “Is it dark?”), whereas questions coded as abstract asked for non-specific information or closely referenced the skill words without a supporting hypothesis (e.g., “Is there a threat?”, “Do I need to fix something?”). After the codebook was complete, two researchers independently coded the questions and discussed diverging codes until they reached agreement. Additionally, we captured the group and gameplay round associated with each question. In some cases, we were able to use this data to code whether questions were asked during rounds that succeeded or failed, and whether players had won or lost the prior round.

Playtesting and Measure Deployment

In all playtests, participants played *Outbreak* in groups of three to five, with a researcher taking the role of the robot player. In L1 and L2, players did not know each other before the playtest. To create familiarity between players, both groups were asked to participate in an icebreaker game (To et al. 2016c) before playing *Outbreak*. In the field playtests, which were conducted in the context of ongoing summer programs, players were typically familiar with one another, so no icebreaker was used. Players were randomly assigned to groups, and playtests were scheduled as part of the regular activities of the program.

Participants were introduced to *Outbreak* as a cooperative board game currently in progress, and told that their early feedback would help the game designers improve the game. It was implied that designers were not present in the room in order to get as honest feedback as possible. Next, one researcher reviewed the rules with the players and played a scripted practice round that included a diverse set of sample questions. The same researcher adopted the role of the robot player for the remainder of the game. The researcher would answer questions about game mechanics if players explicitly asked or if they could not proceed with gameplay. Participants played until they won, lost, or 40 minutes had passed.

After gameplay, we collected emotion data. For the V5 and V8 playtests, each player was given a paper copy of the Feelings Wheel and asked to circle every emotion they had felt during play. The research team then collected the papers for analysis. For the V9 playtests, the researchers demonstrated how to place an event on the emotion map in a way that corresponded to a feeling. Participants were then given the nine event tokens and asked to place each token on a spot on the map that corresponded to their feelings at that point in the game. When participants indicated they were done placing tokens, the researchers photographed the map. If participants did not place any tokens, they were asked a second time if they wanted to complete the measure. If not, the researchers photographed an empty map.

After emotion data had been collected, players participated in a focus group interview. Participants were told that their feedback would be helpful in aiding the game designers to iterate the game and improve it. They were asked what they liked most about the game, what they would change, and any other feedback they'd like to share about the game.

During all phases of the playtest, an additional researcher, seated in the play space, took field notes using the notes template during play, captured feedback during the focus group interview, and made additional observational notes, as described in Measure Development measure development.

It is important to note that our data represents diverse playtests. Some participants played the game only once, while some played multiple times over several weeks; playtests occurred in a range of physical locations, from a formal lab setting to a cafeteria in a science center; and players played multiple versions. Given this diversity of data, it would be inappropriate to perform formal statistical analyses. Instead, we demonstrate that much can still be learned about curiosity and game design from diverse aggregate data.

DESIGNING FOR COMFORT WITH FAILURE

Exploring Comfort with Failure Through Design and Data

In order to explore the concept of comfort with failure, we first needed to operationalize failure within the design of *Outbreak*. Based on our rules design and observation of playtests, we identified three types of failure in the game. First, players could fail to find an antidote in a particular room, which we refer to as "room loss" (V5, V8, V9). Second, players could lose resources such as teammates (V5, V8) or gear (V5, V8, V9), which we refer to as "resource loss." Finally, players can lose the game, either by reaching the end of a countdown to midnight (V5) or by reaching the end of the game board (V8, V9) without finding enough antidotes, which we refer to as "game loss." Room and resource loss occur repeatedly throughout the game. However, game loss can occur only once and reflects players' overall performance.

During lab-based playtests of V5 (L1, L2) and V8 (Y1a, Y1b), we studied players' emotional and social reactions to the design decisions we made around room loss, resource loss, and game loss. Because we did not want to interrupt players between rooms, these playtests relied primarily on observation to understand room and resource loss, which occurred during play. At the end of the game, we collected self-report data on player emotional experience, which reflected their overall experience in the game.

To connect the data more directly to specific types of failure, we collected observational and valence-arousal map data from four playtests of V9 across two separate sessions at the YMCA site. During the first session, we observed two games involving eight students (Y2a, Y2b). A week later, we observed two games involving ten students, seven of whom had participated in the previous session (Y3a, Y3b). All students had previously playtested different games designed by our group in prior sessions. However, because none of the students had played *Outbreak* prior to Y2, we were able to explore how uncertainty and failure were experienced, both as first-time players and on a repeated encounter with the game.

Patterns from the DataIn our earliest playtests of Outbreak with participants from the lab playtests, we observed that failure was a salient concept to the students. Individual player's emotional responses to the threat of failure such as observable anxiety behaviors (e.g., facial expressions, wincing) and vocalized fear over losing often spread to the group, and how the group responded to that – either by amplifying it or dissipating it often had a profound impact on a group norm around failure moving forward in the game.

Failure and Affect

We observed two factors that influenced players' affective relationship to failure. First, we observed that narrative and aesthetic elements had a much stronger effect on players' emotional reactions to failure than we expected. Second, we observed that repeated play changed players' feelings about failure.

Early in the playtest process, we discovered that players felt attached to the resources in the game, and that they were often more willing to accept room loss (e.g. failure to collect antidotes) than resource loss. For example, in group L2, players asked questions such as, "Will we lose the scanner if we send it in?" Although the game's rules prohibit answering the question explicitly, the players decided that their scanner was at risk and chose not to send it into the room. Players correctly identified this decision as one that required weighing a guaranteed failure against the possibility of failure – only by chancing the loss of their scanner could they avoid the guaranteed loss of the room. We observed players experiencing anxiety around this decision, which could affect their willingness to take the risk.

To reduce the level of player anxiety about the risk of failure, we explored the role of narrative and aesthetic factors. Could we change the level of player anxiety using affective manipulations alone? Examining differences between player affective experiences in L1 and L2 suggested that we could. Players in group L1 were visibly distressed during play. Although they claimed in post-game interviews that they enjoyed the game, their Feelings Wheel data corroborated their distress. Of the 37 total emotions circled by four players, 24 were negative; 17 of those fell into the "scared" category, and all four players chose "anxious" to describe their feelings (Table 2). On the other hand, the four players in group L2 circled 49 total emotions, of which 44 were positive. All four players circled "aware" and "confident" to describe their experiences, and no negative emotion was circled by all four players. Our observations confirmed these differences. Players were concerned over the well-being of the game characters and their use of resources; they were sometimes anxious, but never visibly upset.

Group L1				Group L2			
Negative Emotions	**24**	**Positive Emotions**	**13**	**Negative Emotions**	**5**	**Positive Emotions**	**44**
Scared	17	Joyful	9	Scared	4	Joyful	19
Anxious 4 *Scared* 3 …				*Anxious* 3 …		*Aware* 4 *Joyful* 3 …	
Sad	6	Powerful	4	Mad	1	Powerful	15
						Confident 4 *Faithful* 3 …	
Mad	1	Peaceful	0	Sad	0	Peaceful	10
						Peaceful 3 …	

Table 2.Aggregate counts from the lab study groups (L1, L2) Feelings Wheel data. Counts for the two overall categories, positive and negative, are shown, as well as each of the six sub-categories. When three or more participants all circled the same emotion, that emotion is displayed with count data.

What could account for such an extreme difference between L1 and L2, given that the two sessions involved the same version of the game (V5)? During L1, we played a soundtrack of scary music in the background. Players repeatedly mentioned the music during gameplay, and they were visibly unnerved by it. The player response was sufficiently strong that we removed the music during L2 for the well-being of our players. Players in L2 still experienced anxiety, particularly when asked to weigh room loss against resource loss, as noted above. However, they appeared to be more resilient to this anxiety, focused less on the negative impacts of their failure, and had more positive feelings at the end of the game.

Another narrative element that affected players' willingness to take risks was the theming of resources. In earlier versions of the game (V5, V8), game resource cards included both scientific tools, such as a cloaking device or first aid kit, and scientist characters, such as Barbel the anxious ice researcher or Karolina the dependable virologist. Including scientist characters gave us the opportunity to introduce scientist role models who matched our target playtest groups, such as scientists who were female, black, Hispanic, or all three. At the same time, by making characters a collective resource, we hoped to create psychological distance between the players and the fate of their characters, who would serve to heighten the drama of the game. Unfortunately, this psychological distancing did not succeed. We observed that the highest levels of anxiety were associated with negative outcomes for characters. The idea that player choice could result in characters going into a coma was too frightening for our audience. In V9 we removed characters as a separate resource type and saw a reduction in player stress; conversely, if the game were being redesigned for older students, reintroducing threats to scientist characters could increase the level of tension.

Over and above the impact of narrative and aesthetic game elements, we observed that repeated play changed players' affective reactions to in-game failure. As noted earlier, we were able to test the same version of the game (V9) across two different playtest sessions (Y2 and Y3). During these sessions, we collected valence-arousal map data about specific game events, including times when the players failed to complete a room ("When we lost a room" in Table 3). After the second session, players reported affective *dampening*, or a trend toward neutral valence in their emotional reactions, for all game events with one exception – the event involving failure (see Table 3). Players reported feeling more positive about failure events after their second play session, with a decrease in negatively-coded and neutral-coded emotions and a 26.7% increase in positive affect (see Table 3). In other words, playing *Outbreak* a second time reduced emotional responses (i.e., both the high negative and high positive valence) of most game events, but made failure a better experience.

	Positive			Neutral			Negative		
	Y2	Y3	Shift	Y2	Y3	Shift	Y2	Y3	Shift
When we found an antidote	0.714	0.7	(-)	0.143	0.3	(+)	0.143	0	(-)
When I got a new person or item	1	0.7	(-)	0	0.1	(+)	0	0.2	(+)
The first round of asking questions to the robot	0.857	0.6	(-)	0	0.3	(+)	0.143	0.1	(-)
The last round of asking questions to the robot	0.715	0.5	(-)	0.143	0.3	(+)	0.143	0.2	(+)
When the team didn't know who to send into the room	0.572	0.6	(-)	0.143	0.2	(+)	0.286	0.2	(-)
At the start of the game	0.857	0.5	(-)	0	0.3	(+)	0.143	0.2	(+)
At the end of the game	0.714	0.8	(+)	0	0.2	(+)	0.286	0	(-)
Deciding which room to go into	0.572	0.6	(+)	0.143	0.3	(+)	0.286	0.1	(-)
When we lost a room	0.333	0.6	(+)	0.167	0.1	(-)	0.501	0.3	(-)

Table 3. Proportion of game events eliciting positive, neutral, or negative (valence) responses on the valence-arousal map measure across two repeated play sessions (Y2 and Y3).

Our prior work in this area emphasized the role of *uncertainty*, as instantiated in game design decisions, in provoking and supporting curiosity (To et al. 2016a). However, this research suggests that *aesthetic* and *contextual* decisions can change players' affect and hence their willingness to take risks. The same game, deployed in different ways (with or without a scary soundtrack, played once or repeatedly), can produce different affective experiences of failure.

Social Factors

Theories of curiosity suggest that social norms about uncertainty and failure will affect people's experiences of curiosity and their likelihood of expressing curiosity. In our playtests, we were able to deploy our game in two different social settings with different social norms: a Science Center and a local YMCA. We observed that social differences between the groups affected how players engaged emotionally and socially with the game. SC players were highly concerned with failure in ways that paralleled the students in our lab studies L1 and L2. We observed anxiety when players were at risk of losing resources. However, these emotions shaped not only their play decisions, but also their social activity during question-asking and discussion. During the question-asking phase of the game, these students spent most of their time thinking silently, presumably about the "right" questions to ask. As a result, they asked very few questions and received little information. With the little information they had, they would debate back and forth endlessly during the discussion phase and would require light prompting to make a decision to move forward. Their concerns over failure were so immense that it prevented them from failing with grace, and from learning. By comparing these students to the players from the YMCA, we can see that this behavior is not purely driven by game design decisions. YMCA students were not overtly concerned about failure or losing resources, particularly by comparison to the SC and lab groups. They tended toward lightweight, short discussion rounds and rapid decision-making, and would forge ahead quickly through many rooms. While both of these behaviors, reflecting and experimenting, are valid curiosity-relevant strategies, we ideally hope to foster both. Games designed for curiosity therefore require designs that are mindful of the social space they exist in. We want to design social spaces that can evoke the curiosity behavior that is most relevant to the goals of a particular curiosity game.

We note that even though social spaces can be designed to support different types of curiosity-relevant norms, differences in emotional response may be amplified by individual player factors. Because *Outbreak* is a cooperative game, players who are working together may experience “emotional contagion,” or their emotional response being affected by the individual emotional response of other players (Barsade 2002). We observed this behavior in group L1, where one player had a particularly strong emotional response to the scary music. While all players found it unnerving, their response was amplified by seeing the fear displayed by this particular player.

Design Lessons

Helping players manage the aversiveness of potential failure can help prevent it from stifling curiosity. In *Outbreak*, we ask players to embrace risk and uncertainty in order to avoid certain failure. We observed that when players were particularly afraid of risk, they chose certain failure rather than the possibility of failure. Fear of failure also sometimes thwarted strategies to reduce the chances of failure, such as when students became so involved in asking the “right” question that they did not ask enough questions to gather information. Understanding that, in some circumstances, risk can be more intimidating than the certainty of failure can be used to help design for curiosity in other types of games.

Affective responses to failure can be modified by aesthetic game design decisions. We found that aesthetic design decisions, such as narrative and contextual factors had a strong impact on players’ affective experience of failure. Scary music, named characters who were at risk, and first-time play all increased the anxiety level in play. Conversely, table talk, generic items, and repeated play all made failure a more positive experience. Finding the right level of difficulty for a game is often conceptualized as requiring game-mechanical balance; our findings suggest that aesthetics can also be used to balance gameplay when it comes to the perceived risk of failure.

Group norms influence the affective experience of failure and the strategies available to manage it. Players' social norms and the setting in which they are playing affect how willing they are to tolerate failure, to take risks, and to express ignorance in front of a group. For example, our SC and YMCA groups had very different rates of asking questions, even when using the same set of rules. These social norms can be affected by emotion contagion, in which a single player's strong experiences spread to other players. In other types of multiplayer games, designing for players who have outsized or outlier emotions can be a productive way of shifting the norms of the group.

DESIGNING FOR QUESTIONS

Exploring Question-Asking Through Data and Design

To explore this topic, we relied on observational data, valence-arousal map data, and question data from playtests for three different versions of the game in our on-site playtest settings, as well as our lab setting.

In every version of the game, each round of gameplay involves the previously described *question-asking phase* where investigators ask questions of the robot player. The question-asking phase is always limited by a timer. Question-asking mechanics varied between versions in two ways. First, in V5 and V8 players could ask an unlimited number of questions during the question-asking phase. In V9 we introduced battery tokens, which constrained both the number and form of questions. Immediately before each question round, players drew three tokens from a bag. Each token was a small rectangular battery with a question template (e.g., “How many _____?”, “____ need a ____ ____?”) (see Figure 4). In order to ask a question, players turn in a token to the robot player and ask a question matching the template. As discussed below, the robot player needed to use their judgment about how tightly to require the question to match the form. Second, we varied how rooms were displayed to invite curiosity. In V5, the rooms were displayed on a board in a map-style layout. In V8 and V9, the rooms were individual cards drawn from a deck. Cards featured a title and some clue words (e.g., the “Big Office” and “Full of broken ____ and a ____ ____”). (see Figure 2A).

Figure 4:Battery questions with question templates used in the question-asking phase of Outbreak (version nine)

We also use our coded question data to examine the effects of failure on players’ question development within a single gameplay session. Questions are coded as either occurring in the first round, or after a round in which they either failed or succeeded at overcoming a chosen room’s challenge. We use this information to explore the relationship between prior failures or successes in the game and players' decisions to build on, revise, or discard their hypotheses.

Patterns from the Data

From observational data we see that players had highly varying relationships with questions, specifically regarding their level of comfort. In our early playtests with V5 and V8 in the lab and in the field, players were permitted to ask as many questions as possible within the given time limit. While some players took advantage of this and asked questions in a rapid-fire fashion, we saw some players that asked very few or no questions. These players instead seemed to be deep in thought or too nervous or uncomfortable to ask any questions aloud. In an attempt to ensure that every player had the opportunity and motivation to ask questions, in V9 and beyond we distributed battery tokens so that each player was allotted a particular number of questions they could ask. This limited the questions that the more comfortable students could ask and incentivized the less comfortable students to ask questions.

In V9 of the game, we also implemented the question templates. By asking players to fit their questions to the template, we hoped to support players who were overwhelmed by the task of coming up with a question, as well as diversify the questions being asked by players. During game play, we did not strictly enforce that players fit their questions to the template – partly so that students would not feel increased self-consciousness or discomfort with question-asking, and partly because it is logistically difficult for the robot player to check the templates while attempting to answer questions within the timed round. In our analysis of the question data, we examine how closely players matched the given templates when asking questions. In our analysis, only about half of the questions asked perfectly matched the template given. Twelve of the 159 questions across the six game plays used no discernable template at all (i.e., the questions could not be retrofit into any of the existing templates).

The battery tokens were randomly distributed on each round, but we recorded an uneven distribution of usage of the battery token templates across game plays. Of all of the 20 question templates, by far question template Q1, “Is there a _____ ?,” was the most frequently used, with 25 uses over the four plays of V9. By comparison, the next most frequent template, Q4, “____ need ____ ____?,” had 19 uses across those game plays. By contrast, Q20 “When ___ a ____ ____?”, Q19 “_____ _____ the most _____?”, Q7 “How much _____?”, and Q6 “Does the room ____ _____?” all had two or fewer uses.

We observed an increase in the average number of questions asked from V8 with 24 questions per game to V9 with 33 questions per game. This may be taken as an indication that students’ comfort with questions may have increased. However, we must also note that because these data come from repeated game play (albeit with different versions of the game), this pattern may simply have resulted from students’ increased level of comfort and familiarity with the game as a whole.

Finally, we observed differences in question-asking behavior and question content when a question-asking round immediately followed a prior failed round versus a prior succeeded round. Removing all first rounds of question asking, we compared post-success and post-failure questions. In post-success rounds of question asking, questions coded as “building off” were three times more frequent than in post-failure rounds. Similarly, questions coded as “characteristic,” where players ask about a feature of something they have previously discovered, were three times more likely in post-success rounds than in post-failure rounds. Finally, we observed that questions coded as “discovery” were twice as likely in post-failure rounds. These question-asking patterns indicate that when players succeed, they are more comfortable building specific hypotheses and learning more about these hypotheses. In post-failure rounds we see more exploratory behavior, with players prioritizing the pursuit of greater breadth rather than greater depth of information.

Design Lessons

Questions can serve multiple simultaneous roles in supporting and expressing curiosity. Questions are a common tool for reducing knowledge gaps, which is why we centered them as a mechanic for *Outbreak*. However, questions also carry with them implicit hypotheses about the gap the players perceive. Even when players cannot articulate their hypotheses explicitly, they voice them in their questions. Because questions are spoken publicly, they help the group perform collective knowledge assessment; players know what other players are uncertain about, and what they think is worth asking. Finally, because answers are also given publicly, questions help players *help each other* reduce information gaps, not just reduce them for themselves. Even in games where questions are not core to the mechanic, creating moments where question-asking is both encouraged and visibly rewarded can create safe social environments to express curiosity.

Empowering quieter players supports the entire group's efforts to express curiosity. Designs that enforce that all players participate support the entire group in expressing curiosity, without impairing the performance of individuals. As we saw in *Outbreak*, when we switched from a free-form question-asking phase to a structured one where each player was given battery tokens, we witnessed an increase in the average total number of questions the entire group asked. There was both an increase in fluency and better distribution of question-asking amongst players. In other games that require creative participation, enforced participation might temper the influence of an "alpha player" and help the entire group.

Flexibility in enforcing rules fosters curiosity. When players are trying to reduce a knowledge gap, they are sensitive to their ability to effectively use the tools available to them, including questions. Rejecting attempts to close the knowledge gap for violations of minor rules was counterproductive. As we observed in *Outbreak*, the question templates on battery tokens were used loosely. Players typically asked questions that were a close, but not an exact, match. While the robot player rejected questions that had nothing to do with the proffered template, accepting the close-but-not-quite questions helped support player enthusiasm for and fluency with questions. By not formalizing the degree of acceptable deviance into rules, but rather leaving it up to the player's judgment, robot players can implicitly respond to group social norms.

CONCLUSION & FUTURE WORK

This paper explores how game design decisions influence two critical elements of curiosity: the affective experience of failure and question-asking as a method for closing information gaps. In this paper, we present a design model of curiosity that articulates the relationship between uncertainty and curiosity, and defines the role of failure and question-asking within that relationship. We explored ways to instantiate failure and question-asking within a cooperative board game, playtested repeatedly with players in our target demographic, and investigated the impact of game design decisions on their affective experiences of failure and their ability to use questions to close information gaps. We found that affect had a significant experience on players' in-game decisions around risk and failure, as well as on their willingness to express ignorance and take risks socially; players' affective experiences were in some ways more responsive to aesthetic, narrative, and contextual factors than to changes in mechanics. Conversely, changes in game mechanics changed how groups managed their question-asking process, and served to empower quieter players without silencing bolder ones – but flexibility in enforcing the rules and mechanics of the game was key. Designing for curiosity involves a balancing act; when designers can create motivating moments of uncertainty, give players opportunities to face that uncertainty, and equip them with the right tools to resolve that uncertainty they can create positive cycles not only of curiosity but of rich engagement with their games.

Our work to date has studied these questions through iterative design and playtesting with members of our target demographic, middle-school students with marginalized science identities. Our findings can now be used to design larger-scale studies, and to test whether our insights generalize to other audiences. One avenue of future research with *Outbreak* will be to study how the gameplay behaviors and outcomes we observed play out in groups of varying composition, allowing us to understand how factors such as the social and interpersonal dynamics of the group influence players' experiences. In future studies, we can also look at the moment-to-moment processes by which failure and question-asking are constructed in player groups to understand our findings more deeply. For example, the literature on questions indicates that the process of developing questions is as important as the questions themselves. Finally, we can study how our findings can be instantiated in other games, whether explicitly designed to support curiosity or not.

Considering the generalizability of these lessons to other game genres and platforms raises a number of intriguing questions for further consideration and future study. First, how might group processes related to failure, question-asking, and curiosity emerge differently in cooperative games versus competitive games? Second, to what extent is the physical co-location of players in tabletop multiplayer games necessary for producing the outcomes we observed with *Outbreak* (e.g., how critical is the role of nonverbal responses such as facial expression)? Finally, comparing multiplayer to solo game experiences introduces the question of how essential the co-presence of (and/or collaboration with) other players is for producing the affective and behavioral responses that emerged with *Outbreak*. Perhaps appropriately given the topic of this paper, we look forward to exploring these questions in the future.

ACKNOWLEDGEMENTS

Thanks to the SCIPR (Sensing Curiosity in Play and Responding) project and co-PIs Justine Cassell, Jessica Hammer, and L.P. Morency, generously funded by the Heinz Family Foundation. We also thank Anny Fan and Catherine Kildunne for their work in iteratively designing and playtesting *Outbreak*.

BIBLIOGRAPHY

Barrett, L. F. “Feelings or words? Understanding the content in self-report ratings of experienced emotion.” *Journal of Personality and Social Psychology*, *87* (2), 266. Washington D.C., 2004.

Barsade, S. G. “The ripple effect: Emotional contagion and its influence on group behavior.” Administrative Science Quarterly, 47(4), 644-675. 2002.

Berlyne, D. E. “Curiosity and exploration”. in Science vol.153, pp. 25-33. 1966.

Carlin, K. A. “The impact of curiosity on learning during a school field trip to the zoo” (Doctoral dissertation, University of Florida). *Dissertation Abstracts International*, *60*. 1999.

Costikyan, G. “Uncertainty in games.” Boston: MIT Press, 2013.

Csikszentmihalyi, M. “Toward a psychology of optimal experience”. Springer Netherlands, 2014.

Engel, S. The Case for Curiosity. in Educational Leadership, vol. 70, no. 5, pp. 36-40. 2013.

Feldman, D. C. “The development and enforcement of group norms.” in *Academy of Management Review*, 9, no. 1, pp. 47-53. 1984.

Gee, J. P. “What video games can teach us about literacy and learning.” New York: Palgrave-Macmillan, 2003.

Hanington, B., and Martin, B. “Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions.” Rockport Publishers, 2012.

Hunicke, R., LeBlanc M., and Zubek R. “MDA: A formal approach to game design and game research.” In Proceedings of the AAAI Workshop on Challenges in Game AI, vol. 4, no. 1. 2004.

Jirout, J., & Klahr, D. “Children’s scientific curiosity: In search of an operational definition of an elusive concept.” in Developmental Review vol. 32, no. 2 (Jun 2012), pp.125-160. 2012.

Juul, J. “The art of failure: An essay on the pain of playing video games.” Boston: MIT Press, 2013.

Kelley, D. J. “Modeling Emotions in a Computational System.” In *Google It*, pp. 447-461. Springer New York, 2016.

Klopfer, E., Osterweil, S., & Salen, K. “Moving learning games forward: Obstacles, opportunities, & openness.” The Education Arcade: MIT, 2009.

Litman, J. A., & Jimerson, T. L. “The measurement of curiosity as a feeling of deprivation”. in Journal of Personality Assessment vol. 82, no. 2, pp.147-157. 2004.

Loewenstein, G. “The psychology of curiosity: A review and reinterpretation.” in Psychological bulletin vol. 116, no. 1 (Jul 1994), pp.75-98. 1994.

Mahdikhani, M., Soheilahamzehloo., Maryamshayestefard., and Mahdikhani, N. “Student participation in classroom discourse.” vol.5, no. 6, pp. 1422–1431. 2008.

Mohammed, S., and Dumville, B. C. “Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries” in *Journal of Organizational Behavior*, 22, no. 2, pp. 89-106. 2001.

Proulx, T., & Inzlicht, M. “The five “A” s of meaning maintenance: Finding meaning in the theories of sense-making.” in Psychological Inquiry vol. 23, no. 4, pp.317-335. 2012.

Rinkevich, J. L. “The relationship among student creativity, curiosity, and academic intrinsic motivation: A mixed methods phenomenological study of sixth grade students” (Doctoral dissertation, Indiana University of Pennsylvania). 2014.

Rocca, C. H., Krishnan, S., Barrett, G. and Wilson, M. “Measuring pregnancy planning: An assessment of the London Measure of Unplanned Pregnancy among urban, south Indian women.” Demographic research, vol. 23, p.293. 2010.

Schell, J. “The Art of Game Design: A book of lenses.” CRC Press, 2014.

Thelen, E., and Smith, L. B. “A dynamic systems approach to the development of cognition and action.” Boston: MIT Press, 1996.

To, A., Ali, S., Kaufman, G., Hammer, J. “Integrating Curiosity and Uncertainty in Game Design” in Proceedings of DiGRA/FDG ‘16. 2016a.

To, A., Fath, E., Zhang, E., Ali, S., Kildunne, C., Fan, A., Hammer, J., Kaufman, G. “Tandem Transformational Game Design: A Game Design Process Case Study”, 2016 Meaningful Play. 2016b.

To, A., Fan, A., Kildunne, C., Zhang, E., Kaufman, G., Hammer, J. “Treehouse Dreams: A Game-Based Method for Eliciting Interview Data from Children.” In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 307-314). ACM. 2016c.