A Data-Driven, Multidimensional Approach to Hint Design in Video Games

Helen Wauck, Wai-Tat Fu
University of Illinois at Urbana-Champaign
Urbana, IL, USA
{wauck2, wfu}@illinois.edu

ABSTRACT

Hint systems are designed to adjust a video game's difficulty to suit the individual player, but too often they are designed without analyzing player behavior and lack intelligence and adaptability, resulting in hints that are at best ineffective and at worst hurt player experience. We present an alternative approach to hint design focusing on player experience rather than performance. We had 25 participants play a difficult spatial puzzle game and collected player behavior, demographics, and self-reported player experience measures. We found that more exploratory behavior improved player experience, so we designed three types of hints encouraging this behavior: adaptive, automatic, and on-demand. We found that certain players found hints more helpful regardless of whether the hints changed their behavior, and players seemed to prefer seeing fewer hints than the adaptive and automatic conditions gave them. Our findings contribute a deeper empirical understanding of hint design strategies and their effect on player behavior and experience, with practical recommendations for designers of interactive systems.

ACM Classification Keywords

H.5.1 Multimedia Information Systems: Artificial, augmented, and virtual realities

Author Keywords

Hint systems; player experience; player behavior; video games; hints; hint design; game design

INTRODUCTION

Hint systems are prevalent in many commercial, educational, and serious games and are designed to tailor the game's difficulty to suit individual players' needs. Hint designers have used a variety of different methods for generating and delivering hints. Some hint systems give hints on-demand or require the player to "earn" them by making progress, while others analyze player behavior to design adaptive hints that trigger automatically when the player needs help. Hint content can also vary from abstract to concrete. Research on hint systems has studied several different combinations of these hint generation and delivery techniques, but results thus far do not present a clear picture of how different ways of generating and delivering hints affects player behavior and experience in video games.

Even though hints are already widely used in commercial games, existing research focuses mainly on educational games and draws heavily on research from the domain of Intelligent Tutoring Systems (ITS) to inform hint design decisions. When hints are designed this way, the results are a mixed bag. Sometimes hints added to educational games have positive effects such as improved player engagement or learning gains in a specific content area, sometimes they have no effect at all, and sometimes they even have negative effects on player performance and engagement [3, 10, 18].

Part of the problem seems to be that hint design and evaluation typically focuses on only a subset of relevant metrics that do not capture the whole picture. Designing and evaluating hints with the goal of merely increasing players' win rate or the amount of time they spend playing runs the risk of designing hints that miss the central goal of successful games: an enjoyable player experience. Did the player have fun? Why did they have fun? Was it because the game challenged them just the right amount or because they found the content interesting? What kept them playing? These questions cannot be answered without measuring multiple dimensions of player experience; measuring performance or time spent playing alone is not enough [14].

To address this problem, we adopt a broader perspective to study hint design and evaluation in video games that brings together three key metrics of successful game design - player performance, engagement, and subjective experience - that up until now have been studied only in isolation in hint system research. Our approach simultaneously analyzes player experience, engagement, and performance in order to understand the complexities underlying players' reaction to hints.

Our contributions in this work are threefold. First, we present a new data-driven method for designing and evaluating hints that captures the underlying complexities of player behavior and experience and how they are affected by different types of hints in a video game. Second, we provide empirical findings

IUI 2017, March 13-16, 2017, Limassol, Cyprus

DOI: http://dx.doi.org/10.1145/3025171.3025224

from a case study using this approach. Our findings offer deeper theoretical insights into the latent dimensions of player engagement, performance, and experience that are affected by hint interventions. Finally, we suggest practical hint implementation recommendations for game designers based on our findings.

BACKGROUND AND RELATED WORK

We briefly review the history of hint design and evaluation in two types of interactive systems: Intelligent Tutoring Systems and video games, and discuss how our approach to hint design and evaluation differs from and extends previous work.

Hint Design Approaches in Intelligent Tutoring Systems

Hints are an essential component of Intelligent Tutoring systems (ITS), which allow them to adapt to different student behaviors and scaffold the learning experience for individual students, resulting in improved learning [4, 27] across a diverse field of different subjects, such as computer programming [16], physics [23], and algebra [15]. Research on hint systems in ITSs has shown that students benefit most by receiving hints only when they really need them. Most hint systems in ITSs simply give assistance only when the student requests it and scaffold hints such that they progress from more abstract and symbolic to more concrete and specific [27]. Unfortunately, this approach has the downside of allowing students game the system by clicking through the hints rapidly to get to the lowest level "bottom-out" hint, which simply gives them the answer and hinders learning [2].

Researchers have investigated various ways of circumventing this problem by implementing a hint system that detects when a student needs help and automatically triggers a hint at the appropriate time with the appropriate content for the particular problem the student is having. Murray et al implemented adaptive hints in the DT Tutor, which triggered only when a student became frustrated or stopped trying. However, this solution was too computationally expensive and required access to too much student data to be practically applied in most real-world scenarios [17]. Other work has looked at automatically generating the actual content of hints. Paquette et al, for instance, developed ASTUS, an authoring framework for automatically generating next-step hints in a cognitive tutor, and showed that hints generated using this approach could be as effective as instructor-generated hints [19].

Inspired by hint design research in ITSs, researchers have studied how to apply the techniques used for ITS hint design in a video game context, although this field of research is still fairly new. In the next section, we discuss the current state of hint design research in video games.

Hint Design Approaches in Video Games

In recent years, researchers have adopted ITS hint design techniques for educational video games. Conati et al implemented scaffolded, intelligent hints triggered by player behavior in the math game *Prime Climb*. They found that children who received the hints had larger math learning gains than those in a control group who received no hints [10], suggesting that hints have promise as a tool to improve learning in educational games. A follow up study by Conati et al with the same game introduced a new metric, attention to hints, and demonstrated its interaction with player performance, the timing of hint presentation, and players' attitude towards receiving help [9]. The interrelatedness of these four variables indicates the complexity underlying player experience with hints.

O'Rourke et al investigated the effect of different types of hints on player performance in the children's math game *Refraction*. They implemented four hint types in a 2 x 2 design: concrete versus abstract and hints that were either earned with progress or embedded in the game environment. Concrete hints were more helpful to players than abstract hints in the sense that they helped players win more levels of the game. Surprisingly, both concrete hints and abstract hints negatively affected player performance in *Refraction* regardless of whether they were earned or embedded [18]. However, this study evaluated hints on the basis of a single metric: player performance, as measured by the number of levels the player won. It is not clear how hints affected other dimensions of the player's experience that might also be important, such as player engagement.

In a study across three different computer games, Andersen et al also implemented hints in educational games, but used player engagement, a key aspect of successful game design [13, 1], as a more generalizable metric for evaluating them. In this study, player engagement was measured by number of levels completed, time spent in the game, and the rate at which players returned to play again, and hints were implemented in an on-demand fashion in the context of a tutorial. Andersen et al implemented hints for three different games: *Refraction* (the same game used by O'Rourke et al), the platformer *Hello Worlds*, and *Foldit*, a 3D protein folding puzzle game. On-demand hints improved player engagement in *Foldit*, the most complex of the three, but had no effect on engagement in *Hello Worlds* and a negative effect in *Refraction* [3]. This mixed result suggests that game complexity and possibly game genre play an important role in hint effectiveness.

In addition to player performance metrics like winning and player engagement metrics like time spent playing, researchers have also investigated self-reported player experience measures. To evaluate an adaptive computer game AI that used player eye tracking to adjust difficulty level, Wetzel et al asked players to give a self-assessment of their experience along such dimensions as fun, frustration, and level of challenge [28]. In another study, Denisova and Cairns measured players' self-reported *immersion*, the degree to which players feel involved in a game, in an isometric shooting game. They analyzed how immersion was affected by the presence of an adaptive difficulty system that adjusted an in-game timer based on the player's performance [11]. These self-reported measures can add increased authenticity and validity to hint evaluation.

Individual Differences and Hint Design

Hints research in both the video game and ITS domains also points to the need to tailor the game experience to individual player differences. Pereira et al used hints to model player personality traits in a spatial strategy game and found that the personality trait *need for cognition*, related to a person's willingness to engage in cognitive activities, was negatively correlated with the number of hints the player followed in the game [20]. This suggests that individual player differences in personality play an important role in how players respond to hints.

A study by Arroyo et al found that gender and cognitive ability affected what kind of hints were most effective for children in a mathematics tutor. Girls preferred highly interactive hints, whereas boys preferred simpler text-based hints. They also found that children with higher cognitive abilities performed better with abstract, symbolic hints, whereas children with lower cognitive abilities did better with concrete hints [5]. A more recent, large-scale study confirmed these gender differences and revealed additional gender differences in willingness to seek help and in affective and cognitive responses to an in-game pedagogical agent [6]. Therefore, it is important to take into account multiple axes of individual differences in order make hints beneficial for as many players as possible.

Our Approach

Given the mixed results studies on video game hinting systems have shown so far, ranging from positive to neutral to even negative effects, it seems that hints have a complex effect on players that is still not well understood. Our approach to analyzing hint effectiveness aims to capture these underlying complexities by extending prior work in two main ways: incorporating more complex, multidimensional analysis metrics and analyzing individual player differences in hint effectiveness.

Every hint or adaptive system evaluation study in video games that we are aware of has focused on only a small subset of possible evaluation metrics. Conati et al used a pretest-posttest structure to assess learning gains as a result of adding hints to a math game [10, 9]. Thus, hint evaluation was based solely on the game's ability to teach specific math concepts, a measure not generalizable to other types of educational and non-educational games. O'Rourke et al measured hint effectiveness using player performance [18], which has the advantage of being generalizable to many other types of games. However, both of these studies only used a single evaluation metric, which may not fully capture the underlying complexities of players' experience with hints. For example, hints could improve player performance or learning of target concepts by making the game too easy, which might be boring for players and make them less likely to keep playing.

Other prior work analyzed the effectiveness of hints and other adaptive features with multidimensional measures. Andersen et al evaluated on-demand hints using three different behavioral measures of player engagement [3], and Wetzel et al analyzed three different self-reported emotional measures of player experience [28]. These multidimensional metrics provide a more nuanced representation of players' response to adaptive game features. However, analyzing behavioral measures alone requires that explanations for player behavior be inferred, which may lead to misinterpretations of complex player behavior. On the other hand, analyzing self-reported measures alone may be difficult because different players may have different ideas about what affective measures of experience mean in a gaming context. For instance, is frustration always a negative experience in a game? Is a low level of challenge always better?

Denisova and Cairns partially addressed this limitation by combining a self-report measure, immersion, and a behavioral measure, performance, in their evaluation of an adaptive difficulty adjustment system [11]. However, these two measures still may not fully capture other important dimensions of player experience, such as affective response (fun, frustration, boredom) or how different measures of player experience interact with each other.

Another issue that has not been addressed in previous work is the effect of individual player differences on the effectiveness of hint systems in video games. As the studies by Pereira Santos et al and Arroyo et al with ITSs showed, a variety of individual differences between users, such as gender, need for cognition, and cognitive ability, can affect how people respond to hints and what kind of hints they respond the best to [5, 6, 20]. This variation may explain some of the mixed results in studies of adaptive video game features to date.

We present a new approach to hint design and evaluation that uses combined analysis of self-reported player experience, engagement, performance, and individual player demographic information to present a more complex, accurate picture of the impact of different types of hints on different types of players. This approach avoids oversimplifying the complex notion of what makes a video game hint successful.

METHOD

We conducted two online experiments over the course of several weeks. In the first experiment, *Strategy Discovery*, participants played a very difficult puzzle game while we collected data about player behavior, demographics, and self-reported player experience that would help us design effective data-driven hints for the game. We then deployed three different types of data-driven hints (adaptive, on-demand, and automatic) in a new version of the game and ran a second study, *Hint Evaluation*, collecting additional player behavior, demographic, and experience data to assess how our data-driven hints affected player experience, persistence, and performance.

The Game

The puzzle game used in our study was a reproduction of a notoriously tricky puzzle minigame known as "The Master Sword Puzzle" in the popular commercial game, *The Legend of Zelda: Twilight Princess*. We chose this particular game for our study for two reasons. First, its notorious difficulty often caused players to give up and look up a solution online in order to progress - any google search of "Twilight Princess Master Sword Puzzle" or similar terms will confirm this. We felt that the Three Body Puzzle's propensity to produce high levels of frustration and failure indicated that there was room for improvement and players could benefit from in-game hints. Second, there are 64 different solutions to the puzzle, so it presented an opportunity to design hints that go beyond simply giving the player a specific step towards a solution. For the sake of clarifying what the puzzle actually is, we will refer to it as the Three Body Puzzle in this paper.

The design of the Three Body Puzzle is shown in Figure 1. The player is represented by a blue arrow whose movements are synced with two other red arrows. The player must move

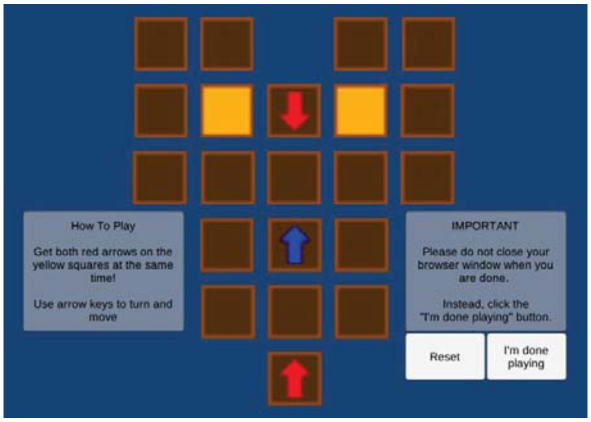

Figure 1. The Three Body Puzzle. The movements and turns the player (blue arrow) makes are synchronized with those of the two red arrows, and the player must use this synchronization to get both red arrows onto the yellow squares at the same time.

around the game board until the two red arrows arrive at the two yellow board squares at the same time. We implemented the game using the Unity Game Engine and the Unity WebGL build for deploying the game online. We hosted the puzzle game online in order to recruit a large set of demographically diverse participants and to facilitate easy collection and aggregation of player behavior and experience data.

Procedure

Figure 2 summarizes our study procedure. Our study consisted of two experiments: *Strategy Discovery* and *Hint Evaluation*. Three different types of hints were designed using data from the *Strategy Discovery* experiment and implemented for the *Hint Evaluation* experiment.

In the first experiment, *Strategy Discovery*, our goal was to use the large amount of data we collected about players' demographics, in-game behavior, and self-reported player experience during gameplay to design adaptive hints for the Three Body Puzzle that would improve overall player experience. We used the data collected from this experiment to find patterns in player behavior associated with a better player experience (more fun, more easiness, less boredom and frustration) and used these behaviors to design data-driven hints. In the second experiment, *Hint Evaluation*, we collected the same data about demographics, behavior, and player experience as in Strategy Discovery. We then evaluated the effect of the hints we designed by comparing self-reported player experience and in-game behavior between the *Strategy Discovery* and *Hint Evaluation* phases.

At the beginning of each experiment, we publicized our study by posting flyers with a link to the study website around campus at a large Midwestern university and off campus at public buildings in the surrounding town. We also advertised the study online via Facebook, email lists, and the r/samplesize subreddit. There were no restrictions on who could participate in the study other than that they had to be adults 18 years or older, and no compensation of any kind was offered to participants, who remained anonymous throughout the study.

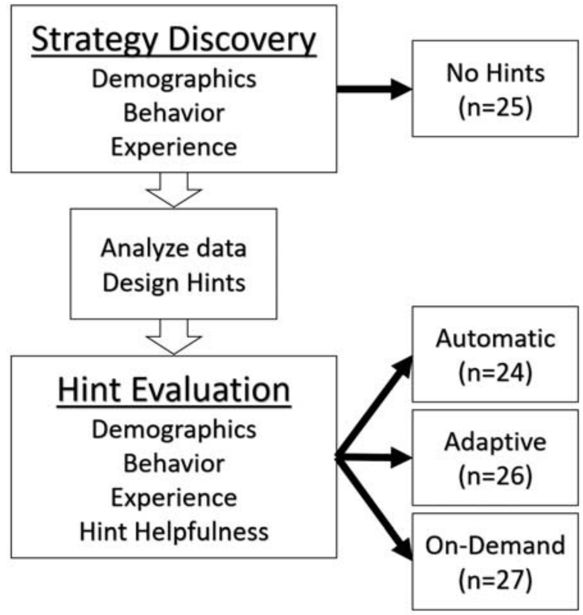

Figure 2. A summary of our experimental procedure.

As soon as participants accessed the study website, they completed an informed consent form and certified that they were 18 years or older before they could proceed. Participants then filled out a short pre-game demographic survey about their age, gender, prior video game experience, and their preferred games or game genres. Upon completing the survey, each participant played the Three Body Puzzle. In the *Hint Evaluation* phase, participants were randomly assigned to one of the three hint conditions: Automatic, Adaptive, and On-Demand. Our implementation allowed players to "reset" the game back to its starting state if they made an error (consistent with the design of the original game) and provided the option to play again after a victory. There was no time or move limit for the game, and players could reset or play again as many times as they wanted before quitting their play session. During gameplay, the online platform recorded player behavior data. Once they decided to stop playing, participants were asked to fill out a post-game survey about their experience playing.

Measures

We collected three types of information from each participant in our study: demographics, self-reported player experience, and player behavior. Table 1 gives a summary of the specific metrics we collected for each of these categories. Demographic information collected in the pre-game survey included gender, age, frequency of video game use (1=no experience, 6=plays daily), and whether or not participants had previously participated in this study (to control for learning effects). We measured player experience by collecting participants' self-reported impressions of how fun, boring, easy, and frustrating they found the game on a 5-point Likert scale in the post-game survey.

The game also collected a set of low-level player behavior metrics, which focused on the player's exploratory behavior, persistence, and performance. We felt that exploratory behavior was the most relevant dimension of player behavior to focus on for the Three Body Puzzle due to the large number of possible solutions (64) to the Three Body Puzzle. Designing more traditional hints that suggest specific, concrete steps toward a solution might restrict the player's creativity, an important part of the overall player experience in a game [25]. On the other hand, exploration of the "problem space" has been shown to enhance visual designers' creativity [12], and might therefore also be useful for enhancing player creativity.

We hypothesized that more exploratory and varied movement patterns across the game board would encourage players to be more creative, generating more new ideas about what to try next and reducing the likelihood that they would get stuck and give up, which in turn would result in a better player experience. New squares explored per move was our main metric of exploratory player behavior. New squares explored per move was calculated by summing up the number of unique squares explored in each of a player's attempts and dividing by the total number of moves the player made across all attempts in order to control for the number of attempts and number of moves the player made.

We also measured player persistence because it is generally considered a good proxy measure for player engagement and would help us understand the relationship between self-reported player experience and in-game player behaviors [8]. We used two different metrics for player persistence: total time spent playing and total number of attempts the player made. Since participants in our study were not compensated for playing the game and were anonymous, and therefore had no obligation to keep playing the game if it did not hold their interest, we felt the time and effort they put into the game were reasonable proxies for player engagement and motivation. Total number of attempts was calculated by counting the number of times that the player either decided to play again after winning (clicked the Yes button in response to the question, "Do you want to play again?") or decided to reset the game back to its starting configuration (clicked the "Reset" button at any time during gameplay).

We measured player performance because we hypothesized that players who won would have more fun due to the satisfying effects of mastery and achievement [24], O'Rourke et al measured player performance in terms of the number of levels completed in *Refraction* since players could not actually "lose" this game [18]. Similarly, in the Three Body Puzzle, players only "lost" the game in the sense that they gave up and quit before discovering the correct solution. Therefore, we measured whether or not players of the Three Body Puzzle were able to eventually find the correct solution to the puzzle (win), regardless of how many attempts they took to do it.

EXPERIMENT 1: STRATEGY DISCOVERY

In the *Strategy Discovery* experiment, participants played our implementation of Three Body Puzzle with no hints. In total, we collected data from 25 players for this phase. We excluded one player from our dataset who made one attempt at the game with zero moves, and further filtered the data to exclude all other attempts where zero moves were made (13 attempts in total), since this indicated that the player was not making a serious attempt or was experiencing some technical issue with the game. This left us with $n = 24$ players for analysis. There were 10 female participants, 12 male, and 2 marked other or left gender blank. Player ages ranged from 18-30 ($Median = 24$), and 79% reported playing video or computer games at least occasionally. All players indicated that this was their first time playing the game. Analysis in this phase focused on how player behaviors influenced player experience.

Demographics	*Behavior*	*Experience*
Gender	New squares explored per move	Fun (1 to 5)
Age	Time spent playing	Easiness (1 to 5)
Game experience (1 to 6)	Number of attempts	Boredom (1 to 5)
Previously participated?	Did player eventually win?	Frustration (1 to 5)

Table 1. Data collected in Strategy Discovery and Hint Evaluation experiments. During Hint Evaluation, we also asked players to rate hint helpfulness on a 5-pt Likert scale.

Hypotheses

We had three main hypotheses in the *Strategy Discovery* phase:

H1a: Players who win the game will report a better player experience (more fun, less boredom, less frustration).

H1b: Players who have a better player experience will persist longer (more time spent playing, more attempts).

H1c: Players who exhibit more exploratory behavior will have a better player experience (more fun, less boredom, less frustration).

Relationship Between Player Experience Variables

First, we examined the relationship between different in-game behaviors and players' self-reported assessments of how fun, boring, easy, and frustrating they found the game. We performed correlation analysis using Spearman's ρ. Self-reported levels of fun emerged as our most salient measure of player experience since fun was significantly correlated with less boredom ($\rho = -0.61$, $p = 0.0015$) and marginally correlated with more easiness ($\rho = 0.35$, $p = 0.09$) and less frustration ($\rho = -0.39$, $p = 0.059$). These results suggest that the best way to improve player experience with hints may be to focus on reducing boredom, increasing easiness, and reducing frustration.

Player Behaviors Associated with Positive Experience

Next, we performed a correlation analysis between player experience and behavior. Significant and marginal results are summarized in Table 2. Contrary to **H1a**, eventually winning the game was not associated with more fun. However, eventually winning was associated with more easiness and less frustration, so player performance may still be relevant to player experience in this game. Our hypothesis that a more positive player experience would increase time spent playing and number of attempts (**H1b**) was also not supported; these behaviors were not associated with any of our four dimensions of player experience. New squares explored per move was

Behavior	*Fun*	*Easiness*	*Boredom*	*Frust.*
Square exploration	0.36 p=0.083	0.44 p=0.03	n.s.	-0.43 p=0.035
Time spent playing	n.s.	n.s.	n.s.	n.s.
Number of attempts	n.s.	n.s.	n.s.	n.s.
Winning	n.s.	0.55 p=0.006	n.s.	-0.44 p=0.03

Table 2. Summary of correlation analysis between player behaviors and experience in the Strategy Discovery experiment. NS = no significant correlation.

significantly correlated with more easiness, less frustration, and marginally correlated with more fun, supporting our hypothesis that exploring more new squares is associated with a better player experience (**H1c**).

From this analysis, new squares explored per move emerged as the player behavior with the most potential to affect player experience. Hints encouraging the player to explore more new squares could improve player experience by reducing frustration, making the game easier, and increasing fun.

Hint Design

Since new squares explored per move was associated with more fun, more easiness, and less frustration, we designed our hints to encourage players to explore more new squares throughout the game.

To determine what specific suggestions to make to players in the hint text and at what point in the game it would be most beneficial to display the hints, we examined how many new squares players generally needed to explore to maximize their levels of fun, our most important measure of player experience. For fun levels 1-3 ($n = 19$), the median value for new squares explored was 0.4, while for fun levels 4 and 5 ($n = 5$), median new squares explored per move was 0.6. We used the median new squares explored value for fun levels above 3 as our criterion for when to display hints and what amount of new square exploration to suggest to players. We designed adaptive hints to display automatically every time the player's new squares explored per move level dropped below 0.6, unless they have already explored all of the board squares, in which case hints would no longer be displayed.

Since repeated, frequent interruptions hurt performance on complex tasks [26], such as the difficult Three Body Puzzle, we decided to only display the hint every 15 moves. Since the hint triggered every 15 moves if the player's new squares per move dropped below 0.6, we designed the hint to advise players to try exploring about 10 new squares over the next 15 moves since this corresponded to a rate of new square exploration just above 0.6, our threshold ($0.6 \times 15 = 9$). To further assist the player in exploring new squares, we had the hint highlight the specific squares on the game board that the player had not yet explored while the hint text displayed. Figure 3 shows the specific text of the hint as well as unexplored square highlighting.

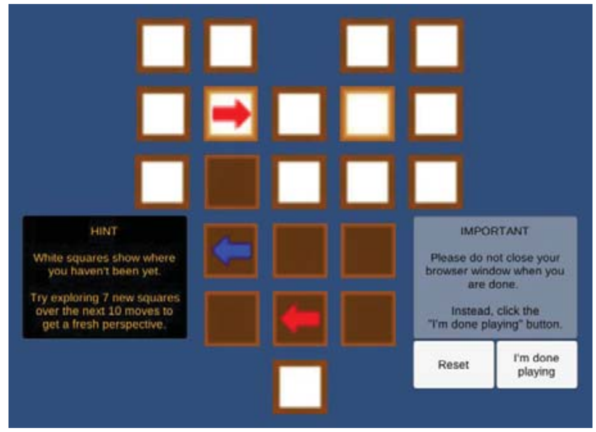

Figure 3. Hint design for the Three Body Puzzle. Squares the player has not stepped on yet are highlighted in white, while hint text (black box) suggests exploring more of these squares.

However, we felt that designing a single type of hint for all players was too simplistic an approach given the significant amount of individual variation between players in terms of new squares explored per move, number of attempts, time spent playing, and self-reported player experience measures. These variables likely have complex interactions since player experience itself is a complicated, multi-dimensional measure with many different potential contributing factors. Therefore, we decided to implement two additional variations on our adaptive hint (on-demand and automatic) to see if different ways of triggering the hint would be preferred by different types of players.

On-demand hints were accessed whenever the player desired via a button added to the interface labeled "Get Hint." Like adaptive hints, automatic hints displayed every 15 moves, but unlike adaptive hints, they displayed regardless of the player's squares explored per move rate. Each time the automatic and adaptive hints display, player control is disabled for 3 seconds to give the player time to notice and process the hint before moving, which hides the hint again.

Before launching the next data collection phase with the implemented hints, we added a hint counter to our in-game data collection and added a new question to the post-game survey that asked players to specify how helpful they found the hints on a 5-point Likert scale (1=not helpful at all, 5=very helpful) and why. Finally, we asked players whether they had actually seen any hints as a manipulation check, enabling us to cross-check their answers with the number of hints the game actually displayed.

EXPERIMENT 2: HINT EVALUATION

In our second experiment, we deployed a new version of the Three Body Puzzle to the study website and advertised the study again to attract new players. Players in this experiment were randomly assigned to one of three conditions, one for each hint type we implemented: adaptive, on-demand, and automatic. In total, we collected data from 90 players. We excluded data from 12 players who did not indicate that this was their first time playing the game and one player who ran into technical problems with the controls, leaving us with 77 observations. Of those, 48% (37) were female, 51% (39) were male, and one marked "Other" for gender. Ages ranged from 18-68 ($Median = 24$), and 77% reported playing video or computer games at least occasionally. In total, 11 players reported not seeing or not using any hints. Two players reported not seeing any hints but received 9 and 3 automatic hints, respectively. We also removed one attempt where the player made zero moves. Combined with the 24 observations from the *Strategy Discovery* experiment, we analyzed data from 101 players total (24 automatic, 26 adaptive, 27 on-demand, and 24 from the *Strategy Discovery* experiment, who saw no hints).

Hypotheses

We had two main hypotheses for the *Hint Evaluation* phase:

H2a: Players in all three hint conditions will have a better player experience (more fun, less boredom, less frustration) than players in the *Strategy Discovery* experiment.

H2b: Players in the Adaptive hint condition will have a better player experience (more fun, less boredom, less frustration) than those in the Automatic and On-Demand conditions.

Effect of Hints on Player Experience and Behavior

Contrary to both **H2a** and **H2b**, we found that no player experience metrics or player behaviors (fun, frustration, easiness, boredom) were associated with the presence of hints or any of the hint types. In addition, 54 out of the 66 players who saw hints (81%) rated the hints 1 or 2 out of 5 (not helpful at all or only a little helpful). Thus, hints did not seem to improve player experience overall, but neither did they seem to make it worse. However, players' written explanations of their hint helpfulness ratings revealed that "helpfulness" as a metric had a complex relationship with player experience, which we discuss in more detail in the **Players' Reasons for Hint Helpfulness Ratings** section.

However, an ANOVA revealed a significant difference in the number of hints seen between the three hint conditions ($F(2,74) = 6.55$, $p = 0.002$). Pairwise Tukey tests showed that players in the adaptive and automatic conditions saw about 1 one more hint per attempt ($\mu = 1.61$) than players in the demand condition ($\mu = 0.70$, $t = -3.73$, $p = 0.0004$). This suggests that the number of hints players prefer to see is less than what the two types of triggered hints showed them. Yet this apparently did not affect any measurable dimensions of player experience, behavior, or performance.

Factor Analysis

We performed exploratory factor analysis to uncover potential latent variables related to player experience and behavior that may have been affected by seeing or using hints. The data used for this analysis consisted of the players who were assigned to one of the three hint conditions in the *Hint Evaluation* experiment ($n = 77$). We excluded data from 11 players who reported seeing or using no hints, leaving us with 66 players for the factor analysis. The player experience and behavior variables included in our factor analysis were players' self-reported measures of fun, easiness, boredom, and frustration, new squares explored per move, how many hints they received (or asked for), and how helpful they found the hints. Since

Measure	*Factor 1: Hints Helped*	*Factor 2: Engagement*	*Factor 3: Strategy*
Fun	0.665	0.199	
Easiness	0.641	-0.239	0.165
Boredom	-0.503	-0.147	
Frustration	-0.702	-0.275	
Square exp.	-0.107	-0.131	0.983
Time played	0.217	0.706	-0.245
Num. attempts	0.160	0.960	0.220
Hints per attempt			-0.658
Hint helpfulness	0.363		
SS Loadings	**1.820**	**1.640**	**1.544**

Table 3. Results from factor analysis. Blank cells correspond to very small factor loadings eliminated by varimax rotation.

factor analysis is not appropriate for categorical variables such as gender, whether the player won or not, and which hint condition they were assigned to, we excluded them from this initial analysis.

For factor analysis, we used the R factanal function with varimax rotation, which prioritizes factors with a small number of large loadings in order to draw out the most significant relationships and reduce noise. Scree plot analysis and an empirical χ^2 goodness of fit test indicated that 3 factors provided the best fit for our data while maintaining eigenvalues above 1 for each factor ($\chi^2(12) = 16.62$, $p = 0.16$). The factor analysis is summarized in Table 3. We used regression scores when calculating the value of each factor for each player.

Factor 1: Hints Helped

Factor 1 had a large positive loading on fun (0.67) and easiness (0.64), and moderate to large negative loadings on frustration (-0.70) and boredom (-0.50). In addition, this factor had small positive loadings on time spent playing (0.21) and hint helpfulness (0.36), a very small positive loading on number of attempts (0.16), and a very small negative loading on new squares explored per move (-0.10). There was no significant loading on number of hints per attempt. Thus, players with high levels of this factor seem to have an overall positive experience with the game and tend to persist a little longer, which may be related to finding the hints they receive helpful. We refer to this factor as *Hints Helped*.

Factor 2: Engagement

Factor 2 had large positive loadings for number of attempts (0.96) and total time spent playing (0.71). This factor also had a weak positive loading for fun (0.20) and weak negative loadings for easiness, frustration, boredom, and new squares explored per move (-0.24, -0.28, -0.15, and -0.13, respectively). There were no significant loadings on hints per attempt or hint helpfulness. Therefore, players with high values of this factor seem to persist for a long time in playing the game, and tend to have a somewhat positive game experience, regardless of how many hints they see and how helpful they find the hints. We therefore interpret this factor as *Engagement*.

Factor 3: Strategy

Factor 3 had a large positive loading on new squares explored per move (0.98), a large negative loading on number of hints per attempt (-0.66), a small negative loading on time spent playing (-0.25), and a weak positive loading on number of attempts (0.22) and easiness (0.16). There were no significant loadings for fun, frustration, boredom, or hint helpfulness. Players with high levels of this factor do not need hints to tell them to explore new squares, spend less time in the game but make more attempts, and find the game a little easier. We call this factor *Strategy*.

This analysis suggests that player experience is a more complex concept than our four-dimensional self-report measures of fun, easiness, boredom, and frustration could capture. By uncovering the latent dimensions of player experience and behavior with factor analysis, we were able to discover hidden patterns in our data that indicate certain types of players found hints more helpful than others. More specifically, 11 of the 12 players who rated hint helpfulness at least 3 out of 5 (1=not helpful at all, 5=very helpful) had above average *Hints Helped* levels. However, this factor analysis alone does not tell us much about what kind of players these people were.

Factor Interactions with Categorical Variables

In order to determine how player demographics interact with the latent dimensions of player experience and behavior we discovered with our factor analysis, we performed additional correlation analyses between hint conditions, factors, whether the player won or not, and the demographic variables gender, age, and previous video game experience.

Less Experienced Gamers Show Hint Placebo Effect

Hints Helped was significantly correlated with winning at least once ($\rho = 0.45$, $p = 0.0002$) and marginally correlated with less game experience ($\rho = -0.21$, $p = 0.088$). suggesting that what determines whether a player finds hints helpful or not may simply be the eventual outcome of the game. This may be a kind of placebo effect; if players see hints, they attribute their performance (whether they win or not) to the hints, even if the hints may not have actually helped them change their strategy. Less experienced gamers may be particularly susceptible to the hint placebo effect.

On-Demand Better Than Triggered Hints for Engagement

Engagement was significantly correlated with younger player age ($\rho = -0.24$, $p = 0.048$), and the on-demand hint condition ($\rho = 0.36$, $p = 0.0066$). Thus, on demand hints may actually have a positive impact on engagement for this particular game, possibly due to players' ability to limit the number of hints they see in the on-demand condition. This, together with our finding that players in the on-demand hints condition used fewer hints overall than players in the automatic and adaptive conditions, suggests that game designers may want to consider implementing on-demand hints first before implementing more intelligent, automatically triggered hints. Analysis of how often players use on-demand hints and at what points in the game could provide game designers with better insights about the right time to trigger adaptive hints during gameplay.

Strategy Not Related to Subjective Player Experience

Strategy was marginally correlated with the automatic hint condition ($\rho = 0.23$, $p = 0.051$). No other variables had significant correlations with any of the factors. Players with high Strategy levels may have a natural intuition for exploring new squares quickly, allowing them to explore the problem space more efficiently and recognize situations where they need to reset the game (or situations where they are close to winning). Therefore, these players did not make very many moves per attempt, which would explain why they tended to see fewer hints overall. The fact that *Strategy* is not associated with any aspect of player experience is inconsistent with our initial finding that more new squares explored per move improved player experience. It may be that the relationship between this particular behavior and player experience was not strong enough to be consistent across both the Strategy Discovery and Hint Evaluation experiments.

In summary, the results of our factor analysis uncovered a latent three-dimensional structure underlying player experience in the Three Body Puzzle, demonstrating that player experience was more complicated than even our multiple metrics of fun, frustration, easiness, boredom, persistence, and performance could capture alone. This three-dimensional structure also revealed information about how different types of players reacted to hints. Of particular interest is the fact that certain less experienced players who won the game had a better experience playing and rated hints as more helpful, indicating a hint placebo effect. In order to determine if the placebo effect was a sufficient explanation for these players' positive experience, we turned to players, written explanations of their helpfulness ratings.

Players' Reasons for Hint Helpfulness Ratings

Overall, the 54 players (81%) who did not find the hints very helpful (ratings of 1 or 2 out of 5) felt that the hints had one of the following problems: they stated the obvious, they were irrelevant to solving the puzzle, or their suggestions were difficult or impossible to follow. For example:

> *"The hint was really not stating anything that I didn't already think about."* (P102)

> *"Half the time the unexplored squares would just get you stuck. They often had little to do with getting the right sequence."* (P125)

> *"I was very hard to get to the squares suggested. It seemed as hard as trying to get to the goal itself."* (P176)

Most of the players who rated hint helpfulness a 3 left the explanation field blank, but the 6 who provided an explanation seemed to be using the rating of 3 as a default rating, stating highly variable reasons like *"I don't know"* (P113), *"Not sure if they were useful"* (P187), or even *"didn't help"* (P116). One player had a more complex reason for rating hints a 3:

> *"It told me that something that I hadn't utilized towards the solution, but it didn't make it immediately apparent that it was giving me a solution, just that I simply haven't been there before. Although it wasn't a extremely helpful hint, it is a hint that I appreciate more than a game pushing me towards the answer. To that end, it was actually a very good thing that the hint wasn't extremely helpful!"* (P105)

The only two players who rated the hints' helpfulness a 4 (no one rated helpfulness a 5) both said that the hints helped them solve the puzzle but were vague about exactly how it helped them do so:

> *"I eventually solved it using the hint, it helped me see my way out of patterns and I understood more that I was the blue [arrow] and helped me figure it out."* (P151)

> *"The text was a little confusing but it helped me to solve it."* (P156)

Overall, these results show that players had several significant problems with hints and felt that hints were not very helpful. However, players' reactions to hints reveal the importance of understanding the subtleties of a "hint helpfulness" measure when the goal of hints is to improve player experience and not necessarily to make the game easier. Players seemed to be interpreting helpfulness to mean that the hint gave direct advice about how to reach the solution rather than whether or not the hints actually helped them win. Since the hints were designed to give indirect, more general strategic advice, it makes sense that players would not necessarily know whether the advice actually helped them win or not. In addition, players did not seem to know how to interpret what 3 meant on a helpfulness scale of 1 to 5. The ambiguity and uncertainty of players' given reasons for rating hints' helpfulness the way they did lend support to the notion of a hint placebo effect. Players who gave hints a helpfulness rating 3 or higher usually could not give a specific reason for why the hint helped them (or did not even know if the hint helped or not) despite rating helpfulness significantly higher than 1 ("not at all helpful"), suggesting that they had an intuitive sense that the hint helped them in some way. These player assessments of hint helpfulness and its relationship with player experience underscore the importance of using multiple metrics to understand the complex effect hints can have on not only player behavior, but also player perceptions.

DISCUSSION

Our use of three key metrics for hint design and evaluation, player performance, engagement, and self-reported experience, revealed the complex effects of hints in our case study with the Three Body Puzzle. These complex effects included the hint placebo effect we saw only for less experienced players who won, players' apparent desire to use fewer hints than our automatic and adaptive hint conditions gave them, and the difficulty of interpreting what it means for a hint to be helpful, both for players and for game designers. Our case study also demonstrated a general approach to hint design that addresses the underlying complexity of player experience by analyzing the impact of hints across our three general metrics. Since player performance, engagement, and self-reported experience are very general metrics which can be interpreted in many different ways, they can be measured in nearly any kind of video game, making our approach generalizable to many video games beyond the one we analyzed in this study.

It is important to note that one of our most interesting results, the presence of the hint placebo effect for less experienced players who won the game, would never have been revealed if we had not combined three different player behavior and experience measures in our analysis: subjective self-reported experience with the game, performance, and persistence. This highlights not only the importance of each individual measure we used, but also the idea that for player analytics, the whole is truly more than the sum of its parts. Using overly simplistic, unidimensional measures for evaluating hint effectiveness by themselves may miss important effects that can only be revealed by studying the interactions between multiple behavior and experience measures.

Our results aligned well with the finding in ITS literature that adaptive hints can be tricky to design in a way that is simultaneously easy to implement and useful to users. However, Anderson et al's finding that on-demand tutorial hints could have positive, neutral, and negative effects depending on the complexity of the game [3] suggests that on-demand hints may be just as difficult to implement correctly. The effectiveness of on-demand hints seems to be highly dependent on a factors specific to the game, such as genre, complexity, and target audience. These mixed results may also simply be a result of the content of the hint - whether it is more concrete or abstract, or even the specific wording of the content. Additional comparisons of on-demand hints across different types of game genres, player demographics, and hint content are needed to understand the complexities underlying the mixed results observed both in our study and in previous work.

We also found evidence that individual differences in prior gaming experience can impact how players respond to hints when latent dimensions of player experience are revealed. However, there are likely many different relevant demographic variables affecting players' reaction to hints. Arroyo et al, like us, studied gender, but in the context of hint presentation, a variable we did not manipulate in this study. Arroyo et al and Pereira et al also looked at differences in cognitive ability, personality traits, and affect, measures we did not analyze [5, 6, 20]. A larger scale study with a sufficient sample size to examine many different individual differences between players would shed light onto the relative contributions of different demographic variables in a given player's response to hints.

DESIGN IMPLICATIONS

Our findings have three main practical implications for game designers. First, the hint placebo effect we observed for less experienced players who won and the absence of this effect for many other players suggests that different types of players benefit from different types of hints. For more novice gamers in our study with high levels of the *Hints Helped factor*, winning may have been more important for a good player experience, whereas certain other players, particularly those with high levels of the *Strategy* factor, seemed to focus more on the exploration aspect of the Three Body Puzzle.

These individual player differences provide supporting evidence for two of game designer Richard Bartle's four types of players: achievers, whose primary goal is to achieve the goals of the game, and explorers, who want to understand the game's breadth and enjoy discovery [7]. Game designers should use this knowledge to design hints that address individual player differences, but should also be aware that the player types we observed in this study may not generalize to other types of games with different goals and mechanics. Therefore, game designers should focus on identifying player types specific to their game before designing hints. One way to do this could be to explore methods used in ITS research for modeling user behavior to identify specific types of players, and then design hint presentation and content tailored to each type.

Second, our finding that players in the on-demand hint condition saw hints less often and had higher levels of the *Engagement* factor suggests that good parameter tuning for automatic and adaptive hints is very important for ensuring their usefulness to players. Thus, we recommend that game designers implement on-demand hints first to see how many times and at what points players typically request help, and then use this data to inform the design of automatic, adaptive hints that are more likely to align with player preferences. Implementing adaptive hints without first experimenting with on-demand hints may work as well, provided there is a thorough analysis of player behavior data beforehand, but this approach may be riskier, as our results revealed.

Finally, game designers who want to assess hint helpfulness using player self-report measures should make sure to clearly explain to players what is meant by "helpfulness" if they want to reduce ambiguity in player responses. However, ambiguous or even negative reactions to hints can still be useful to designers; in our study, they helped reveal more complex effects of high-level, indirect hints on player experience, such as the balance between being too helpful and not helpful enough that some of our players mentioned. To account for this complexity, game designers may want to consider assessing multiple dimensions of helpfulness. One possible way to do this would be to ask players directly to what extent they felt the hints affected how fun, easy, boring, and frustrating they found the game.

LIMITATIONS AND FUTURE WORK

This work had several limitations that open the door for future work. First, we analyzed data from only one game. Future work could investigate the extent to which this data-driven, player experience-centric approach to hint design is generalizable across different games and game genres. Second, this study only analyzed one type of hint content and three types of hint display. Future work could investigate different levels of specificity in hint instructions to see whether players prefer more abstract or more concrete hint suggestions. Third, we analyzed how often participants played video games, but not what kind of video games they played. It would be interesting to investigate whether the genre of game a player prefers affects how they respond to different types of hints. Finally, while we collected a large amount of low level player behavior data for this study, we analyzed all behaviors as either totals or averages across all attempts each player made. Future work could study player behavior through the lens of time series from a player's first game to their last game and include additional low-level measures such as a measure of "distance" between individuals' solutions as in [16, 21, 22] to gain deeper insights into player behavior patterns.

CONCLUSION

This work makes three main contributions to the study and design of intelligent hint systems for video games. First, we contribute a new data-driven approach to hint design and evaluation that uses player experience, performance, and engagement together assess the effectiveness of hints in video games. Second, we present a case study demonstrating how to implement this approach in practice, which reveals that player experience is a complex, multi-dimensional concept dependent on latent variables that reflect individual differences between players. Third, the results of our case study reveal that hints may have a placebo effect for players with less prior game experience, leading them to believe the hints helped even if they didn't actually change the player's in-game behavior. Our work contributes to the body of knowledge about the psychology of different kinds of intelligent and non-intelligent hinting systems in games and provides game designers with actionable insights to guide hint design.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE - 1144245.

REFERENCES

1. Ernest Adams. 2014. *Fundamentals of game design*. `https://books.google.com/books?hl=en`

2. V Aleven and K R Koedinger. 2000. Limitations of student control: do students know when they need help? *Intelligent tutoring systems* (2000), 292–303.

3. Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Rich Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popovic. 2012. The impact of tutorials on games of varying complexity. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, New York, New York, USA, 59. DOI: `http://dx.doi.org/10.1145/2207676.2207687`

4. John R. Anderson, Albert T. Corbett, K.R. Koedinger, and R. Pelletier. 1995. Cognitive Tutors: Lessons Learned. *The Journal of Learning Sciences* 4, 2 (1995), 167–207.

5. I Arroyo, J E Beck, B P Woolf, C R Beal, and K Schultz. 2000. Macroadapting animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In *Intelligent Tutoring Systems, Proceedings (Lecture Notes in Computer Science)*, K Gauthier, G and Frasson, C and VanLehn (Ed.), Vol. 1839. 574–583.

6. Ivon Arroyo, Winslow Burleson, Minghui Tai, Kasia Muldner, and Beverly Park Woolf. 2013. Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology November 2013* 105, 4 (2013), 957–969. `http://ovidsp.ovid.com/ovidweb.cgi?T=JS`

7. Richard (MUSE Ltd) Bartle. 1996. Hearts, Clubs, Diamonds, Spades: Players who suit MUDs. *Journal of*

MUD research 6, 1 (1996), 39. DOI: http://dx.doi.org/10.1007/s00256-004-0875-6

8. Elizabeth A. Boyle, Thomas M. Connolly, Thomas Hainey, and James M. Boyle. 2012. Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior* 28, 3 (2012), 771–780. DOI:http://dx.doi.org/10.1016/j.chb.2011.11.020
9. Cristina Conati, Natasha Jaques, and Mary Muir. 2013. Understanding attention to adaptive hints in educational games: An eye-tracking study. *International Journal of Artificial Intelligence in Education* 23, 1-4 (2013), 136–161.
10. Christina Conati and Xiaohong Zhao. 2004. Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game. *Proceedings of the 9th international Conference on Intelligent User Interfaces* (2004), 6–13. http://portal.acm.org/citation.cfm?id=964442.964446
11. Alena Denisova and Paul Cairns. 2015. Adaptation in Digital Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press, New York, New York, USA, 97–101. DOI: http://dx.doi.org/10.1145/2793107.2793141
12. Kees Dorst and Nigel Cross. 2001. Creativity in the design process: Co-evolution of problem-solution. *Design Studies* 22, 5 (sep 2001), 425–437.
13. R. Garris, R. Ahlers, and J. E. Driskell. 2002. Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming* 33, 4 (dec 2002), 441–467. DOI:http://dx.doi.org/10.1177/1046878102238607
14. Wijnand A. IJsselsteijn, Yvonne A.W. de Kort, Karolien Poels, Audrius Jurgelionis, and Francesco Bellotti. 2007. Characterising and measuring user experiences in digital games. (2007).
15. Kenneth Koedinger, John Anderson, William Hadley, and Mary Mark. 1997. Intelligent Tutoring Goes To School in the Big City. *Human-Computer Interaction Institute* (1997). http://repository.cmu.edu/hcii/5
16. Timotej Lazar and Ivan Bratko. 2014. Data-driven program synthesis for hint generation in programming tutors. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 8474 LNCS. Springer Verlag, 306–311.
17. R. Charles Murray, Kurt VanLehn, and Jack Mostow. 2004. Looking Ahead to Select Tutorial Actions: A Decision- Theoretic Approach. *International Journal of Artificial Intelligence in Education* 14, 3 (2004), 235–278. http://www.ijaied.org/pub/930
18. Eleanor O'Rourke, Christy Ballweber, and Zoran Popovií. 2014. Hint systems may negatively impact performance in educational games. In *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14*. ACM Press, New York, New York, USA, 51–60. DOI: http://dx.doi.org/10.1145/2556325.2566248
19. Luc Paquette, Jean-François Lebeau, Gabriel Beaulieu, and André Mayers. 2012. Automating Next-Step Hints Generation Using ASTUS. Springer Berlin Heidelberg, 201–211. DOI: http://dx.doi.org/10.1007/978-3-642-30950-2_26
20. Carlos Pereira Santos, Vassilis-Javed Khan, and Panos Markopoulos. 2016. Inferring A Player's Need For Cognition From Hints. In *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16*. ACM Press, New York, New York, USA, 76–79. DOI:http://dx.doi.org/10.1145/2856767.2856805
21. Thomas W. Price and Tiffany Barnes. 2015. An exploration of data-driven hint generation in an open-ended programming problem. In *CEUR Workshop Proceedings*, Vol. 1446. CEUR-WS.
22. Kelly Rivers and Kenneth R. Koedinger. 2014. Automating hint generation with solution space path construction. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 8474 LNCS. Springer Verlag, 329–339.
23. Vasile Rus, Rajendra Banjade, Nobal Niraula, Elizabeth Gire, and Donald Franceschetti. 2017. A Study On Two Hint-level Policies in Conversational Intelligent Tutoring Systems. Springer Singapore, 171–181. DOI: http://dx.doi.org/10.1007/978-981-10-2419-1_24
24. Jesse Schell. 2014. *The Art of Game Design: a book of lenses*. CRC Press.
25. Olli Sotamaa. 2007. Perceptions of Player in Game Design Literature. *Situated Play, Proceedings of DiGRA 2007 Conference* (2007), 456–465. http://www.digra.org/dl/db/07311.59383.pdf
26. Cheri Speier, Joseph S. Valacich, and Iris Vessey. 1999. The Influence of Task Interruption on Individual Decision Making: An Information Overload Perspective. *Decision Sciences* 30, 2 (1999), 337–360. http://doi.wiley.com/10.1111/j.1540-5915.1999.tb01613.x
27. Kurt Vanlehn. 2006. The Behavior of Tutoring Systems. *Int. J. Artif. Intell. Ed.* 16, 3 (2006), 227–265. http://dl.acm.org/citation.cfm?id=1435351.1435353
28. Stefanie Wetzel, Katharina Spiel, and Sven Bertel. 2014. Dynamically adapting an AI game engine based on players' eye movements and strategies. *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems - EICS '14* (2014), 3–12.