BY BRENDAN MAHER

GOOD GAMING

SCIENTISTS ARE HELPING TO TAME TOXIC BEHAVIOUR IN THE WORLD'S MOST POPULAR ONLINE GAME.

NEXT STOP — *A KINDER INTERNET?*

I t took less than a minute of playing *League of Legends* for a homophobic slur to pop up on my screen. Actually, I hadn't even started playing. It was my first attempt to join what many agree to be the world's leading online game, and I was slow to pick a character. The messages started to pour in.

"Pick one, kidd," one nudged.

Then, "Choose FA GO TT."

It was an unusual spelling, and the spaces may have been added to ease the word past the game's default vulgarity filter, but the message was clear.

Online gamers have a reputation for hostility. In a largely consequence-free environment inhabited mostly by anonymous and competitive young men, the antics can be downright nasty. Players harass one another for not performing well and can cheat, sabotage games and do any number of things to intentionally ruin the experience for others — a practice that gamers refer to as griefing.

Racist, sexist and homophobic language is rampant; aggressors often threaten violence or urge a player to commit suicide; and from time to time, the vitriol spills beyond the confines of the game. In the notorious 'gamergate' controversy that erupted in late 2014, several women

A team from South Korea won the 2013 *League of Legends* **world championship in Los Angeles.**
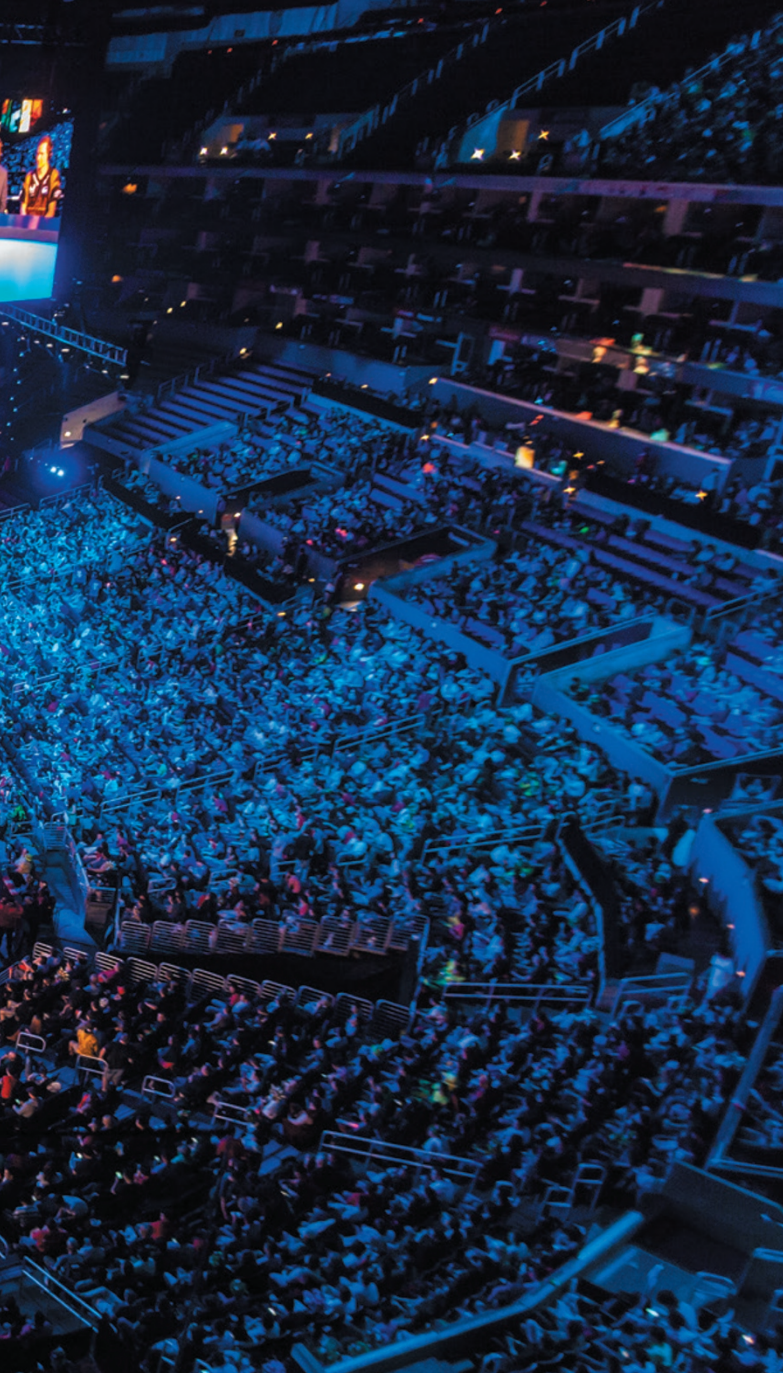
involved in the gaming industry were subjected to a campaign of harassment, including invasions of privacy and threats of death and rape.

League of Legends has 67 million players and grossed an estimated US$1.25 billion in revenue last year. But it also has a reputation for toxic in-game behaviour, which its parent company, Riot Games in Los Angeles, California, sees as an obstacle to attracting and retaining players. So the company has hired a team of researchers to study the social — and antisocial — interactions between its users. With so many players, the scientists have been able to gather vast amounts of behavioural data and to conduct experiments on a scale that is rarely achieved in academic settings.

Whereas other game companies have similar research teams, Riot's has been remarkably open about its work — with players, with other companies and with a growing collection of academic collaborators who see multiplayer games as a Petri dish for studying human behaviour. "What's most interesting with Riot is not that they're doing it but that they're publicizing it and have an established way of sharing it with academics," says Nick Yee, a social scientist and co-founder of Quantic Foundry, a video-game-industry consulting firm in Sunnyvale, California.

Riot's findings have helped to reveal where the toxic behaviour comes

from and how to steer players to be kinder to each other. And some say that the work may translate to digital venues outside the game. "The work they do is extensible to thinking about big questions," says Justin Reich, an education researcher at the Massachusetts Institute of Technology in Cambridge, "not just how do we make online games more civil places, but how do we make the Internet a more civil place?"

BIG BUSINESS

Jeffrey Lin, the lead designer of social systems at Riot, is the public face of its research programme. He has been playing video games online since he was about 11 years old and had long wondered why so many of his fellow gamers put up with toxic behaviour. "Everybody you talk to thinks of the Internet as this hate-filled place," he says. "Why do we think that's a normal part of gaming experiences?"

In 2012, Lin was finishing a PhD in cognitive neuroscience at the University of Washington in Seattle and was working for the game company Valve in nearby Bellevue when a friend and fellow gamer introduced him to the co-founders of Riot, Marc Merrill and Brandon Beck. They had recognized that toxic behaviour was a major drag on players' experience, and they wanted to solve the problem with science. So they hired Lin as a game designer, essentially giving him the keys to a juggernaut in the online gaming world.

League of Legends, Riot's only game, was released in 2009 and currently attracts 27 million players each day. It is by far the most popular of a growing segment of games referred to as eSports, a world in which elite players form professional teams, win university scholarships and take part in million-dollar tournaments in sporting arenas. The final of *League of Legends*'s 2015 world championship in Berlin drew 36 million viewers online and on television, rivalling the audience of the finals of some traditional sports.

The game can be intimidating to newcomers. Players control one of more than 120 characters called champions, each of which has specific abilities, weaknesses and roles. Teams are usually made up of five players, who must cooperate to kill monsters and opponents, collect gold to purchase magical items, capture territory and eventually destroy the other team's base.

Matches last about half an hour on average, so having a poorly performing player on a team can be aggravating. And the game requires coordination between players, for which it provides an in-game chat function. If someone makes a mistake, he or she will generally hear about it fast. Players can report their teammates for being toxic, and this can result in a temporary or permanent ban from the game. But working out how to distinguish a few frustrated grumbles or good-natured trash talk from the kind of vitriol that is worthy of punishment is a difficult task.

To tackle it, Lin needed to make sure that he had a good picture of where such toxicity was coming from. So he got a team to review chat logs from thousands of games each day and to code statements from players as positive, neutral or negative.

The resulting map of toxic behaviour was surprising. Common wisdom holds that the bulk of the cruelty on the Internet comes from a sliver of its inhabitants — the trolls. Indeed, Lin's team found that only about 1% of players were consistently toxic. But it turned out that these trolls produced only about 5% of the toxicity in *League of Legends*. "The vast majority was from the average person just having a bad day," says Lin. They behaved well for the most part, but lashed out on rare occasions.

That meant that even if Riot banned all the most toxic players, it might not have a big impact. To reduce the bad behaviour that most players experienced, the company would have to change how players act.

Lin borrowed a concept from classic psychology. In late 2012, he initiated a massive test of priming, the idea that imagery or messages presented just before an activity can nudge behaviours in one direction or another.

The Riot team devised 24 in-game messages or tips, including some that encourage good behaviour — such as "Players perform better if you give them constructive feedback after a mistake" — and some that discourage bad behaviour: "Teammates perform worse if you harass them after a mistake". They presented the tips in three colours and at different times during

the game. All told, there were 216 conditions to test against a control, in which no tips were given. That is a ridiculous number of permutations to test on people in a laboratory, but trivial for a company with the power to perform millions of experiments each day.

Some of the tips had a clear impact. The warning about harassment leading to poor performance reduced negative attitudes by 8.3%, verbal abuse by 6.2% and offensive language by 11% compared with controls. But the tip had a strong influence only when presented in red, a colour commonly associated with error avoidance in Western cultures. A positive message about players' cooperation reduced offensive language by 6.2%, and had smaller benefits in other categories. Riot has released just a few of these analyses, so it is hard to make broad generalizations.

From a scientific standpoint, says Lin, the results from the priming experiments were "epic", and they opened the doors to many more research questions, such as how various tips and colours might influence players from different cultures. But the behavioural improvements were too modest and too fleeting to change the culture of the game. Lin reasoned that if he wanted to make the community more civil, then players would have to have a say in devising the norms. So Riot introduced the Tribunal, which gives players a chance to serve as judge and jury to their peers. In it, volunteers review chat logs from a player who has been reported for bad behaviour, and then vote on whether the offender deserves punishment.

The Tribunal, which started in 2011, gave players a greater sense of control over establishing community norms, says Lin. And it revealed some of the things that triggered the most rebukes: homophobic and racial slurs. But players who were banned from the game were often unsure why they had been punished, and continued to act negatively when the bans were lifted. So Lin's team developed 'reform cards' to give feedback to banned players, and the company then monitored their play. When players were informed only of what kind of behaviour had landed them in trouble, 50% did not misbehave in a way that would warrant another punishment over the next three months. When they were sent reform cards that included the judgements from the Tribunal and that detailed the chats and actions that had resulted in the ban, the reform rate went up to 70%.

But the process was slow; reform cards might not show up until two weeks to a month after an offence. "If you look at any classic literature on reinforcement learning, the timing of feedback is super critical," says Lin. So he and his team used the copious data they were collecting to train a computer to do the work much more quickly. "We let loose machine learning," Lin says. The automated system could provide nearly instantaneous feedback; and when abuse reports arrived within 5–10 minutes of an offence, the reform rate climbed to 92%. Since that system was switched on, Lin says, verbal toxicity among so-called ranked games, which are the most competitive — and most vitriolic — dropped by 40%. Globally, he says, the occurrence of hate speech, sexism, racism, death threats and other types of extreme abuse is down to 2% of all games.

"If the numbers they put out there are correct and true, it seems to be working well," says Jamie Madigan, an author in St Louis, Missouri, who writes about the psychology of gamers. And that's because the reprimands are specific, timely and easy to understand and act upon, he says. "That's classic psychology 101."

OPEN DATA

Riot's research team is constantly experimenting with other ways to improve interactions in the game. Sportsmanlike behaviour can earn players honour points and other rewards. Tinkering with chat features

helped, too. And the team is planning to use the Tribunal to train the game's algorithms to detect sarcastic and passive-aggressive language in chats — a major challenge for machine learning.

From the start, Riot has also made much of its data available for others to investigate. Jeremy Blackburn, an avid gamer and computer scientist who works at Telefonica Research and Development in Barcelona, Spain, mined data on 1.46 million Tribunal cases to develop his own machine-learning approach for predicting when player behaviours would be deemed toxic. Together with Haewoon Kwak at the Qatar Computing Research Institute in Doha, he found that the most important factor — beyond the specific words used in the toxic messages — was how well the opposing team performed[1]. Blackburn, who is interested in studying cyberbullying, hopes to look more at how different cultures judge behaviour. Some evidence, he says, suggests that it is common for Korean gamers to gang up on and berate the poorest-performing players, for example. League data may bear this out. "We saw there was a lot more pardon for this verbal-abuse category."

Rachel Kowert in Austin, Texas, is a research psychologist on the board of the Digital Games Research Association. She is impressed by the work and especially by Blackburn and Kwak's unfettered access. "It's awesome for the researchers. You can't put a price on real data," she says.

Other companies also have data that scientists would like. Blizzard Entertainment in Irvine, California, makes the popular online fantasy game *World of Warcraft*, which many regard as a treasure trove for data on complex social interactions. But few people outside the company have been able work with the data, and most of those who do are subject to stiff non-disclosure agreements. (Blizzard did not respond to *Nature*'s requests for comment.)

By contrast, Riot talks about its data at gaming conferences, and when it collaborates with researchers there are few restrictions on publishing. It also has an outreach programme, visiting universities to establish collaborations. And last May, Lin presented data at the annual meeting of the Association for Psychological Science in New York City to drum up more interest.

"IT'S AWESOME FOR THE RESEARCHERS. YOU CAN'T PUT A PRICE ON REAL DATA."

Even with those efforts, the company's research has yet to achieve broad recognition among behavioural scientists. "Hopefully they will come to more conferences where people are studying behaviour," says Betsy Levy Paluck, a social psychologist at Princeton University in New Jersey. Although she was not familiar with Riot, she says that the company seems to be working out how to do high-powered, big-data research in psychology, which has been a major challenge.

Daphné Bavelier, a cognitive neuroscientist at the University of Geneva in Switzerland, met Lin at the conference in New York City. Her research has suggested — to the joy of many gamers and the agony of their parents — that some games, particularly fast-paced first-person shooters, can improve a handful of cognitive abilities, such as visual attention, both within and outside the games[2]. She plans to collaborate with Riot to study how players tackle the steep learning curve in *League of Legends*.

The team-based nature of the game could also be useful to scientists. Young Ji Kim, a social scientist at the Massachusetts Institute of Technology's Center for Collective Intelligence, was able to recruit 279 experienced teams from *League of Legends* to fill out surveys and work together on a battery of online tests that were designed to explore team dynamics and the factors that make teams successful. (By providing an in-game incentive worth about $15, Riot helped her team to get thousands of sign-ups in a couple of hours, she says.) The preliminary results suggested that the teams' rank in the game correlates with their collective intelligence — a measure that generally tracks with things such as social perceptiveness and taking equal turns in conversation.
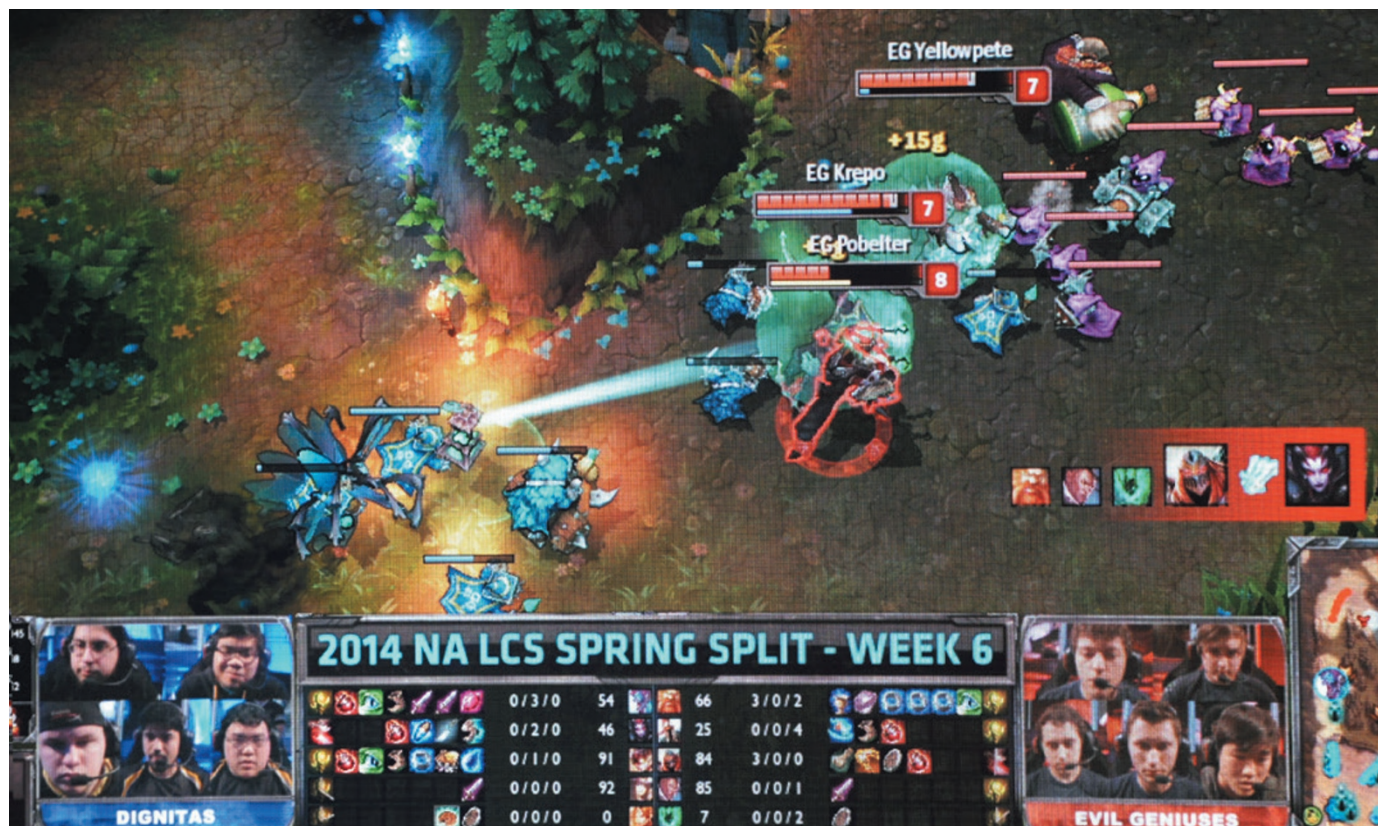
The enthusiasm that players show for participating in experiments such as these may be attributable to Lin, who writes frequently about Riot research and can often be found answering players' questions on Twitter and other social media. Being upfront and public about the efforts is important, says Bavelier. Although most digital companies run experiments on users, they are often less transparent. Facebook, for example, published a study about how behind-the-scenes tinkering with news feeds can manipulate user emotions[3], and received significant backlash from users. "We need to learn from some of the mistakes of others to make sure that the users are aware of what we're doing," says Bavelier.

Riot has an internal institutional review board that evaluates the ethics of all its experiments. Although not a conflict-free arrangement, it at least suggests that the research is being reviewed with an eye towards participant protection. Academic collaborators also need to get approval from their local boards.

VIRTUAL VIOLENCE

Lin has lofty goals for his teams' research and interventions. "Can we improve online society as a whole? Can we learn about how to teach etiquette?" he asks. "We're not an edutainment company. We're a games company first, but we're aware of how it could be used to educate."

Parents, lawmakers and some scientists have fretted for decades that video games, particularly violent ones, are warping the minds of children. But James Ivory, a communication scientist at Virginia Polytechnic Institute and State University, in Blacksburg, says that much of the attention on violence has missed the biggest impact that games have. "Researchers are slowly starting to wise to the idea that it may not be as important to think of what it means for someone to pretend to be a soldier than whether they're spending their time spewing racial or homophobic slurs."

By the age of 21, the average young gamer will have logged thousands of hours of playing time. That fact alone makes dichotomies such as 'real world' and 'digital world' ring false — for many, game-playing is the real world. And, says Ivory, "the strongest influence these games have on people is how they interact with other people".

League of Legends **requires teammates to communicate and cooperate.**

Some researchers are cautious about trying to apply lessons from the game to other settings. Dmitri Williams, a social scientist and founder of Ninja Metrics, an analytics company in Manhattan Beach, California, warns that games have very specific incentive structures, which could limit how well these experiences map to the wider world. "People behave well in real life because if they offend someone or screw up, they have to deal with the consequences." So, the manipulations that work to curb bad behaviour in *League* may be meaningless elsewhere.

And there are still considerable challenges for Riot. Players continue to complain about toxic behaviour or what they deem to be unwarranted punishments. And a blog called 'League of Sexism' argues that the suggestive portrayal of female characters in the game contributes to a strong current of sexism in the player community. "It's difficult for players to identify sexist behaviour when sexism is built into the game's very imagery," says a representative for the blog, who wished to remain anonymous. Although Lin's efforts are "admirable and likely industry-leading", the representative says, many games are still "awash with verbal harassment, griefing and overall negative behaviour from teammates and opponents". Lin says that Riot artists are aware of these concerns and that they have made efforts to portray female characters in a stronger and more-powerful way.

Although Riot boasts that serious toxic behaviour infects only 2% of games, somehow I managed to experience it within a minute of playing for the first time. But immediately after "FA GO TT" popped up on my screen, something interesting happened. Another player chimed in with, "Calm down". Perhaps it was a sign that Lin's efforts to engineer a more civil, self-policing digital space is starting to work. Or maybe it was just a friendly teammate reminding us all that it's just a game. ■ SEE EDITORIAL P.549

Brendan Maher *is a features editor with* Nature.

1. Blackburn, J. & Kwak, H. Preprint at http://arxiv.org/abs/1404.5905 (2014).
2. Green, C. S. & Bavelier, D. *Nature* **423,** 534–537 (2003).
3. Kramer, A. D. I., Guillory, J. E. & Hancock, J. T. *Proc. Natl Acad. Sci. USA* **111,** 8788–8790 (2014).