Graphically Speaking

Editor:
Miguel Encarnação

Methods for Game User Research

Studying Player Behavior to Enhance Game Design

Heather Desurvire
User Behavioristics Research

Magy Seif El-Nasr
Northeastern University

The past decade has seen the game industry evolve from a niche to a major sector of the entertainment industry. In 2010, the Entertainment Software Association reported that the computer game and videogame industry generated over US$4.9 billion in added value to the US gross domestic product.[1] Additionally, games have become a central form of popular media that has impacted the lives of both youth and adults; 68 percent of American households play videogames.[1] This broadening of games' appeal has motivated the industry to create innovative forms of play; these new forms allow a variety of interactions and accommodate players of all ages, intellectual abilities, and motivations. So, designing a good game requires understanding players and their interaction with this new interactive digital experience.

Consequently, *game user research* (GUR) has emerged as an important area of investment for the game industry and a rich domain for collaboration between user experience professionals, game developers, and academic researchers. GUR has grown into a community, which has a presence as an International Game Developers Association special-interest group, holding annual international summits and workshops (see gamesuserresearch.org). The largest human-computer interaction organization, ACM SIGCHI, has expanded to add a branch on games.

Game user researchers (we'll just call them "researchers" from now on) help game designers improve the player experience to both meet the designers' goals and, most important, make gameplay more pleasurable. To identify and understand the player experience, these researchers employ a variety of methods spanning both qualitative and quantitative data collection and analysis. Here, we examine several of the most important methods and show how we applied two of them in a case study.

Current Methods

When researchers think about identifying the player experience, they think about how players interact with the game and how players experience and label this interaction. Researchers have borrowed a variety of methods from human-computer interaction and have connected them with gameplay theory and design to uncover these experiences and determine where a game creates a good player experience and where it falls short. Designers then use this information to improve the game as needed.

Some of the most common and effective methods are *think-aloud*, *Rapid Iterative Testing and Evaluation* (RITE), *heuristics*, *playtesting*, and *A/B testing*. (Another effective method is game analytics, which we discussed in a previous article.[2]) Figure 1 shows where these methods fit in game production.

Think-Aloud

In 1982, Clayton Lewis introduced the think-aloud protocol to user experience professionals.[3] Peter Wright and Andrew Monk later introduced it as a method for usability testing.[4] This protocol has been modified for the gaming industry; it's one of the most widely used and effective ways to produce actionable results for game designers.

In think-aloud, a player verbally expresses thoughts, feelings, actions, and experiences during gameplay. The researcher conducts the sessions, typically in a laboratory setting (see Figure 2), probing the player when necessary to share

 Published by the IEEE Computer Society

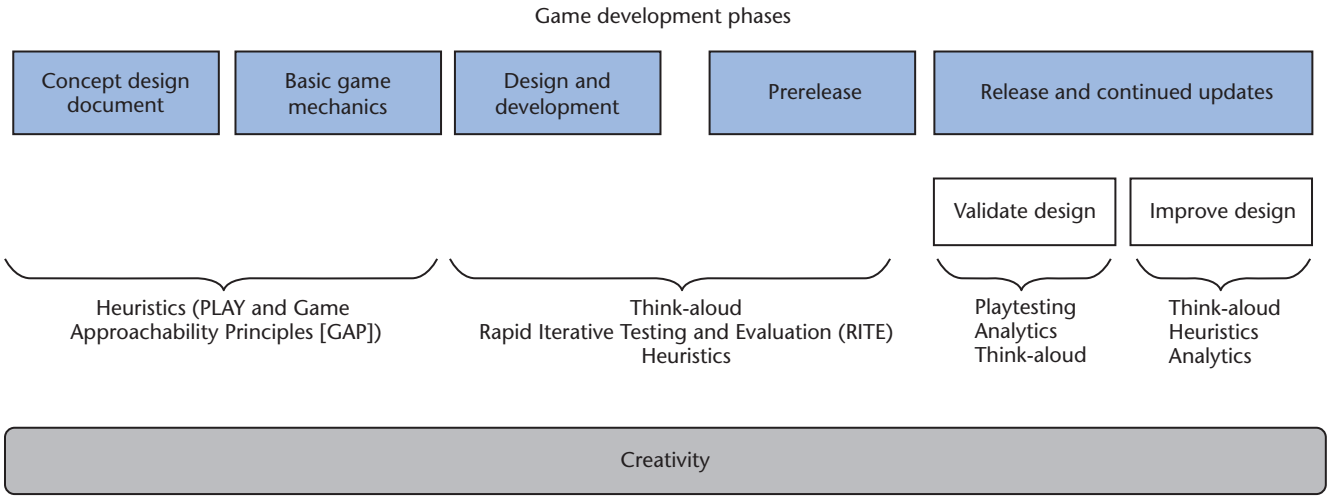

Figure 1. Using game-user-research (GUR) methods during game production. GUR methods can improve the game experience to both meet the designers' goals and make gameplay more pleasurable.

experiences aloud. Because researchers encourage the players to verbalize their immediate thoughts as they play, the method reduces bias and inaccuracies common to player self-reports. It's important for researchers, not the game designers, to conduct these studies. This lets the designers focus on problem solving rather than dealing with the difficulty of making decisions due to being emotionally attached to their designs.

The researcher collects this data and objectively reports it to the designer, who can then concentrate on creative solutions. Both researchers and designers may observe live sessions or video highlights, with the researchers identifying and illuminating the results for the designers.

Think-aloud works with new players and in both single and multiplayer situations. Additionally, it's fast and cost-effective.

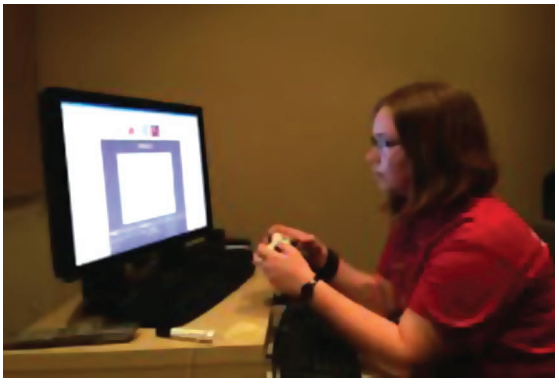

Figure 2. An example lab setup for testing. When testing, we focus on the players' experience of the game. We observe both the gameplay screen and the players. We use what we observe about their stance, facial expressions, in-game interactions, and verbal expressions.

RITE

Microsoft Games Research developed RITE during usability testing for the popular game Halo 1 and has used it for subsequent Halo releases and a variety of games. Companies such as THQ, Disney, Activision, and Electronic Arts also regularly use it.

In RITE, during game testing, researchers employ think-aloud. However, once a problem is identified, it's fixed, even after only one user, with testing resuming following that fix.

RITE uncovers issues that otherwise aren't easily identified. For example, in a game-testing session, if a player can't see an entrance to another room to perform the next puzzle, it might be due to a playability or usability hurdle. Using RITE, the researcher can discover that the door is hidden because the entrance looks more like the background graphics than a door. Correcting this problem will let the player naturally find and play the next puzzle.

Heuristics

Traditionally, heuristics are a quick-and-dirty way to identify product usability problems early in production. In industry, pressure often exists to meet aggressive deadlines, with limited budgets. Heuristic evaluation originated with the 10 heuristics that Jakob Nielsen identified and validated.[5]

The things that set videogame testing apart from the testing of productivity products are the elements that contribute to a game being fun, which go far beyond simple usability. One such element is challenge. If all player errors were prevented, which they would be according to the productivity heuristics, the game wouldn't be fun anymore. In contrast, the possibility of errors is part of gameplay. Other elements include pace, flow, immersion, engagement, game mechanics, sound track, sound effects, camera angle, narrative, and emotional connection.

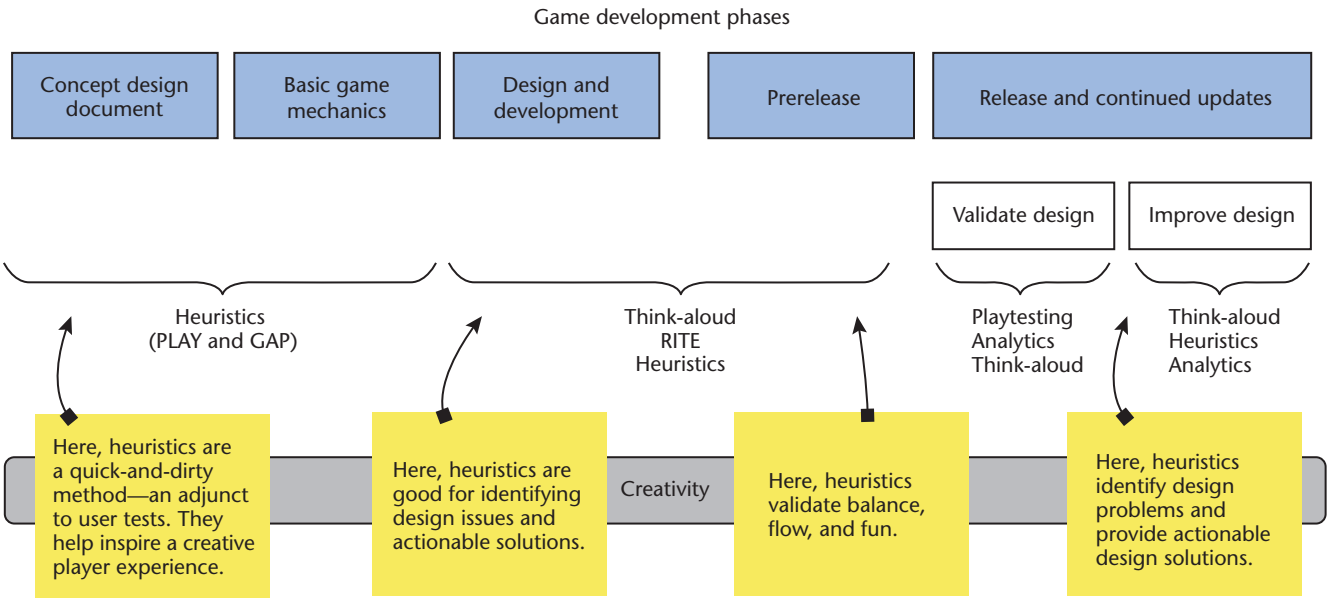

Figure 3. The yellow notes show the use of heuristics at different parts of game production. The case study described in the main article shows some examples of these uses.

PLAY heuristics. To provide heuristics for games, researchers have investigated using and modifying usability heuristics. Heather Desurvire and her colleagues moved heuristics further, developing a set of heuristics called Heuristic Evaluation for Playability (HEP).[6] She and Charlotte Wiberg revised HEP into PLAY, a validated set of foundational heuristics that can be applied to many game types and genres.[7] These were based on Noah Falstein and Hal Barwood's research for the 400 Project (www.finitearts.com/Pages/400page.html), which included theories of fun and narrative, and interviews with designers of AAA (big budget) games.

Game Approachability Principles. Tutorials and introductory levels in games are a particular area for which designers can use heuristics. These are often the first place that players quit playing or develop a negative perception of a game owing to being stuck. Often, players in such circumstances say they're bored because they don't think they can win.

Having helped develop over 30 AAA games in situations in which it was too late to change the tutorial and introductory levels, Desurvire and Wiberg developed Game Approachability Principles (GAP).[8] These heuristics were based on literacy and learning principles as applied to videogames, such as in James Gee's research,[9] and psychological schools of learning, such as cognitive and social learning.

Similar to the "theorem" attributed to Atari's Noah Bushnell—"All the best games are easy to learn and difficult to master"—GAP aims to let players easily access a game such that learning the tools of play is fun and mastering the game is possible. So, an example heuristic is, "Give the player sufficient practice with new actions before punishing him or her for failing." Another example represents giving the information to the player in context and in time: "There's a reason to learn and understand the information given to the player."

Employing heuristics. Figure 3 illustrates using heuristics during game production. Researchers review the heuristics one by one while performing a walkthrough of the game, always keeping the player demographic in mind. They mark whether each heuristic has been violated or adhered to. If a heuristic has been violated, they suggest a fix that matches the game designer's intent. In short, the designer still owns the design, whereas an objective researcher identifies or predicts problems. Finally, game designers themselves can use heuristics when creating the initial concept, and iterations of the heuristics help them consider good game design principles and stimulate new, better ideas. Armed with GAP and PLAY, researchers have been able to identify design issues and recommend fixes, validating what works well and enhancing the designers' creativity by suggesting improvements.

Heuristic evaluation's benefits seem obvious: identifying any game pain points as well as validating good design elements quickly and cost-effectively. The downside is, of course, that it isn't based on observing real players. This underlines the fact that the method should be an adjunct to, not a substitute for, other methods.

Playtesting

Traditionally, game publishers had market research-

ers employ focus groups to identify players' perceptions of their games. Companies such as Disney, Sierra Online, and LucasArts did this and called the process playtesting. However, as researchers developed their craft and evolved methods from product research, playtesting evolved to focus on testing groups of players playing a game over a long time period. In this way, researchers can identify problems and better identify the players' perception of fun. Such research is best performed when a game is ready—that is, close to its prerelease date.

Playtesting involves many players at separate stations. They play through all or a part of the game (for example, the first hour). Gameplay can continue over several days or several weekends to complete the game. Playtesting measures the players' engagement and identifies pacing and balance problems. Playtesting methods include surveys, interviews, and player observation. Most researchers use several methods and triangulate between them.

Also, some researchers, such as Mike Ambinder at the Valve Corporation,[10] use physiological sensors to gauge the players' affective state. However, these methods haven't been experimentally validated to show coherent correlations between game events and affect. They are also expensive and might interfere with the test's ecological validity. Data analysis requires an expert and is usually extensive and time-consuming.

Additionally, researchers are developing standardized survey measures that extend the ones used in the social sciences. For example, Microsoft developed a scalar measure to collect players' perception of their enjoyment. Throughout testing sessions, this method uses two short scales to measure playability and fun. Similarly to other psychology experiments, using short, quick scales lets players become accustomed to the questionnaires so that they can use them without having to think hard or read the answers each time. In other words, when players become used to a simple three-point scale (such as 1 = not fun at all, 2 = kind of fun, and 3 = really fun), they'll automatically volunteer feedback using that scale.

To further reduce player self-consciousness and to access the player's true experience, some researchers remotely observe the players, having them score their level of fun as they play. The researchers then compare the scores the players give as they play with the ones they give afterward. In addition, the researchers observe the players' facial and body expressions to determine whether the players' self-rated scores match. So, even if the scores' absolute values aren't found to be accurate, the relative values will be. Imagine that a player rates a game using the scale we described in the previous paragraph. The player gives a rating of 2 while playing the game but gives a 1 or 3 rating afterward; that change is significant relative to the previously reported score. Comparing the absolute scale scores between the players would be tentative; identifying the relative differences is more salient and meaningful.

A/B Testing

This method, similar to in-game metrics in that the analytics are taken from within the game, compares two game versions to find out which one produced better results. This helps gauge specific differences between the versions. It often involves hypotheses posed to determine whether some game aspect made a difference—for example, whether a change in an interface item affected conversion rates (that is, when users convert to paying users). To use A/B testing, researchers and designers must have two distinct versions of the game and a specific, clearly measurable outcome. Measurement techniques vary. For online games, researchers often use the frequency of a specific choice (for example, signing up for the game after playing a demo).

Case Study: Employing Think-Aloud and Heuristics

This case study is based on work with a AAA game design studio (any specifics are privy to the company and not the public). When providing overall game usability testing using think-aloud, we identified a catastrophic design problem. The design team could easily learn and use the new game mechanics essential for play. However, the players weren't able to learn the mechanics, even with assistance, so they stopped playing.

The design team intended the new game mechanics to enhance gameplay by introducing a new way to operate the Xbox 360 and PlayStation 3 controllers. These mechanics conflicted with how players usually use the controllers. Players typically use the letter or color buttons and analog sticks for camera control and avatar movement. Most games use the same mapping, thus minimizing the interface knowledge transfer from game to game. However, because this new game relied on an atypical use of the controller, players needed to unlearn the old way.

This problem was uncovered only through think-aloud (at point 1 in Figure 4). This was because the design team and its associates had learned the new controller scheme easily, so they assumed other players would too. There were two options to correct this problem: redesign the controller

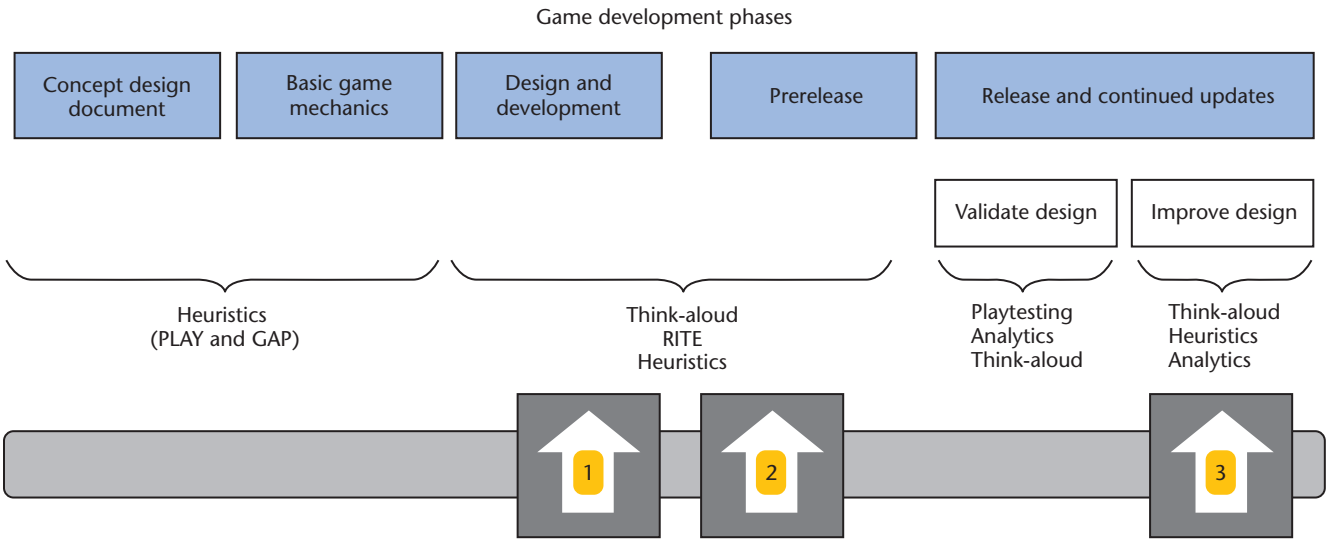

Figure 4. Using think-aloud during development. At point 1, think-aloud uncovered several problems with the player experience—the players weren't easily learning the new mechanics crucial to gameplay. At point 2, the game designers used heuristics to design a tutorial in which players easily and enjoyably learned the mechanics. At point 3, the user researcher and designers again used think-aloud to verify that the tutorial was successful, in terms of the players both learning the mechanics and enjoying the game.

to map to the usual setup, or devise a tutorial to help players unlearn the old way and learn the new way. The team opted to devise a tutorial. So, they went back to the design phase (point 2 in Figure 4) and used PLAY and GAP to design a tutorial that was acceptably fun to play and, most important, could easily teach the new controller scheme.

We spent a week with the design team. We spent one day teaching them GAP. The next few days, using GAP, we and the team developed a new map of learning goals for the players. We broke the game mechanics into sections that were easy to learn and use. From there, we interwove story elements and the perceived gameplay challenge into the context of each section, on the basis of GAP.

For example, one GAP principle is that players learn best from seeing a skill demonstrated, practicing it, and gaining the confidence that they can perform it. One way to achieve that last item is to show a nonplaying character (NPC) demonstrating the skill or using hidden game mechanics and then giving the player an opportunity to mimic the NPC. Giving players the opportunity to practice the mechanics without perceivable consequences gives them more confidence they can perform it in more critical situations in the game. In addition, providing verbal reinforcement or encouragement, such as "Excellent," helps players feel more confident they can continue and will improve. The team identified each element they needed to teach and set up this structure for learning: demonstration and practice without consequences, thus increasing self-efficacy.

The design team also found creative ways to apply GAP to teach the new game mechanics. This included integrating instruction with the narrative and helping to establish players' emotional connection to their avatar. So, when the teaching began, the gameplay did too.

For example, the team identified mechanics in which the player learned to operate a weapon used against only a specific enemy. Players first viewed a nonplaying character using the weapon for that enemy, which demonstrated that the weapon and enemy existed. If the player didn't find the weapon and correctly use it within a certain time period, a visual showed and demonstrated which controller buttons to press.

The team also considered whether having to perform the tutorial would annoy the more advanced players. Designers generally want to have more people playing their games. So, for new or intermediate players, learning the game mechanics might be a greater barrier to playing the game than being slightly annoyed would be for hard-core players. Hard-core players are more likely to quickly press through a tutorial. This is a tradeoff in which erring on the side of easier game entry through the tutorial will increase the number of players. The key then is to lessen the potential annoyance for hard-core players.

The team did this by weaving the tutorial elements into actual gameplay and building the story elements. By using GAP to structure learning, the team built fun and gameplay into the tutorial. They used PLAY and their creativity to create elements of fun and intrigue and to move the story forward. This helps hard-core players have a more enjoyable experience in the tutorial while teaching them and less-experienced players to play the

game. These practices all ultimately aim to have all players continue to play the game.

Finally, the team designed a location where players could practice. First, the game showed several enemies in succession to make the players feel challenged (including realizing that losing was possible) but also to convince them that they could be successful once they learned the new mechanics. Once players learned the mechanics, they were free to engage in real game combat with real consequences. The team used many elements of narrative design, level design, and character design, stimulated by GAP, to enhance play within the parameters of learning the basic mechanics.

The team used think-aloud again on the new design (point 3 in Figure 4). All the players learned the new mechanics easily; only one needed to replay the tutorial to learn them. In the first testing session, the players had felt that the tutorial wasn't fun. They reported that the new design was fun and engaging. However, only a minority of the players said they would enjoy replaying the tutorial. That's to be expected because the tutorial aimed not to create the most possible fun but only to make learning more fun. If the team had more time to refine the design, the fun levels would likely have increased further. Most important, the players were able to learn the essential game mechanics.

Game user research is still in its inception. Researchers will no doubt devise many other methods to advance our understanding of how players play games and thus help achieve better designs. Some approaches might be decided by the videogame industry on the basis of budget and time considerations. Heuristics are fairly inexpensive in terms of cost and time. Think-aloud tends to be a quick-and-easy way to test games on a small sample of users and gain deeper insight into fixing player experience issues. Employing manual metrics and programming hooks into the game code for analytics of user gameplay provide another way to learn about in-game behavior.

Researchers are concentrating particularly on the triangulation of methods using analytics, which can reveal quantitative aspects of in-game behavior, and the other qualitative and observational measures we discussed in this article. We believe a combination of methods will give a better indication of the context of play and more insight on players' experience, emotions, motivations, and attention. This in turn will help researchers and designers deduce a game design's success and help improve the game.

References

1. S.E. Siwerk, *Video Games in the 21st Century*, Entertainment Software Assoc., 2010; www.theesa.com/facts/pdfs/VideoGames21stCentury_2010.pdf.
2. M. Seif El-Nasr et al., "Game Analytics for Game User Research, Part 1: A Workshop Review and Case Study," *IEEE Computer Graphics and Applications*, vol. 33, no. 2, 2013, pp. 6–11.
3. C.H. Lewis, *Using the "Thinking Aloud" Method in Cognitive Interface Design*, tech. report RC 9265, IBM, 1982.
4. P.C. Wright and A. Monk, "The Use of Think-Aloud Evaluation Methods in Design," *ACM SIGCHI Bull.*, vol. 23, no. 1, 1991, pp. 55–57.
5. J. Nielsen, "Heuristic Evaluation," *Usability Inspection Methods*, J. Nielsen and R.L. Mack, eds., John Wiley & Sons, 1994, pp. 25–62.
6. H. Desurvire, M. Caplan, and J.A. Toth, "Using Heuristics to Evaluate the Playability of Games," *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ACM, 2004, pp. 1509–1512.
7. H. Desurvire and C. Wiberg, "Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration," *Proc. 3rd Int'l Conf. Online Communities and Social Computing* (OCSC 09), Springer, 2009, pp. 557–566.
8. H. Desurvire and C. Wiberg, "Master of the Game: Assessing Approachability in Future Game Design," *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, ACM, 2008, pp. 3177–3182.
9. J. Gee, *What Video Games Have to Teach Us about Learning and Literacy*, Palgrave Macmillan, 2003.
10. M. Ambinder, "Biofeedback in Gameplay: How Valve Measures Physiology to Enhance Gaming Experience," presentation at 2011 Game Developers Conf., 2011; www.valvesoftware.com/publications/2011/ValveBiofeedback-Ambinder.pdf.

Heather Desurvire *is the principal at User Behavioristics Research and an adjunct faculty member in the University of Southern California's Interactive Media Department. Contact her at heather3d@gmail.com.*

Magy Seif El-Nasr *is an associate professor at Northeastern University's College of Arts, Media, and Design and College of Computer and Information Sciences, the Director of Game Educational Programs and Research at Northeastern, and the Director of Game Design in the College of Arts, Media, and Design. Contact her at magy@neu.edu.*

Contact department editor Miguel Encarnação at lme@computer.org.