Unraveling Toxic Behavior in Online Games

A glimpse into the long term solutions for toxic gaming

By Alex Massenzio

With toxic behavior being one of the main reasons players quit most online multiplayer games, creating systems that combat these actions has never been more important. Game studios have recently began hiring teams of designers, scientists, and engineers to research the reasons behind online harassment in order to help reduce the amount of overall toxicity occurring within their respective games. These findings have shown that toxic behavior mainly stems from a lack of social cues due to anonymity, players being poor at self-assessment, and a lack of consequences for their behavior. While implementing changes to fix these issues, studios such as Riot Games and Valve have found that having quick response times to bad behavior, encouraging players to question their own behavior, and having said players define for themselves what is proper action have all helped solve the long term goal of fixing players' behavior. This approach does not only reduce the amount of toxic players, but also greatly reduces the amount of players who commit the action again in the future.

Toxic Behavior in Gaming

While toxic behavior in gaming has existed for some time, the most common area it occurs in is within online multiplayer gaming. More specifically, any online game that requires teamwork and some sort of text or voice communication can be especially susceptible to these problems. Some games that fall into this category include League of Legends, Overwatch, and Dota 2. All three of these games not only require active communication throughout each match, but if teammates underperform, the opposing team gains some type of advantage. Under these circumstances, not one sole player can win without the support of their team. However, players end up becoming more focused on how well their team is performing throughout the match. When these circumstances are in place, toxic behaviors are bound to happen.

Contrary to popular belief, the bulk of toxic behavior does not come from consistently toxic players, but most likely players that rarely engage in this type of behavior. In 2012, only 1% of all players in League of Legends were *consistently* toxic, and these players only contributed to 5% of the total toxicity generated day-by-day.[5] This means that to actually put a dent in the level of toxic behavior, they would need to find out why these otherwise normal players would lash out.

During an interview, Jeffrey 'Lyte' Lin, the former Lead Designer of Social Systems at Riot, spoke about a major cause of toxic behavior is due to the missed social cues from not being face-to-face with your teammates. "When you go on a basketball court, you don't really see this kind of toxicity ... You can often see when something you say is considered out of line,

you can see if you're causing some issues or you causing kind of harm or pain in somebody else in-person."[4] Without these social cues, players are forced to only rely on the little information they can get from their current match. If a player is underperforming due to having a rough day, their teammates will not be empathetic because all they see is our player "not trying".

Other reasons why players might be toxic include the idea of deindividuation and a lack of consequences for their actions. Deindividuation is when in groups, we tend to follow what the group wants to do, resulting in losing our self-awareness and sense of responsibility of our own actions. This can result in players believing their personal performance in a given match has little to do with their team failures. Also, people might become less aware of how toxic they are actually being. This idea, along with certain games not enforcing consequences for their behavior greatly adds to the amount of toxicity found within these genres.

Strategies to Decrease Toxic Behavior

Seeing these hurdles to overcome, gaming studios such as Riot Games, Valve, and Blizzard have created teams of designers, scientists and engineers specifically aimed at reducing toxicity levels in their games. Each team has taken slightly different approaches to the problem, and in return each has found different successes and failures. With the massive amount of players and matches that occur every day on these companies' servers, they have been able to perform large-scale tests to see whether these ideas have worked or not. Some of these tests have had majorly successful, while others were either thrown away or actively hurt the playerbase for a short while.

One simple but highly effective tactic that Valve and Riot both utilize is the use of priming. Priming is the act of showing a message before doing something to influence the action of the viewer. In Valve's case, they used a trick in where after each match they would give two dialog screens to the user. The first one asked the player to rate how well their teammates cooperated. Valve didn't actually care about this answer, they asked it so they could then give the next survey question: "Please rate the cooperation *you* displayed towards your teammates in the last match." The reason they asked this is to create cognitive dissonance in toxic players. Cognitive dissonance is when someone experiences discomfort from contradicting thoughts, and when this happens we seek to fix it. In this case, players who didn't play nice with their teammates are forced to either own up to themselves not having good teamwork, or put a high score down and acknowledge to themselves that they had poor rapport. This simple two question survey resulted in around 137,000 fewer reports per day, which comes out to about a 12.5% decrease of overall reports given out daily.

Riot Games has also come up with quite a few solutions to toxic behavior in their game, League of Legends. First off, they also made use of priming by simply posting 'tips' aimed towards having good teamwork on their loading screen. Just adding in a few sentences reduced negative attitude reports by 6.2%, verbal abuse reports by 8.3%, and offensive language by 11%.

Besides priming, Riot also created a system called the Tribunal to tackle the effects of deindividuation. The tribunal was a system in which players could opt-in to judge verbal abuse

reports made in matches in return for small in-game bonuses. This made the community define what is proper or improper to say in match. When the Tribunal deemed someone guilty of verbal abuse, the offending player would get a temporary ban, along with a ‘reform card’ which would inform them of why they were banned, as well as feedback to prevent being banned in the future. The tribunal, along with reform cards increased the reform rate of offending players to 70%[5].

The tribunal system, even with its great numbers still did face a major problem, turnover rate. A single report making its way all the way to the end of the process could take up to a month. To fix this, Riot opted to remove the tribunal system in favor of a newer machine learning system. This system attempts to learn certain behaviors on a case-by-case basis by taking in hundreds of thousands of data points each game, across all of their games. The system learns as much as it can, from what people write in in-game chat, to the way the player plays their character. After learning what words and actions are considered toxic behavior, the system could then ban and give feedback within an hour of a report. The incredibly fast feedback loops resulted in the reform rate reaching a high of 92%, quite a big improvement over the tribunal’s 70%.

Conclusions

When designing reporting and anti-toxicity systems in the future, there a few takeaways we can obtain from these examples. All of these systems and experiments that are successful in long-term improvement have a common property, which is *to make the player question their own behavior*. Banning players without teaching them what they did wrong will only result in a band-aid solution where the player will eventually come back without having learned anything. When the player questions their behavior, there is a much higher chance that they will not repeat the same offense.

Another aspect of these systems to note is that they consist of many smaller tests. As we can see from Riot Games, trying out many smaller experiments has helped them greatly decrease the influx of toxic behavior in League of Legends. Something as simple as adding a few helpful reminders, or making the player answer a few questions just to think about their behavior can do wonders for the health of the community.

As small as these changes are, we have to be careful of how fragile they can be at times. If a few dialog boxes can decrease offensive language used daily, then it is just as feasible for a minor feature to increase it as well. When designing any part of an online experience, we should consider the psychological effects it can have on the playerbase. In the future, hopefully more game studios will take this approach as they develop new multiplayer experiences.

References

1. M. Becker, “Dota 2 attempts to curb community toxicity for new players,” *App Trigger*, 29-Jul-2017. [Online]. Available: https://apptrigger.com/2017/07/29/dota-2-attempts-curb-community-toxicity-new-players/. [Accessed: 2017].

2. M. Gault, “How Valve Tricked Players into Being Less Toxic,” *Motherboard*, 06-Nov-2016. [Online]. Available: https://motherboard.vice.com/en_us/article/qkjned/how-valve-tricked-players-into-being-less-toxic. [Accessed: 2017].

3. A. Jones, “Dota 2's matchmaking system no longer depends on your behaviour score,” *PCGamesN*, 05-Oct-2017. [Online]. Available: https://www.pcgamesn.com/dota-2/dota-2-matchmaking-behaviour-score. [Accessed: 2017].

4. J. Madigan, “Podcast 4: Toxic Behavior,” *The Psychology of Video Games*, 20-Apr-2017. [Online]. Available: http://www.psychologyofgames.com/2015/06/podcast-4-toxic-behavior/. [Accessed: 2017].

5. B. Maher, “Can a Video Game Company Tame Toxic Behavior?,” *Scientific American*, 31-Mar-2016. [Online]. Available: https://www.scientificamerican.com/article/can-a-video-game-company-tame-toxic-behavior/. [Accessed: 14-Nov-2017].