



The Science of Playtesting

We go behind the scenes at Valve, Bungie, and Epic Games to see how user feedback helps shape the game development process.

Last updated by Laura Parker (/profile/Lozzica/) on September 9, 2012 at 11:06PM

 [Comments](#)



(<https://www.facebook.com/sharer/sharer.php?u=https://www.gamespot.com/articles/the-science-of-playtesting/1100-6323661/>)

(<https://twitter.com/intent/tweet?text=The Science of Playtesting&related=&url=https://www.gamespot.com/articles/the-science-of-playtesting/1100-6323661/>)

(https://share.flipboard.com/bookmarklet/popout?v=2&title=The Science of Playtesting&url=https://www.gamespot.com/articles/the-science-of-playtesting/1100-6323661/&utm_source=gamespot.com&utm_medium=article-share&utm_campaign=publisher)

Related



[+ Follow](#)

(/halo-combat-evolved/)



[+ Follow](#)

(/halo-2/)



[+ Follow](#)

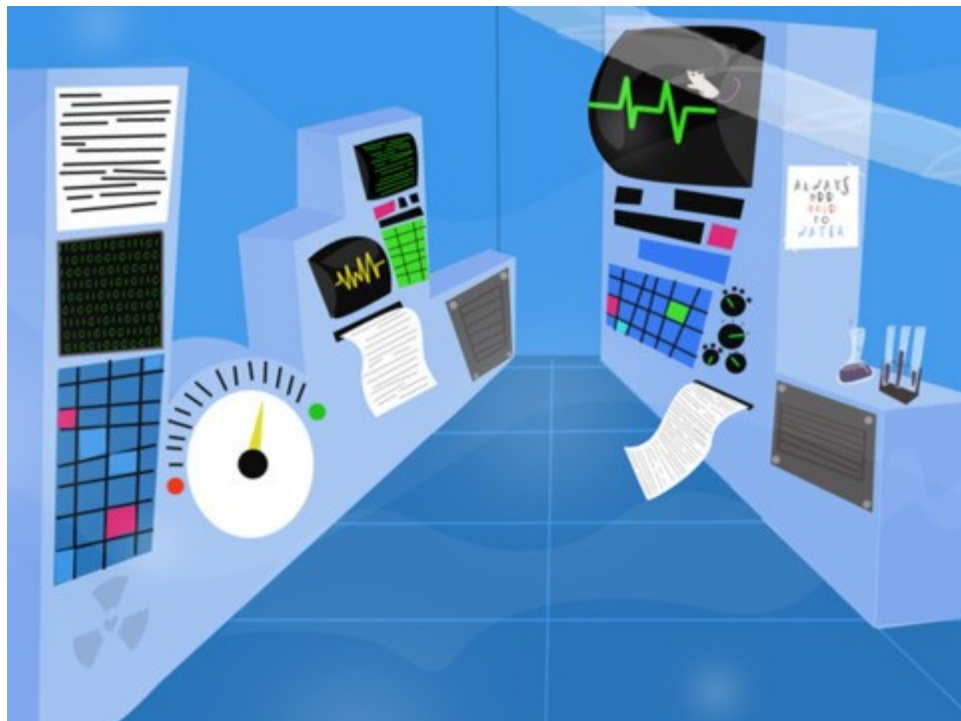
(/halo-3/)

rippling across the skin of one of the candidates. Whoever he is, his nerves have just given him away.

The playtesting rooms inside Valve's headquarters resemble a large, untidy science lab. PC monitors line the wall, spewing out steady streams of graphs, patterns, and charts. Some keep track of candidates' heart rates; others record their eye movements. A pile of resistors and voltage batteries lie discarded in one corner of the room; candidate information reports, video recordings, audio transcripts, written questionnaires, and the results of thousands of hours of direct observation lie in the other. Valve takes its playtesting seriously.

While playing video games for a living may sound like the dream ticket for any avid gamer, for those involved in the process, it represents much more than just another routine part of game development. Testing games before they are released gives game developers a rare insight into the end user experience, helping them validate the quality of the game and isolate potential problems. We now see games as much more than simple products; we see subjective experiences that affect individuals in different ways. For this reason, video game testing not only has to be more rigorous and pervasive than product testing in other industries, but also more precise. Can feelings be measured? Why does one player enjoy a particular game, while another does not? How can user feedback be used to make games better?

Publishers and developers are constantly seeking answers to these questions, trialing a variety of playtesting methods to get the best results. In this feature, GameSpot will go behind the scenes of three major studios--Valve, Bungie, and Epic Games--to find out how tools like science and psychology are helping game developers better understand the nature of player behavior.



Playtesting room or science lab?

A Healthy Dose of Perspective

Three years ago, Valve hired experimental psychologist Mike Ambinder to head up the playtesting department of its development arm. With a B.A. in computer science and psychology from Yale University and a Ph.D. in psychology from the University of Illinois, Ambinder was the perfect man to lead Valve in a new direction in the field of user research.

Using his background in visual recognition, Ambinder began treating Valve's playtesting sessions as a series of scientific experiments designed to test various game design hypotheses. Since most psychological research revolves around trying to isolate mechanisms of behavior to find out why people are motivated to do the things they do, Ambinder thought game playtesting should strive to do much the same thing: use extractable data to make an altogether better user experience.

"For us, playtesting is the most important part of the game development process," Ambinder says. "It's not something we save for the end of the development, or use as a quality assessment (QA) or balancing tool. Instead, it is the dominant factor that shapes our decisions about what to release and when to release it."

This is a relatively new attitude for the games industry. While playtesting has always been a vital part of the game development process, the role it plays in shaping the final user experience has never been as important as it is now. Traditionally, playtesting methodologies

focused on video games as products rather than as variable experiences that can affect players in different ways. Even as video games became more complex and publishers began to employ dedicated quality assessment teams rather than simply having the development teams playtest their own games, user feedback practices remained fixed on highlighting problems surrounding the more objective aspects of game design--coding errors and bugs--rather than exploring the experience of what it's like to actually play the game. Methodologies like functionality testing (looking for general problems in the game's overall design), compliance testing (checking that the game complies to publishers' technical and legal requirements), compatibility testing (testing the game on various configurations of hardware and software), localization testing, public beta testing (which lets users pick up any errors the developers may have missed), and regression testing (testing to make sure previously reported bugs have been eliminated) have become industry standards, with publishers and developers mixing and matching various methods to suit individual project needs.

But things are changing. The growth of gaming audiences and the subsequent push towards more diverse gaming experiences has led some publishers to rethink traditional playtesting methods in favor of something a little more relevant. According to Ambinder, the games industry is starting to move towards more innovative ways of gathering data, willing to spend more time, energy, and resources on its accumulation.

"I think more and more companies are starting to see the value in hiring folks with backgrounds in psychology or related fields that provide skilled training in extracting meaningful data from playtesters," Ambinder says. "For us, playtesting is crucial, as it is the most effective and honest means of validating our products. We would be foolish to release a game that went through minimal playtesting, as we could have little confidence in its quality had it not gone through rigorous testing. To that end, we start playtesting as soon as we have something playable, and we basically never stop as we are constantly updating our products after shipping."

When developers first became interested in user research some 10 or so years ago, the standard practice was not to waste too much time on it; therefore, only the first hour of the game would undergo outside playtesting. Things are a little more serious these days. Most publishers would consider it madness to release a game that hasn't been subjected to hundreds of hours of rigorous playtesting, combing over every single part of the game right

up to its release. A decade's worth of knowledge has come down to one thing: perspective.

After spending years as a playtester himself, Epic Games senior game test manager Prince Arrington saw the value of perspective in the work of those around him.

"I always got a kick out of giving feedback to developers and then seeing my suggestions come to life in-game. But sadly, it has the tendency to be one of those things that can easily be undervalued. It's very easy for us, as developers, to get too close to our projects and fall into the trap of not realizing that our baby isn't perfect. This often leads to poor design remaining poor. The value in having outside feedback is that it's always nice, if you're open to constructive criticism, to get those checks throughout the development cycle so that you can get the confirmation that you're making something kick-ass...or otherwise, the painful realization that you're not."

()

Bungie playtested its Halo games with people who had never even seen a shooter.

Just like Ambinder, Arrington's professional career is rooted in psychology. Earning his psychology degree from North Carolina State University, Arrington left academia to become a contract tester when his father told him that he played games so much that he should probably start making them. During his three years as Epic's QA manager, Arrington has come to see playtesting as a valuable tool in keeping developers in check.

"Playtests are as much a part of the development cycle as design meetings or code reviews,"

Arrington says. "While different in execution, it's something that should be done daily, with different combinations of participants with varying skill sets willing to give constructive criticism, with the sole purpose of making the game as good as humanly possible.

"Developers work on a project for so long, there's always the potential to lose perspective on what's working and what's not working. Without this form of genuine feedback, developers have a tendency to drink the Kool-Aid and become accustomed to the inefficiencies and flaws of a project, which leads to a failure to explore more suitable paths."

This perspective is often offered by people who have never played video games before. Both Valve and Epic aim for a wide demographic when deciding whom to bring in, from seasoned gamers, to people who have never played a shooter before, right down to non-gamers. With the gaming audience growing every day, developers have come to understand the value of reaching out to all skill levels, fighting to keep existing audiences while trying to snare new ones.

"We have a sign-up page for playtesters on Bungie, and we bring in a whole range of people: people who are experts, people who play casually, everyone," says John Hopson, the user research design lead at Bungie. (He too has a psychology degree). "For Halo testing we even bring in people who have never played a shooter before. It's painful to watch because they have a lot of trouble, especially with things that we don't even think about anymore, like coordinating the use of two thumbsticks or knowing where the buttons are. But Halo is supposed to be a game that is fun for everyone, so it's necessary to see how people who have never played a shooter before react to it."

Arrington says if developers lean too far in one direction they risk skewed playtesting results, ones that don't take into account information from players who have the potential to account for a large portion of the game's user base.

"If you fail to account for players new to your game or genre, it's typically those players that will stop playing, or avoid your game, if the barrier to entry is too high. So this is very important for making the game easily accessible to the casual gamer as well as challenging enough for the veteran, hardcore gamer."



It's Not Rocket Science

Back inside Valve's playtesting rooms, Ambinder is busy measuring data. Surveys, question-and-answer sheets, verbal reports, direct observations, gameplay data, design experiments, and a handful of biofeedback data--Valve's latest foray into perfecting the process of playtesting--are painstakingly collected and categorized. It is this latter methodology that has made Valve a leader in the field of refining and improving the playtesting process. In the field of science, biometrics is a set of methodologies that can be used to identify individuals based on the measurement of intrinsic physical and behavioral traits; in the world of video games, the same set of methodologies can be applied to ascertain how much people are enjoying themselves while playing a game. Cue Ambinder, voltage battery in tow.

"We became interested in the use of biofeedback both as a playtesting methodology and as potential user input to gameplay because the idea of quantifying emotion or player sentiment seems to have utility," Ambinder says. "On the playtesting side, recording more objective measurements of player sentiment is always desired. People sometimes have a hard time explaining how they felt about various things, and memories of feelings and events can become conflated. Conversely, if you have a more objective measurement of arousal or engagement, you can get a clearer picture for how people are emotionally consuming your game."

()

Playtesting in Portal 2 helped to reveal that not all gels are good gels.

This is how it works: groups of playtesters are gathered and brought into Valve's headquarters where they are asked to sit down and play a specific game for an hour or two, depending on the title. While this is happening, a separate group of developers observe and record the behavior of the playtesters. No assistance is offered--the goal is to create as naturalistic an environment as possible. At the end of the playtesting sessions, the players may be asked to fill in surveys (so Valve can record long-term trends), before being ushered in to begin individual question-and-answer sessions (or group question-and-answer sessions for multiplayer games). From time to time, Valve brings out the big guns: eye-tracking (done remotely via a set of cameras built into the monitor while players are in the playtesting session), heart-rate monitoring, and skin conductance testing that measures the electrical pulses on the skin to determine how engaged someone is with the task he or she is performing. The point of all this is to gauge individual players' experience as they make their way through a game: Do they really like the game they're playing? Does a particular part arouse their interests more than others?

"At the moment, we're very curious about the measurement of skin conductance, a correlate of physiological arousal," Ambinder says. "By identifying the peaks and valleys in the arousal waveform, as well as the eliciting events that may cause them, one can start to gain insight into how various game mechanics and game events lead to both positive and negative changes in player state. In addition, one can attempt to predict subjective enjoyment and

frustration metrics by looking at the quality of the arousal waveform produced during the playtesting session."

Valve says all the data collected in the playtesting process is useful in one form or another, be it to fix problems with the game itself or to make the game more appealing by including more of what players seem to respond to. For example, visual outlines in *Left 4 Dead 2* were added after playtesting revealed that verbal communication among teammates was neither efficient nor accurate enough to convey relative location information. Several experimental paint types in *Portal 2* were also done away with after playtesting groups revealed that the quality of the experience of playing the game was negatively affected by their inclusion.

Of course, not everyone in the industry is as open to biometrics as Valve is. Some publishers continue to rely on a process of trial and error to handpick playtesting methodologies best suited to the kinds of games they are producing. Take Bungie, for instance. It's hard to find a game franchise that has relied on the feedback of users as consistently as Halo. Hopson has been paying attention to how players unlock doors for seven years now.

"We watch how our playtesting groups play our game, and record them," Hopson says. "We notice every little detail, from how they pick up a weapon and use it to how they go about unlocking a door. Other times we bring in groups and have them play the game and we collect data about how hard a campaign mission is, or which map is easier to navigate, and so on. And for other questions, we go to data mining [collected on Xbox Live servers]. This gives us an idea of how players play the game, and things like what maps they are best on, and so on."



Is biometrics the best of the playtesting methodologies?

Hopson records everything: how players die during the game, where they die, and what weapons were used to kill them. From this, he moves on to individual question-and-answer sessions where he will ask players about their favourite missions in the game and their reaction to things like difficulty levels and how they solved problems during the set time period. Hopson and his team attend every development, design and tech meeting in the company, constantly looking at the best ways to apply user research. For example, it doesn't matter if only one or two people are stumped by a particular issue: if it's severe enough, Hopson highlights it as important. The aim, as with all playtesting methodologies, is to discover what players think about the game. But even applying basic psychological principles, such as paying attention to how much information a player can retain at any one time, or how much memory a player uses during the playtesting process, won't produce as accurate a result as the biometrics approach. So why isn't Hopson getting out the heart-rate monitors?

"Some publishers are going down this route, but I'm not sure biometrics is the way to go. We've had a lot of time to experience what works and what doesn't, and biometrics doesn't tend to add a lot to the techniques we're already using. You could learn this stuff by just asking people. I mean, what we want to know is whether people are having fun or not. And just asking them, or watching them play the game, can determine that. We don't need super-precise accuracy. For example, I can ask you right now to describe whether the room you're sitting in is warm or cold. And you could tell me. I don't need to stick a thermometer in the room to verify that."



The More the Merrier

In its purest form, playtesting is a simple process: have people play the game, and then ask them what they thought. The real trick is getting as many opinions as possible. By its very nature, playtesting makes this hard to do--it's not easy to consistently find people to physically bring in to the premises, watch them play the game, and then interview them, often individually, about it. While playtesting has become more sophisticated over the years, development teams still have a limited number of resources to dedicate to the game-making process; coordinating daily groups of gamers into the building and looking after them, down to the smallest details of where to sit them and what to feed them, continues to remain a logistical nightmare. While publishers are slowly finding ways to work around this (Bungie hosts regular playtesting "weekends," for example, with two straight days of user testing producing a substantial amount of data), more organic solutions are becoming increasingly popular.

Both the Halo and Gears of War franchises are known for their sizable and meticulous public beta testing (the Halo: Reach multiplayer beta saw 2.7 million players hit the servers, eclipsing the Xbox Live record for a beta previously set by Halo 3), which helps developers like Bungie and Epic gather a whole array of useful information, from bandwidth size to matchmaking system preferences in multiplayer. The popularity of these public betas also means developers are rewarded with a hefty amount of user feedback via forums, email, Twitter, Facebook, in-game chat, and so on. This is mostly where Arrington and the Epic team are treated to more subjective, what-is-it-like-to-play-this-game moments, which he says can often be more insightful than cold, hard data.

()

Epic uses the Gears of War public betas to gather all sorts of useful information.

"I'm kind of old school and think that there's something pure and genuine about letting players play, observing what they do and engaging with them," Arrington says. "I don't know that ours is the most effective method out there, but it's a viable system, and it works really well here. Some swear by biometrics. Others isolate their testers in a room with a one-way mirror, record their actions, and give them an anonymous questionnaire to fill out. All methods have their benefits and drawbacks, but ultimately we have to choose what feels best for us."

It turns out public betas are a pretty useful tool: having 20 million multiplayer matches played by 1.3 million players around the world can tell developers a lot about the game they are about to release, things they might never have thought to check had it not been for a 10-year-old kid from San Paolo. For the Gears of War 3 public multiplayer beta, Arrington and his team tracked every single bullet fired (23 billion) while observing how players managed themselves in the game, from how weapons were being used to how maps were being played, exploits that were found, match durations, rule confusions, and so on. Based on the data received, Epic's development team went to work changing a large number of the matchmaking algorithms to account for the proper distribution of servers across the different regions, as well as improving gameplay aspects like opening up the spawn points to avoid spawn camping, reducing the goal score for King of the Hill to speed up matches, and tweaking the overall balance of the weapons to make them more fair during combat. But it

didn't stop there--during every single day of the Gears of War 3 beta, someone from Epic jumped in the game to see how things were shaping up.

"It's one thing to see a bunch of charts and stats and to hear secondhand how the game feels, but until you're in there yourself, it's hard to really know what's really a big issue or not," Arrington says. "It just gives you a more knowledgeable perspective on how it's supposed to be so you can better compare that against how it actually is at that moment."

As much as developers can benefit from online public betas, this methodology does little to help in tracking the day-to-day changes that occur during the development cycle of a game. While it would be impossible to generate the same amount of data gathered in a public beta test in an internal playtest environment, the level of insight that developers are afforded by daily playtests cannot be achieved through something as large-scale as a public beta. In-house playtesting begins long before developers put out the call to eager gamers; it's part of a rigorous and intensive process that these days makes up a very important part of a development team's production schedule.

()

The Epic Games quality assessment team.

At Epic, Arrington says the goal is to make sure that everyone in the company gets multiple opportunities to give feedback on as many of the game's features as possible, including levels, AI, weapons, characters, modes, user interface, visuals, and so on. Arrington's team typically runs multiple playtests daily, usually lasting one hour per session, and tracks every developer's

personal profile to monitor their progress and to check, among other things, who exactly shot them in the face.

During these playtesting sessions, Arrington has one of his team sit with the developers who are testing the game, recording comments and looking out for three basic types of data: feedback, bugs and glitches, and statistics and heat maps. The kinds of statistics tracked on the back end are things like number of downs, kills, deaths, and revives over the course of the project; heat maps offer a top-down image of the level, tracking activity on the map and providing level designers with a visual representation of high-combat areas and low-activity areas. Immediately after each playtest is over, the participants sit and discuss their experiences with the level designers while Arrington and his team take notes.

"A lot of the data collected in our playtests actually ends up being used. During the playtests, not only is my team documenting everything that occurs, but the designer of the level and the lead level designer are also there either shoulder surfing or playing with the participants. Once it's all over, they will go through all of the data that we've collected and incorporate any changes that they feel will have a positive impact on the quality of the game to the end user. Often, these changes are implemented quickly enough for the next playtest, which really allows us to iterate quickly."

()

Playtesting in the Future

A byproduct of games becoming more complex, mature experiences in the last decade is the exponential growth in the skill level of gaming audiences. Players have become more adaptive in their ability to respond to changes in the gaming environment, and the dominance of first-person shooters has contributed to an increase in the gaming mind: faster reflexes, better performance, and the ability to correctly judge and prepare for random in-game situations with relative ease. No one is in a better position to judge the evolution and growth of gamers

than those who spend most of their time watching them play. Ambinder, Arrington, and Hopson have all noticed the same trends arising from their study of player behavior over the years: players are getting smarter, quicker, and more efficient.

"In terms of extracurricular activities, gaming has become as regular a pastime as playing outside (remember that?)," Arrington says. "As a result, players definitely have a higher game IQ than when we were playing Donkey Kong and Defender in arcades back in the day. As gamer IQs have increased, developers have increased the complexity of their games. Just compare the first Madden game to last year's version: while the game of football hasn't changed, the complexity in the execution has increased exponentially. The bottom line is that more intelligent gamers require more challenging games and vice versa."

There's also the hardware to consider: gaming consoles have become more complex. (Just compare the Atari 2600 with its joystick and one button to the Xbox 360 controller with its two thumbsticks, two shoulder buttons, two triggers, a D pad, and five face buttons). This has contributed to physical increases in the skill level of gamers, from hand-eye coordination to reflex times. The fact that most games these days use similar control schemes means gamers can quickly adapt from one genre to the next without having to relearn the entire process from the start, meaning that gaming proficiency is more widespread than it has ever been.

()

Gaming used to be so simple. What happened?

"We've been exposed to this familiarity in games for so long, we inherently, in whatever game

we're playing, know that the blinking item is important, that there might be loot in the crate, that developers probably hid collectibles somewhere inconspicuous, and that shooting the red barrel is going to ruin someone's day," Arrington continues. "I think my son destroys me at fighting games to this day because I put a controller in his hand at a very early age. Had I known that he would grow to be that kid that crushes my self-esteem by letting dad win at Super Street Fighter IV, I would have made him take up chess."

The prevalence of online gaming has also changed the gaming mind: developers can no longer analyze games at the level of the individual player without taking into account online and social capabilities. So, for example, if Hopson and his team are looking at a group of 15 players in a multiplayer match and one of them is misbehaving, it becomes impossible to simply isolate that player to find out why he's behaving like that--the group has to be observed as a unit.

"We've certainly placed more emphasis on the fact that people have to experience the whole game," Hopson says. "Once upon a time, the game industry used to accept a failure rate that no other entertainment industry accepted--for example, everyone who walks into a cinema expects to see the whole film, not just part of it. I think we used to deliver games that were so difficult that not everyone could finish them, and there was no alternative to that. But now, we have the ability to offer a whole spectrum of difficulty levels for every kind of player."

The last few years have seen both publishers and developers branching out with playtesting methodologies, but always with the same goal in mind: more accurate data. Ambinder, Arrington, and Hopson all recognise the beginning of an exciting time in the field of playtesting, a time for taking risks and seeking even more novel ways of reaching further into the minds of players than ever before.

"The last few years have seen a broad expansion in the talent brought in to facilitate playtesting, and with the addition of these specialists, it is likely that the best breakthroughs are ahead of us," Ambinder says. "I think modeling of future gameplay behavior based off of prior behavior is an emerging field that has yet to reach maturity. It seems likely that our predictive tools will only improve in the years to come. I think we'll also get better at quantifying the subjective experience of playtesters and hopefully become more adept at isolating facilitators and inhibitors of enjoyment, engagement, fatigue, lack of challenge, and

so forth. It is tough creating an objective environment for playtesting, and if we become more effective at removing sources of bias, we should be able to acquire richer and more meaningful data."

()

As games become more complex, so too will playtesting methods.

While Valve has clearly embraced a more scientific approach to playtesting, Arrington doesn't want the process to become too clinical.

"I'm sincerely hoping that it remains a good mix of both organic and clinical playtesting. My fear is that developers or publishers will start looking for shortcuts or that magic bullet, and so they'll stop taking risks and start going for the sure thing. I'm praying that playtesting remains a tool and doesn't become a weapon of mass destruction. The process as it is now [at Epic] keeps us honest and takes a lot of the ego out of the equation. Whether we're playtesting internally or bringing people in, there's nothing like listening and reacting to people that are passionate about what we're doing, and it helps us make some of the best games in the industry."

Filed under: [Halo: Combat Evolved \(/halo-combat-evolved/\)](#) [Halo 2 \(/halo-2/\)](#) [Halo 3 \(/halo-3/\)](#) [Gears of War \(/gears-of-war/\)](#) [Gears of War 2 \(/gears-of-war-2/\)](#) [Left 4 Dead 2 \(/left-4-dead-2/\)](#) [Portal 2 \(/portal-2/\)](#) [Gears of War 3 \(/gears-of-war-3/\)](#)

Got a news tip or want to contact us directly? Email

news@gamespot.com ([mailto:news@gamespot.com?subject=Hi GameSpot! I have a news tip for you!](mailto:news@gamespot.com?subject=Hi%20GameSpot!%20I%20have%20a%20news%20tip%20for%20you!))

 **Upvote**

 **View Comments (0)**

© 2018 CBS Interactive Inc. All rights reserved.
Privacy Policy (<https://legalterms.cbsinteractive.com/privacy>)
Ad Choice (<https://legalterms.cbsinteractive.com/adchoice>)
Terms of Use (<https://legalterms.cbsinteractive.com/terms-of-use>) Help (/help/)
Advertise (<http://www.cbsinteractive.com/advertise>)
Partnerships (<mailto:cbsi-partnerships@cbsinteractive.com>)
Careers (https://cbs.avature.net/cbsinteractivcareers/SearchJobs/?3_56_3=%22260%22)



(<https://www.gamespot.com/>)