



Big playerbase, big data: On data analytics methodologies and their applicability to studying multiplayer games and culture by Ben Egliston

Abstract

Rapport with big data is something of a methodological rarity in empirical work on videogames, particularly within humanities oriented literature; an unusual omission considering the scope of many multiplayer game environments. Addressing this, the present work ventures the question 'how can research into multiplayer videogames benefit from the use of big data'? I offer a response through a case study of Valve Software's multiplayer game Dota 2, presenting a number of approaches which draw on player data analytics. In addition to mapping out frameworks for empirical research, I explore the theoretical dimensions of porting analytics based approaches to studies of multiplayer videogames, charting perceived incompatibilities between analytics approaches and popular ontologies of play, and how the prevalence of relational ontologies of play privilege particular modes of empirical inquiry.

Contents

- [1. Introduction](#)
- [2. Why do we need big data in the study of multiplayer games?](#)
- [3. Platform specificity](#)
- [4. Approaches](#)
- [5. From APIs to paratexts: Emerging ethical and legal issues](#)
- [6. Summarising remarks](#)
- [7. Game studies' disciplinary and methodological tensions](#)
- [8. Temporal orientations of big data: Immediacy of ethnographies and player analytics' past alignments](#)
- [9. Conclusion](#)

1. Introduction

In recent years, big data has radically altered the landscape of research in the humanities and social sciences. Ostensibly forging a new research paradigm (Kitchin, 2014), and propelling forward the humanities, big data research has proffered an array of novel insights into digital life and culture. Videogame research, however, has been surprisingly limited in its participation in much of the academic discourse surrounding the uptake of big data driven, analytics-related methodologies. This absence of analytics driven research is particularly pronounced in humanities oriented approaches to game studies, such as those rooted in cultural and media studies. This is particularly unusual, given the existence of data analytics infrastructures in many contemporary multiplayer games, as well as the longstanding tradition of developers using analytics to track player feedback internally for quality assurance purposes (Derosa, 2007; Seif El-Nasr, *et al.*, 2013).

Addressing the paucity of humanities oriented literature working with player analytics, this paper proposes the question, 'how can videogame research leverage and benefit from these emerging methodological assay?' As a point of focus, I explore how large scale research into multiplayer videogame spaces could apply data analytics. I provide an account for the productive use of humanities driven analytics game studies, using a

series of detailed examples from Valve Software's Dota 2 (Valve, 2013), a highly popular multiplayer game [1]. I outline applications of its numerous player data logging platforms. While the present methodology has been developed for the study of Dota 2, a by-product of a larger, ongoing research project, I attempt to broadly synthesise methods and outcomes in a way that can be ported to other multiplayer videogames. Importantly, I orient this work toward researchers of games located within humanities based disciplines. Researchers of games in fields of cultural or media studies (for instance) could productively apply similar methods and tools to those discussed in exploring videogames and their related cultural, social, technical, somatic or temporal phenomena.

This paper focuses on two principal methods of sourcing data: the utilisation of third-party data aggregation platforms (typically used by players, in order to guide their own play — characterised as a kind of paratext), and the use of Dota 2's proprietary Web Application Programming Interface (API); and the way it is used to create custom-purpose analytics tools. Both approaches avail researchers to archival nexuses, host to both player-game interaction data and textual communication data. I outline the ways in which these platforms (and platforms of their nature) can be used, and have been used, in the study of multiplayer game spaces, and their inherent practices and cultures. In addition to underlining existing applications of work, I speculate on ways in which prospective research could draw on the affordances of player analytics in the study of multiplayer games. For example, I argue that databases of player statistics represent a point of ingress into histories of play and videogame culture, offering answers to many pressing questions in the study of games.

Further to sketching out the methodological utility of player data analytics in game studies, I address a number of other important issues pertaining to the uptake of this approach. I briefly identify a number of problematic ethical and legal issues emerging from the adoption of nascent data analytic software which researchers should be equipped to deal with. Beyond identifying broader ethical and legal problems, I also articulate particular methodological limitations with the two proposed approaches to data collection. This paper presents both a pragmatic and theoretical look at emerging methods for researching multiplayer videogames. Accompanying outlines of methods and potentialities of application, I situate data analytics amongst existing, prevailing traditions of researching videogames in the humanities; with particular focus placed on ethnomethodological work situating the researcher as player. Further, I explore how popular ontologies used to frame the experience of play (as well as broader theories of media consumption) can orient researchers toward ethnographic lines of inquiry.



2. Why do we need big data in the study of multiplayer games?

Over the course of the past decade, the composition of playerbases has shifted considerably; game spaces are now populated by a vast and variegated spectrum of individuals, many of whom are far from aligned with the sensibilities of the traditional 'gamer' (Golding, 2014; Anthropy, 2012; Shaw, 2011). With this broadening audience, unsurprisingly, has come larger playerbases. Games like Dota 2 boast concurrent player stats peaking well into the millions (Good, 2015).

It is clear that videogame communities are no longer a niche subset of a broader screen-media-ecology. It is then necessary, I argue, that the existing methodological toolkit for exploring multiplayer game environments and modalities be updated, so that researchers are better equipped to deal with the contemporary cultural milieu of playing together.

Unsurprisingly, one of the benefits (although, limitations have been defined in the literature, see boyd and Crawford, 2012) of big data is the scope of the data itself (Kitchin, 2014). Additionally, the benefits of the heterogeneity of the data, facilitated by its sheer scale, have been noted by scholars (Kitchin, 2014). While, as I note in the coming sections, there were indeed methodological challenges in *realising* the heterogeneity through the data, the computational turn in studies of media would appear generally useful in collating data that is both vast and varied.

In terms of games research there is a fairly limited body of work, emerging from roughly 2012 and beyond, using quantitative utilities of analytics and logged player data, principally situated in research fields of human-computer interaction, computational intelligence, data science and computer science (Lim and Harrell, 2015; Drachen, *et al.*, 2014; Seif El-Nasr, *et al.*, 2013). Indeed, while emerging analytics related work in game studies owes much to these kinds of applications, the research questions, values and expected outcomes of the research differ considerably. Using analytics oriented approaches in (humanities) game studies could be productively applied in thinking about the emergence and formation of play and cultures of play. A popular humanities oriented approaches to game studies is to situate play as a field of relations; an array of intermingling forces, including the social and cultural, as well as materialities, (somatic) physicalities and temporalities. Through the numerous examples I provide in this paper, I argue that studying games from such conceptual standpoints can be enhanced through analytics related methods; bridging the values and outcomes of humanities approaches to game studies with the quantitative data-driven methodological rubrics of information sciences.

To be sure, there is immense value in 'small data' (as I argue later). However, my purpose in this paper is to make a case for the analysis of large sets of computationally retrieved player data in game studies. Ultimately, I contend that analytics represent a point of ingress for researchers into practically developing rigorous projects situated around large bodies of players interacting with a videogame.

3. Platform specificity

The claims I make in this paper are generic by nature (within the field of multiplayer games), yet find foothold in my example of Dota 2. Granted, there are obvious problems with the proposition and development of frameworks specific for the analysis of a single platform, which could result in limited degrees of transferability. However, to overcome issues of platform-specificity and limited transferability, the approaches and provocations I offer can easily be ported other relevant multiplayer games. Games such as League of Legends — ostensibly, the most popular game in the world (Funk, 2013) — are functionally similar and could easily draw in methods I discuss. In terms of academic relevance, the body of research situated around the MOBA genre is steadily growing (Donaldson, 2015; Ferrari, 2013; Egliston, 2015; Drachen, *et al.*, 2014). In addition, the growing cultural salience of multiplayer games — which we can see through emerging fixtures such as e-sports (Taylor, 2012) — indicates that while methods have been developed with particular ground in mind, this field is by no means fallow.

4. Approaches

There are two broad camps into which I situate methodological approaches to player data analytics. The first is the use of analytics based paratexts, objects surrounding the videogame proper, as a form of *ad hoc* research device. The second, is the practice of directly leveraging the game's API, and the way this can be used to create custom data-tracking applications. These approaches both deal with large bodies of player data, but provide the researcher with findings varying in degrees of granularity. I outline these approaches, provide examples of use (and speculate on future uses), and identify their respective limitations.

4.1. Reaching into the periphery: Paratexts and analytics

Paratextuality, as per Gérard Genette's original formulation, is the idea that literary texts possess a significant field of textual relations. Texts, essentially, are part of a network with other textual bodies surrounding them. Paratextuality, according to Genette, serves a functional, instructive purpose: to further shape and enhance our understanding of the text, framing and mediating our perception [2]. There is a growing literature situating videogame paratexts as sites of importance for researchers. Drawing principally from Genette's work, scholars of videogames have, over the last decade, appropriated paratextuality in various ways as a means to study videogames, practices and players.

Mia Consalvo (2007) argues that various kinds of videogame paratext are central as both objects of play and objects of study. Consalvo contends that gaming paratexts are central in the play of games, serving "a specific role in gaming culture ... they instruct a player in how to play" [3]. In arriving at this position, Consalvo interrogates a number of 'gaming paratexts', including videogame magazines. Importantly, Consalvo negotiates paratextual spaces as sites of research significance; many scholars since have methodologically utilised paratextual media in game studies (Carter, 2014; Kirkpatrick, 2012). Further solidifying the research potential of paratextual material, Thomas Apperley argues that "the relationship between digital games and their paratexts is an example of how the convergent audience uses other media, especially the internet, to collaborate on, conduct and coordinate research" [4]. This body of work does well to argue that the texts surrounding games can tell us much about the game-proper and its associated practices.

In the present work, these paratexts are repositories of player data. Web-based data logging applications function as an important kind of paratext; useful in shaping understandings of games for players and researchers alike. Situated at the periphery of the game itself, these platforms provide utility to players in searching for ingame strategies and researchers (as *ad hoc* research tools) in identifying trends in play. One of the analytics platforms discussed in the present work purports to "give you the tools you need to improve" (Dotabuff, 2016).

Despite the potential that these applications show, such *ad hoc* research tools are relatively rarely utilised in studies of games. However, there is a fairly limited literature using player paratexts like Dotabuff to make claims on games (Egliston, 2015a).

4.1.1. Examples

I will outline three applications (Dotabuff, YASP and Datdota), their existing, provisional uses as well as their prospective uses as research apparatus. All three are free to use (at a base level), Web-based applications. This, I feel, is especially important and influenced selection considerably. As researchers of digital media, situated within the field of the humanities and social sciences, many of us are inadequately equipped to undertake computationally demanding tasks such as API scraping. Platforms such as the following are beneficial insofar that they enable big data research for those outside of computationally oriented disciplines. Indeed, computational literacy is a notable barrier to quantitative, data-related research in the social sciences (Zelenkauskaitė and Bucy, 2016), which can be translated to humanities. In line with the publically visible, easily navigable and parsable interface data that Zelenkauskaitė and Bucy (2016) describe, gaming paratexts-like Web apps provide many kinds of researchers, with varying degrees of computational literacy, the opportunity to augment or guide their research with large-scale player data.

4.1.2 Dotabuff and YASP

Dotabuff is a popular stat-tracking tool in the Dota 2 community. Games of Dota 2, and the myriad player inputs which occur are parsed, consolidated and presented diagrammatically and proportionally. A Web API is used which parses replays of Dota 2 matches. Dotabuff affords the user the ability to explore trends occurring across the world in Dota 2, allowing users to answer questions such as the following: What strategies are people using? What characters and items are proving effective? Where and when are new playstyles emerging? What playstyles are *unpopular*? Table 1 provides an overview of how and what Dotabuff data can tell researchers.

Hero pickrates, for instance, are incredibly illuminating when thinking about how groups of players are engaging with the game. In Dota 2, heroes are the characters that players control; each of which are relatively unique and have their own gameplay affordances. By measuring the popularity (or unpopularity) of heroes, we can roughly gauge the *styles* of play and, moreover, user experience. The indexes of play here achieve similar ends to the behavioural metrics for purposes of player profiling in computer and data science oriented games literature (for example, Seif El-Nasr, *et al.*, 2013). It is worth noting, also, that these kind of behavioural metrics, generated through an analysis of how players *play* have existed in studies of games for quite some time. Richard Bartle's eponymous Bartle Test (which provides a taxonomy of types of game players) is key (Bartle, 1996). Further, it has given seed to other work employing and developing it over large scale research (Williams, *et al.*, 2008). Evidently, there is a clear precedent for using data-driven player profiling. Marrying these techniques with playerbase-wide data, would allow researchers to develop broad understandings of playerbases.

Table 1: Indexical markers of playstyle, provided by Dotabuff application.		
Data	Function	Potential indexical marker or application
Hero	Search for matches with a specific hero (Character) being played	Playstyle; how people play the game
Lane	Location of player on map (first ~10 minutes of game)	Lanes generally correspond to a particular style of play
Win rate	The rate which particular heroes, lanes, are won	Efficacy of playstyle

Importantly, not only does Dotabuff provide player data analytics, but retains this data. Dotabuff has functioned as a repository of game data covering the game's entire lifespan (of roughly four years, at time of writing). Access to years of game data facilitates longitudinal research alongside other important affordances which I will later outline. Important to note, however, is Dotabuff is an opt-in tool (games where at least one player has an account registered with Dotabuff will be parsed).

Previous work (a pilot study for the research project, from which the present work draws) has profitably applied Dotabuff as a research tool. The platform was used to identify associations between distinct groups of players, indexing similarities in game approaches through player choices in game (Egliston, 2015a).

Another research application of the platform used Dotabuff to explore regional play. Players are regionally separated in Dota 2, with players generally choosing to play on the region closest to their geographical location (in order to minimise latency between their network connection and the game server).

Building on the aforementioned playstyle framework, tentative work highlighted that play is thoroughly regional, and shaped by cultural contexts and local technical infrastructures. For instance, Australia was located as a particularly sterile site for strategic innovation in competitive Dota 2 due to a perceived inability to communicate and develop playstyles through livestreamed media, due to poor internet infrastructure (Egliston, 2015b).

The ability to access region specific game data could be profitable in the development of both empirical research and theory of games and play. There is a growing body of research which has situated play as influenced by contextual factors, such as social and locational fields (Hjorth, 2010; Chan, 2008; Apperley, 2010). Using data sets which focusing on the geographical location of the player builds on this relevant area of research; suggesting that there is indeed a kind of geographical identity to play. [Figure 1](#), using Dotabuff, shows differences in playstyle based on the index 'hero pickrate', across the server regions Australia and U.S. West (along with the aggregate global pickrates).

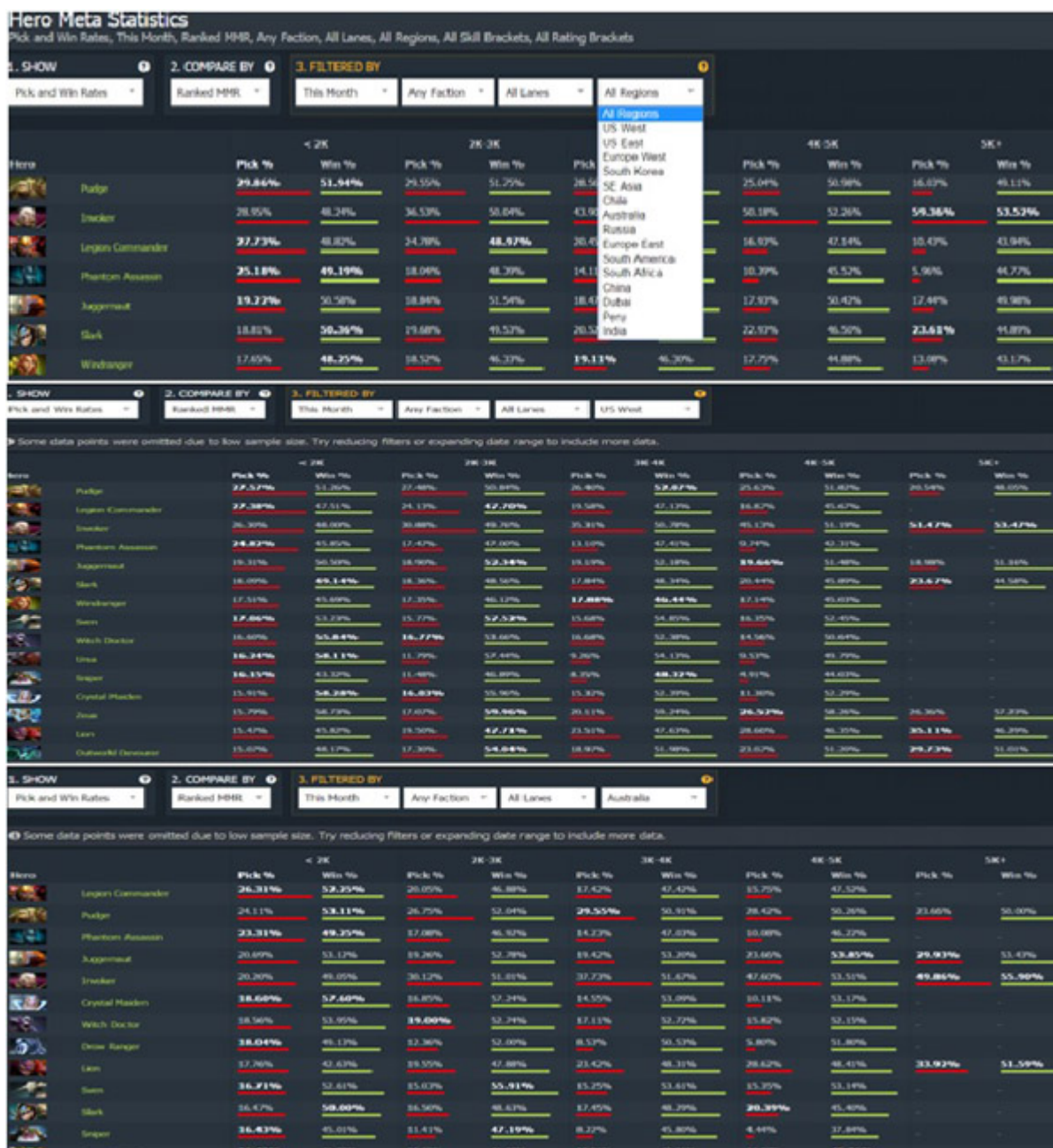


Figure 1: Regional data, Aggregate global pickrates, U.S. West and Australian server pickrates.

YASP is another useful analytics platform, providing large scale player statistics. Unlike Dotabuff, statistics do not extend to the game's earliest iterations so the data available is smaller. Compare, for instance, the tracked games using the 'Invoker' hero across Dotabuff and YASP. Dotabuff reports roughly 220 million games played, where YASP reports 30 million. The detail of analysis is significantly more thorough, however. For instance, YASP provides, as a base feature, the ability to map interactions between specific characters ingame. It also allows the user (registered with YASP) to generate word clouds based on the ingame text they have read. [Figure 2](#) shows this, taken over a series of 176 games. Prospective research could marry analytics based platforms like YASP with situated fieldwork; mapping out textual interactions between researcher and players. Indeed, multiplayer game spaces are just as much social in character as they are configurative possibility spaces. Through these kinds of emerging apparatus, researchers can explore social dynamics on a large scale. Parsed text communication could be subject to further content or lexical analyses [5].



#	class	#	property	value
---	-------	---	----------	-------



4.1.3. Datdota

Datdota collates player statistics solely from professional, tournament level games, occurring in the e-sports arena. As others have explored with respect to traditional sports (Hutchins, 2015), there has been a considerable shift in value of sports (or e-sports) data as of late; with outlets such as Datdota (and Hutchins' examples of baseball analytics) providing a source of data-based entertainment for sports followers. While almost identical in functionality to Dotabuff and YASP (except much more limited in scope; data sets of professional play are smaller than those of the broader playerbase), Datdota provides a number of visualisation options; for instance, heat maps of player movements in a particular game or, heat maps of player movements over a series of games. See [Figure 4](#). In addition, Datdota provides the option to model the game's map (that is, the space in which play occurs) on the Cartesian plane. Now, through this tool, it is possible to not only map out styles of play, but also to quantify how players engage with space (which could, potentially, in combination with the previous indexes of playstyle, be used to formulate a more refined operationalisation of play). Previous work (similarly situated around Dota 2) has employed methods to investigate spatio-temporal phenomena through quantitative means, looking at relatively large sets of data (Bauckhage, *et al.*, 2014). While this work located heatmaps as a relatively facile analytical tool (providing more in terms of visualisation rather than deep quantitative data) they still provide the ability to quantify and sketch out the array of ways in which players move through the game world, an inroads toward detailed and large scale research on spatial engagement with games.

In addition, DatDota provides utility in quantifying key player-game interface; particularly the somatic (haptic) aspects of play. Using key metrics developed to map and analyse haptic interaction (specifically, the actions per minute, or APM, metric), researchers are provided a gauge of players' level of tactile engagement and manual dexterity. Sourced from DatDota, researchers can explore aggregate APM of various playstyles. The way in which the hand is a site of transaction between the human and the material, mediating experience and perception is an idea of which much has been made. From the existential phenomenology of Martin Heidegger (1962), to Maurice Merleau-Ponty's (2002) phenomenology of perception, to more recent formulations of haptic-technocorporeality (recent work on videogames by Kirkpatrick, 2009; Ash, 2013), there is a considerable literature concerned with theorising the hand. Beyond this, within disciplines like media studies, there is a rich lineage of work looking at the relationship between the form of media, and the human user (McLuhan, 1974, and beyond) and an even wider lineage of media technology effects theory (see, Innes, 1950, to contemporaries like Kittler, 1990).

Through the collation of data showing player APM, researchers are provided a quantified snapshot of interface between the somatic rhythms of the user and the technological device (important to note; data-methods don't completely obscure the way important, physical constituents of play function. Rather, they present quantified readings). [Appendix 1](#) is a tabulated set of playstyles and their respective APM over a series of matches in Dota 2.

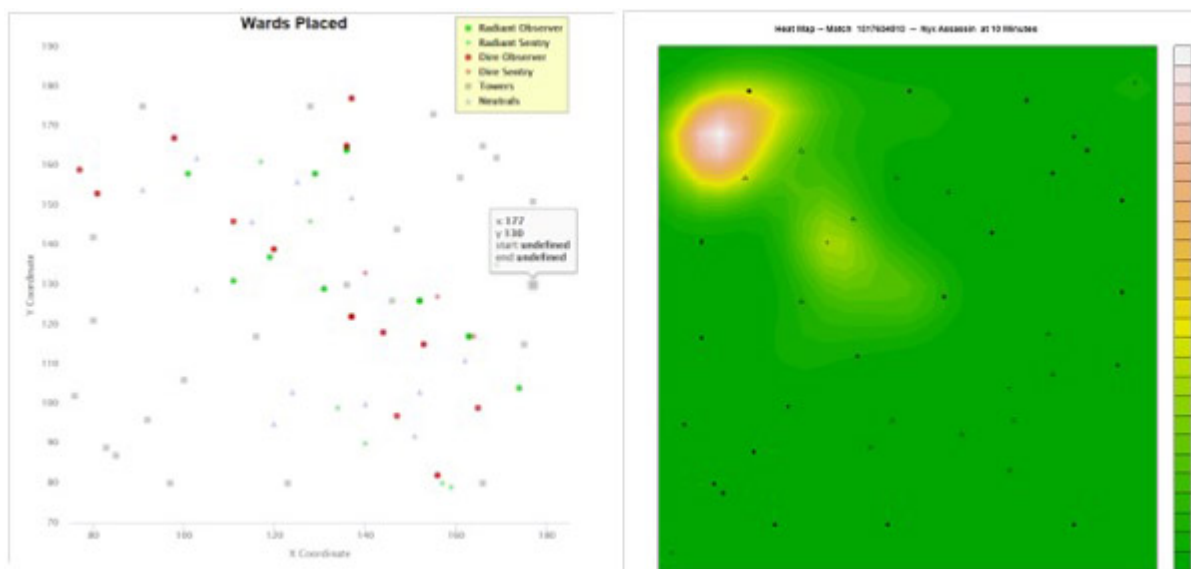


Figure 4: Visualisation of players' spatial engagement with game. Source: Datdota.com.

4.1.4. Example data set exercise

To show how researchers can utilise these platforms, I present an example drawing from data collected and hosted by Dotabuff and Datdota. The particular question I am attempting to answer through this data set is 'to what extent are gamers in the amateur, public arena of play influenced by events in the game's e-sporting

community?’ Here, I am concerned with generating a snapshot of not only playstyle but the influence of broader cultural practices and media structures.

Taken from a larger set of data, from an ongoing study, Tables 2 and 3 present statistics of two distinct groups of Dota 2 players: professional e-sports level players ([Table 3](#)) and amateur gamers ([Table 2](#)). Data was taken from a two-month period (February–March 2016) surrounding the e-sports event, showing the immediate influence of highly visible e-sports events (Event ranging from 3 February 2016–6 March 2016). Using data from Datdota to populate the ‘professional player’ dataset and Dotabuff to generate the ‘amateur player’ dataset, I am able to map out associations between distinct and important constituents of the game’s playerbase, as well as show the influence of broadcasting media, which showcase e-sport play.

Undertaking analysis of amateur player data, over the period February–March 2016, I was able to sketch out the ways in which approaches to gameplay were influenced by those broadcast in the e-sports arena (see [Table 2](#)). Results from an earlier, pilot study yielded similar results; suggesting associations between groups of players and broader media structures such as broadcast e-sport (Egliston, 2015a). This study invoked theories of paratextuality alongside its empirical findings, using analytics driven results to situate paratextuality as an important dimension of playing modern multiplayer games, providing a media-theoretical account of new gaming modalities.

Table 2: Amateur pickrates surrounding the broadcast of the ‘Shanghai Major’ Dota 2 tournament.

Source: Dotabuff.com.

Hero	Pickrate March 2016 (%)	Pickrate change (%) since previous month(February–March)
<i>Invoker</i>	36.53	+3.67
<i>Sven</i>	18.65	+5.30
<i>Outworld Devourer</i>	17.00	+5.01
<i>Spectre</i>	21.19	+3.62
<i>Ember Spirit</i>	11.26	-5.78
<i>Nature’s Prophet</i>	15.39	+1.51

Table 3: Pickrates of heroes by professional players at ‘Shanghai Major’ Dota 2 tournament, 3 February 2016–6 March 2016.

Source: DatDota.com.

Hero	Pickrate at tournament (%)
<i>Invoker</i>	42.05
<i>Sven</i>	27.27
<i>Outworld Devourer</i>	27.27
<i>Spectre</i>	18.18
<i>Ember Spirit</i>	17.05
<i>Nature’s Prophet</i>	40.91

While these figures are just a snapshot of data fetched for the study, analytics based platforms have so far proven highly generative in mapping out large-scale user activity around events in the game’s community. Future work could productively apply techniques in order to conduct research into other important play-influencing variables; such as content patches or widespread technical glitches. From these findings, researchers in the humanities have the ability to further develop and refine theories of what videogames fundamentally involve and demand.

4.1.5. Limitations of Web applications

While these paratextual Web applications have proven fruitful in provisional work, and do show promise as an accessible approach to studying player data, there are limitations that should be addressed.

Like much scholarship which has harnessed big data, the vast nature of player data can make claims and associations appear more tenable than they are in actuality. boyd and Crawford make this point with reference to research using Twitter user analytics. Particularly, they note that sets of big data are not

representative of all people, and that it would be “error to assume ‘people’ and ‘Twitter users’ are synonymous: they are a very particular sub-set” [6].

Like the Twitter example, the population of opt-in Web applications are far from representative of the game’s entire playerbase. The playerbase of Dota 2 is enormous, and while these platforms only require one player to be opted in to gather data (players who have not opted in are represented as anonymous, but stats are still parsed) they provide an incomplete account of the playerbase.

To return to one of the points with which I started, contemporary videogames are host to a varied ecology of players. Traditional conceptualisations of games and play are eroding (Apperley and Jayemanne, 2012; Shaw, 2011). Certainly, the dismantling of ideas of the traditional gamer is important and conducive to diversity in games. However, it introduces methodological challenges. In using opt-in platforms like Dotabuff, research is only drawing from an incomplete set of potential participants. Platforms like Dotabuff are developed to appeal to relatively experienced or oriented player set, providing them with means to analyse and reflect on their own play (and the play of others) in order to refine their own abilities in future play. By dealing with such platforms, there is indeed the risk of collecting data from an incomplete sample size; potentially excluding those who may only deal with the game in a cursory sense (and skewing claims made).

Drawing from a sample of players who are reasonably well-acclimated with the game recalls T.L Taylor’s idea of instrumentalised power play. Indeed, play is contextual in character, influenced by cultural, social and temporal variables [7]. As such, a coverall approach here is reductive, and by effacing kinds of play we are obscuring a great many ways to understand the culture, identity and sociality of play.

The omission of significant groups is an import shortcoming of opt-in platforms, and an obstacle in allowing researchers to make statements about playerbases and game cultures. Ultimately, while we may be equipped to deal with the *scale* of data, the heterogeneity of userbases and gaming approaches is not *wholly* accounted for using Web applications.

Moreover, the *kind of data* yielded is worth thinking about when using Web applications and metadata aggregators. Developers of data aggregators filter raw game data through a particular criteriological lens. This ‘data cleaning’ [8] represents a kind of secondary informational gatekeeping via user-generated visibility of information (drawing on Singer, 2014). Essentially, in using platforms like Dotabuff, *et al.* particular game data is made visible (based on app developers’ subjective criteria), and distributed amongst relevant networks through the Web app. Particular kinds of play are, indeed privileged; particularly those involving mechanical interaction. However, as recent work tells us, there are many significant game modalities that extend beyond the reaches of conventional ‘gameplay’ (Apperley and Jayemanne, 2012).

With no doubt is there considerable value in pre-visualized, aggregate data. It is visible, digestible and usable for most researchers. However, it is important to be mindful of what is *disaggregated*.

4.2. Dota 2 API

The second method of inquiry involves accessing the Dota2 WebAPI; a method much more grounded in computational finesse. A detailed walkthrough of the API is beyond the scope of this paper (and would not be particularly useful in transferability of claims). Instead, I underline key affordances, and show how it has been leveraged by developers.

The value of APIs as academic research tools has been realised in the efforts of social media studies (Lomborg and Bechmann, 2014). Studies situated around Twitter, for instance, have leveraged APIs in order to study communicative phenomena (Bruns and Liang, 2012).

The Dota2 WebAPI, like the APIs of many other platforms, is useful in identifying (with varying degrees of granularity, as I will discuss) ways in which individuals interact with the game. Unlike certain social media platforms, and other popular videogames, Dota 2’s API is publically accessible. This enables researchers and players alike to retrieve game history and details in JSON or XML format (to be modeled visually in other applications). The WebAPI allows researchers to extract information surrounding many aspects of the game, which can be operationalised as metrics for research. [Appendix 2](#) shows a number of strings and traceable events in the Dota 2 API. In addition, it shows the function of these strings as well as ways in which they could be used and, with particular emphasis on establishing indexical markers of play.

Certainly, much of the legwork in harvesting and aggregating data is done by Web applications. However, should researchers need to explore specific, and undocumented phenomenon in games, APIs present a strong explorative approach. Web developers and software engineers have made use of the API to track highly specific interactions (e.g., the datadrivendota application allows users to track health of player characters over a series of games). In addition, the API allows users to parse (with significant, manual work for the researcher) *all public* games, rather than just those recorded by a Web app (contingent on a user opting in); addressing the Web app limitation of diminishing some of big data’s heterogeneity. While we have the potential for more precise results, the ends to which researchers must go to extract and retrieve this kind of ‘covert data’ (to borrow from Zelenkauskaitė and Bucy, 2016) are considerable.

4.2.1. Limitations

As I have noted above, the manual labour involved in utilising the API can be significant. In addition to the resources expended in attaining data, researchers will also need to be acclimated with coding language and protocol. Unless considerably well versed in these protocols, the opportunity cost in attempting to manually obtain data is fairly high (in comparison to using existing tools). Users on the DotaDev forum also cite rate

limiting (how many calls users can make to the API) as a shortcoming of the platform, and as others have noted, something that should be considered when utilising APIs for research (Bruns and Liang, 2012).

Another limitation of APIs as a research tool is that it requires a degree of cooperation on the part of the software developer (not only with respect to making public the API, but also making clear its functionality via documentation). The Riot API's documentation (for League of Legends) clearly articulates API functionality, for instance (unlike the Dota 2 WebAPI).

This reliance on developer upkeep, and the positioning of researchers as end users has proven problematic in collating data using APIs. A recent example is particularly telling of this. In early June 2016, Valve's Dota 2 API was down due to technical issues; proving problematic for Web apps which fetched data using the API [9]. Where data-aggregating Web apps present a kind of informational gatekeeping, in that data is refracted through a particular subjective lens, there is a different kind of gatekeeping at play here; one around fundamental access to information.

Another limitation pertains to the nature of the data collected. Like webapps, the WebAPI is limited to fetching data on public games. While not a significant omission with the scope of total games played, privately played games (against both players and AI bots) still represent sites of importance. This recalls the emphasis Kate Crawford places on thinking about formulations of 'active' online participation and what is constitutive of participation [10]. Crawford, along with danah boyd, later argue that "Twitter Inc. has revealed that 40 percent of active users sign in just to listen (Twitter, 2011). The very meanings of 'user' and 'participation' and 'active' need to be critically examined." [11]. The same should be done for practices constitutive of play (the implication that players who play against AI are not 'real players', and thus are disaggregated from the playerbase, is troubling). Data sets and the metrics they provide cannot account for every important kind of videogame practice.



5. From APIs to paratexts: Emerging ethical and legal issues

As videogame specific analytics platforms emerge, so do issues pertaining to the ethics and legality of deploying such technologies for purposes of research. Practices of player data collection, retention and presentation represent nebulous ethical and legal ground.

Many of the issues associated with adopting nascent approaches to player data analytics are analogous to previous studies situated around big data in digital media research. Issues of privacy and user consent are unsurprisingly pertinent; areas which have been well debated and crystallised within existing literature (boyd and Crawford, 2012; Markham and Buchannan, 2012).

One particular ethically unclear area, on which I wanted to place particular focus, is that of the use of data generated by others. Users have essentially reported data theft, with data being ported to new platforms without permission [12]. While these resources prove generative, as researchers can we ethically (and legally) be part of this?



6. Summarising remarks

In summary, methods involving player data analytics, emerging from publically available proprietary APIs and Web-based data aggregators, have the potential for use in studies of games. These platforms allow researchers to scrape raw data, analyse pre-visualised data, and identify metrics of play; enabling the development of variable operationalisation. It is likely that the use of ready-made tools, the kinds of paratext I discussed prior, will be more conducive to performing videogame research in the realm of the humanities and social sciences, simply due to a requisite technical literacy in dealing with game APIs.

I have also outlined the limitations of these approaches. Importantly, and as has been amply ventilated elsewhere, big data is not necessarily the best, most complete or rigorous set of data (see Manovich, 2012; boyd and Crawford, 2012; Bruns, 2013, for more nuanced writings on limitations), and while these approaches show promise in providing accounts of game-related phenomena, they also highlight some of the critical flaws.

Additionally, the approaches presented here, like almost any research methodology or apparatus, are not exhaustive, and should not be deployed in isolation; they demand more rigorous methodological support. As recent work tells us, "Further efforts should be made to situate big social data outcomes as complementary to existing approaches rather than regarding them as analytical islands" (Zelenkauskaitė and Bucy, 2016). Certainly, many of the proposed approaches here encourage this, with Web apps providing relatively easy and digestible access to large-scale data; facilitating the marriage of research in humanities oriented disciplines (and overwhelmingly qualitative approaches) and more quantitative, data-driven research.

In an ideal research world, it *would* be best practice to congeal interdisciplinary approaches. However, as I outline in the next section, there are tensions, arising out of ontological formulations of videogames and play, which inhibit wider uptake of quantitative data-related assay in humanities game studies.



7. Game studies' disciplinary and methodological tensions

Over the course of the previous sections, I have outlined provisional and speculative applications of quantitative, analytics-related approaches to videogame research. The prevailing sentiment is that studies of multiplayer videogames would indeed benefit from a wider adoption of computationally-retrieved, player analytics related methods; with the potential to yield important insights into practices and cultures of play.

Despite data analytics becoming an increasingly relevant aspect of videogames, there is a notable paucity of work taking advantage of this growing data infrastructure. While it is possible that nascent approaches to player data analytics have merely not been realised (as Bauckhage, *et al.*, 2014 note) or that uptake is limited due to literacy [13], I argue that there is another potential reading, lying in the ontological location of games as interactive or played objects (see Keogh, 2014, for a good critique of these ideas); and in order to explore how meaning is created, we must, putatively, watch players *play*. To further unpack this, perhaps the 'detached' nature of player analytics research (while providing a rigorous documentation of events) represents an obfuscation of many of the important networks of relations situated around play with which players are engaged.

Play has been located as the result of interface between networks of materialities and bodies (Ash, 2013; Sudnow, 1983; Kirkpatrick, 2009). Others situate play as a cybernetic circuit (Swink, 2009). Recent work has channelled actor network theory in order to explore the assemblage of multiplayer games, interrogating the intermingling network of things and forces influencing gameplay (Taylor, 2009).

Indeed, it is fairly well established within games literature that play is the result emerging from contestations between players and other peripheral objects and forces. However, I argue this has resulted in researchers wanting to situate themselves in ways to directly observe the stage on which interactions, between human player, game software, physical material and machine, all unfold.

Certainly, analytics based approaches do estrange the researcher from a number of important constituents of play, specifically physical materialities. Or, perhaps researchers, particularly in the humanities, are not willing to reduce the many rich interfacial properties of play to a platform's statistical output (for instance, my previous example of quantifying the somatic undercurrents of play).

Ultimately, I argue that disaggregating researcher and play space, and drawing distance between the researcher from the familiarly material, bodily and other constituents of play provides, in many ways, a unique vantage point from which to observe play. Overwhelmingly though, it seems that videogame research approaches are very much freighted with conceptions of games as a complex and tightly interwoven network of actors; and the direct observation of this interface is key.

In addition to perceived incompatibilities at an ontological level, there are issues at a broader methodological level within cultural studies approaches to studying games. Methodologies, as Lawrence Grossberg, Gary Nelson and Paula A. Treichler tell us, "always bear traces of their history" [14]. This is no less relevant in researching games and the valorisation of ethnographic study; approaches situated around the direct observation of game communities, players and their practices.

The idea of the embedded, participant or observer researcher in ethnography has been a mainstay in cultural studies research for some time. Having realised the potential of audience experience and perception as a field worthy of study in and of itself, popularised and valorised by scholars like Henry Jenkins (1992), Grossberg, *et al.* argue that researchers moved toward embedded, participatory research techniques in order to best study that audience [15].

The idea of the researcher as a participant, having populated broader studies of media and culture, unsurprisingly materialised as a dominant approach to studying multiplayer games. Formative work in the study of games was quick to adopt (or suggest the adoption of) this methodology. Espen Aarseth (2007), essentially argues that playing videogames is a fundamental approach to research offering more significant insight into games and game practices than other non-situated methods [16]; and one that has been followed quite closely in ethnographic accounts of play (particularly multi-player instances) since. The modes of experience and perception of videogame players are manifold; games are vast possibility spaces and things which are engaged with in different capacities (Apperley and Jayemanne, 2012). Researchers have argued that key ways to explore the myriad kinds of engagement with games is through ethnomethodological means. Apperley and Jayemanne advocate, and map the utility of, situated ethnographic methods, arguing "for game studies, ethnography offers an approach that acknowledges the complex contexts in which game play takes place." [17] While I've argued that player analytics can offer generative and nuanced insights into contexts and styles of play, there are a great many things that are left obscure by computationally driven, quantitative approaches.

Many ethnographic interventions into game spaces have offered valuable insights into multiplayer games and their cultures (Boellstorff, *et al.*, 2012) as well as videogame communities of practice (Carter, 2014). Ethnographies of game communities have proven pivotal in evaluating aspects of multi-user play, such as sociocultural issues of xenophobia (Goodfellow, 2015) and the intersection of physical and virtual spaces (Apperley, 2010). Others have deployed autoethnographic accounts of play, underlining the experiences of coming to terms with new game environments and the myriad possibilities within (Giddings and Kennedy, 2008). These studies showcase the varied research potentialities of situated-researcher ethnographies and outline a number of aspects that would difficult to adequately document by scraping an API or drawing from

sets of player data. Understanding play as an assemblage of manifold forces is indeed something researchers are made particularly attentive to through ethnomethodological thick description.

My intention in this section has been to telegraph the ways in which ontologies of play, as well as prevailing ideas in media and cultural studies have informed research direction. The merits of data analytics, I have argued, are obscured by popular ways of thinking about and researching games. The outcomes of analytics, I've argued, are around broad portrayals of game communities and their practices. Certainly, researchers are still made attentive to key interactions between important stakeholders in play (e.g., charting the haptic via APM metrics), however, ultimately, these approaches sit incongruously alongside research interests in micro-level interactions (research interests which are more aligned with qualitative, ethnographic inquiry).



8. Temporal orientations of big data: Immediacy of ethnographies and player analytics' past alignments

Over the course of this paper, I have argued that humanities oriented games research would benefit from more analytics driven work; allowing researchers to make quantitatively driven claims about large groups of players.

Beyond the benefits of significant scale and heterogeneity in sample sizes, I argue that big data holds value as a research apparatus for videogame studies due to its temporal orientation. For the most part, researchers are drawing on recordings of the past. By contrast to the numerous examples of ethnography in the previous section, big data approaches outlined in this paper do not possess the same sense of immediacy. Through observational, or participatory approaches to studying multiplayer games, the researcher (and the research) is very much geared toward watching play *happen*. While the idea of watching *in situ*, and bearing witness to phenomena unfolding, is often lauded as one of the draw cards of the approach (Reeves, *et al.*, 2009), we occlude an important aspect of play; *past* play.

I argue that the aggregation of player data holds value as an apparatus in generating *histories and archaeologies of play*. Scholars have, as of late, been entangled in explorations of the rich historical dimensions of past videogames. Work has emerged exploring technical histories of games (Montfort and Bogost, 2009), and approaches invoking media archaeology (Apperley and Parrika, 2015). Others have written, more broadly on the need for conservational efforts in the field of videogames (Swalwell, 2013). Much of this work, however, focuses on some form of past materiality of the videogame object, tracing trajectories of being.

One such area of gaming's past, however, which has remained difficult to explore, is that of past play practice. How best do videogame researchers reliably and rigorously explore interactions within a videogame that are so ephemeral? How can we document meaningful cultural traditions occurring within these spaces? It is well noted that multiplayer spaces, particularly those hosted online, can cease to be playable; as was the case with Second Life's Okapi Island (Tringham and Ashley, 2015). Other recent example have begun to emerge, for instance, Blizzard's once monolithic MMORPG, World of Warcraft. Closure of servers — particularly, fan-hosted servers of the game in an earlier iteration — represent an erasure of over a decade-long culture (Johnson, 2016). Tools similar in nature to those discussed in the present work [18] could be used to keep visible important histories of videogame communities and practice.

Through data analytics platforms, a workable, reliable and rigorous solution is presented.

In addition to the obvious benefits an alignment toward the past poses to formulating longitudinal studies of play, the ability to examine digital traces has applicability for the study of the *culture* of play. The ability to draw on an archive of text based, ingame correspondence could prove invaluable in exploring behaviour in multiplayer space, in addition to using play as a marker of behaviour. Potentially, researchers could map out shifts in community values and attitudes. Multiplayer games, Dota 2 especially, could be quite a productive testbed here; exploring recent formulations of 'toxic technocultures', emerging from writing on Gamergate, for instance (Massanari, 2015). Recent work has noted the Dota 2 community, in particular, is particularly caustic (Osterweil, 2016). Researchers could apply quantitative content analysis approaches, or perhaps implement more technically sophisticated methods of lexical analysis; such as latent semantic analysis.

The temporal alignment of archived data could prove pivotal in addressing issues of game play conservation, while aiding in the construction of both longitudinal (and historical) accounts of videogame culture; specifically, the important, yet fleeting cultures of networked game space.




9. Conclusion

Multiplayer videogames are an important part of the contemporary media landscape. Host to playerbases as huge as they are diverse, player data analytics are a potent apparatus for researching not only play, but game cultures, experiences and practices of both past and present. Through quantitative, and large-scale fieldwork, researchers of videogames in humanities disciplines such as media and cultural studies, can expand on an array of key research interests. As I've noted, sociocultural relations could profitably be mapped using quantitative game-text data; charting contestations between gaming stakeholders. Further, a

window into the increasingly diverse environments of contemporary gaming are afforded (even if limited in some ways). Popular media studies approaches could also be enhanced, such as media-technologies-effects theories. For instance, media studies' evergreen interest in techno-corporeality can be quantitatively developed via analytics research; with metrics existing to measure relations between hardware, software and the somatic rhythms of the human player. Further, media-theoretical accounts of experience and perception can be developed (as pointed out by the brief visitation of e-sports data).

Beyond providing an account of ways in which data analytics in multiplayer games can function and enhance game studies, using the example of Dota 2, interrogations of the game studies research climate were presented, alongside numerous ontologies of videogames, and the ways in which they have seemingly inhibited the wider uptake of player data analytics.

All things considered, the idea of appropriating data analytics for purposes of academic games research (principally in the humanities) is still one relatively loosely sketched out. However, the provocation made in this paper is still key; these approaches have the potential to enhance existing viewpoints and generate new perspectives not necessarily made visible through established or popular means.

This provocation is one backed up by an emerging field of work on game analytics within computationally-oriented games literature, as well as a fairly sizeable body of research situated around user analytics of other digital media. It is expected that future applications in game studies will yield rich and novel insights into games and their cultural dimensions. 

About the author

Ben Egliston is a doctoral candidate in the Digital Cultures Program at the University of Sydney, New South Wales. His research interests are broadly situated around videogames. He is currently occupied with examining how modalities of broadcasting and spectating videogame play have come to temper the ways in which players approach, experience and perceive videogames.
E-mail: benegliston [at] gmail [dot] com

Notes

1. While I feel that Dota 2, as an object of study, is fairly interchangeable, and peripheral to the suggestions and provocations being made in this paper, I will briefly outline the game. Dota 2 is a MOBA (Multiplayer Online Battle Arena) game for the PC. The game is adversarial in nature, and involves two teams of five a side attempting to destroy the others' base: a large structure surrounded by powerful armaments. Players interact and engage in combat with others through characters called 'heroes'. There are, at the time of writing, 111 unique heroes in the game, each with a relatively distinct mechanical flavour.

2. Genette, 1997, p. xviii.

3. Consalvo, 2007, p. 22.

4. Apperley, 2010, p. 125.

5. To note, analyses of communicative trends in multiplayer environments have been undertaken in the past. The work of Herring, *et al.* (2009) explored textual communication by looking at chat logs.

6. boyd and Crawford, 2012, pp. 12–13.

7. Taylor, 2003, p. 67.

8. See boyd and Crawford, 2012, p. 10.

9. A number of developers took to the game's official Reddit forum, see https://www.reddit.com/r/DotA2/comments/4l5qo6/dota2_live_api_issue_that_maybe_most_of_you_are/.

10. Crawford, 2009, p. 532.

11. boyd and Crawford, 2012, p. 669.

12. waterandshade, 2015, noting YASP essentially scraping Dotabuff's data instead of generating own data using API.

13. As studies looking at schisms between social sciences and information science research suggests, see Zelenkauskaitė and Bucy, 2016.

14. Grossberg, *et al.*, 1992, p. 2.

15. Grossberg, *et al.*, 1992, p. 15.

16. Aarseth, 2007, p. 6.

17. Apperley and Jayemanne, 2012, p. 10.

18. For instance, World of Logs or WarcraftLogs in World of Warcraft, offering similar utility to Dotabuff.

References

Espen Aarseth, 2007. "Playing research: Methodological approaches to game analysis," *Artnodes*, number 7, at <http://www.uoc.edu/artnodes/7/dt/eng/aarseth.html>, accessed 4 January 2016.

Anna Anthropy, 2012. *Rise of the videogame zinesters: How freaks, normals, amateurs, artists, dreamers, drop-outs, queers, housewives and people like you are taking back an art form*. New York: Seven Stories Press.

Thomas Apperley, 2010. *Gaming rhythms: Play and counterplay from the situated to the global*. Amsterdam: Institute of Network Cultures, at <http://networkcultures.org/blog/publication/no-06-gaming-rhythms-play-and-counterplay-from-the-situated-to-the-global-tom-apperley/>, accessed 21 June 2016.

Thomas Apperley and Jussi Parikka, 2015. "Platform studies' epistemic threshold," *Games and Culture*. doi: <http://dx.doi.org/10.1177/1555412015616509>, accessed 21 June 2016.

Thomas Apperley and Darshana Jayemane, 2012. "Game studies' material turn," *Westminster Papers in Communications and Culture*, at <http://www.westminsterpapers.org/articles/abstract/10.16997/wpcc.145/>, accessed 11 June 2016.

James Ash, 2013. "Technologies of captivation: Videogames and the attunement of affect," *Body & Society*, volume 19, number 1, pp. 27–51. doi: <http://dx.doi.org/10.1177/1357034X11411737>, accessed 28 March 2016.

Christian Bauckhage, Rafet Sifa, Anders Drachen, Christian Thureau and Fabian Hadiji, 2014. "Beyond heatmaps: Spatio-temporal clustering using behavior based partitioning of game levels," *2014 IEEE Conference on Computational Intelligence and Games*, pp. 1–8. doi: <http://dx.doi.org/10.1109/CIG.2014.6932865>, accessed 13 June 2016.

Richard Bartle, 1996. "Hearts, clubs, diamonds, spades: Players who suit MUDS," at <http://mud.co.uk/richard/hcds.htm>, accessed 11 June 2016.

Tom Boellstorff, Bonnie Nardi, Celia Pearce and T.L Taylor (editors), 2012, *Ethnography and virtual worlds: A handbook of method*. Princeton, N.J.: Princeton University Press.

danah boyd and Kate Crawford, 2012. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, Communication & Society*, volume 15, number 5, pp. 662–679. doi: <http://dx.doi.org/10.1080/1369118X.2012.678878>, accessed 28 March 2015.

Axel Bruns, 2013. "Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship," *First Monday*, volume 18, number 10, at <http://firstmonday.org/article/view/4879/3756>, accessed 1 October 2015. doi: <http://dx.doi.org/10.5210/fm.v18i10.4879>, accessed 1 October 2015.

Axel Bruns and Yuxian Eugene Liang, 2012. "Tools and methods for capturing Twitter data during natural disasters," *First Monday*, volume 17, number 4, at <http://firstmonday.org/article/view/3937/3193>, accessed 1 October 2015. doi: <http://dx.doi.org/10.5210/fm.v17i4.3937>, accessed 1 October 2015.

Marcus Carter, 2014. "Emitexts and paratexts: Propaganda in *EVE Online*," *Games and Culture*, volume 10, number 4, pp. 311–342. doi: <http://dx.doi.org/10.1177/1555412014558089>, accessed 17 October 2015.

Dean Chan, 2008. "Convergence, connectivity, and the case of Japanese mobile gaming," *Games and Culture*, volume 3, number 1, pp. 13–25. doi: <http://dx.doi.org/10.1177/1555412007309524>, accessed 2 January 2016.

Mia Consalvo, 2007. *Cheating: Gaining advantage in videogames*. Cambridge, Mass.: MIT Press.

Kate Crawford, 2009. "Following you: Disciplines of listening in social media," *Continuum*, volume 23, number 4, pp. 525–535. doi: <http://dx.doi.org/10.1080/10304310903003270>, accessed 29 September 2015.

Scott Donaldson, 2015. "Mechanics and metagame: Exploring binary expertise in *League of Legends*," *Games and Culture*. doi: <http://dx.doi.org/10.1177/1555412015590063>, accessed 18 September 2015.

Dotabuff, 2016. "Dotabuff," at <http://www.dotabuff.com>, accessed 18 September 2015.

Datdota, 2016. "Datdota," at <http://www.datdota.com>, accessed 18 September 2015.

Phillip Derosa, 2007. "Tracking player feedback to improve game design" (7 August), at http://www.gamasutra.com/view/feature/129969/tracking_player_feedback_to_.php, accessed 2 January

2016.

Anders Drachen, Matthew Yancey, John Maguire, Derrek Chu, Iris Yuhui Wang, Tobias Mahlmann, Matthias Schubert and Diego Klabajan, 2014. "Skill-based differences in spatio-temporal team behaviour in Defence of The Ancients 2 (DotA 2)," *Proceedings of 2014 IEEE Games, Media, Entertainment Conference*, pp. 1–8. doi: <http://dx.doi.org/10.1109/GEM.2014.7048109>, accessed 1 April 2016.

Ben Egliston, 2015a. "Playing across media: Exploring transtextuality in competitive games and esports," *DiGRA '15: Proceedings of the 2015 DiGRA International Conference*, at http://www.digra.org/wp-content/uploads/digital-library/122_Egliston_Playing-Across-Media.pdf, accessed 11 April 2016.

Ben Egliston, 2015b, "Playstyle and place: On the territorial identity of tactics in Dota 2," paper presented at DiGRA Australia 2015: Inclusivity in game studies (Sydney, 29–30 June).

Simon Ferrari, 2013. "From generative to conventional play: MOBA and *League of Legends*," *DiGRA '13: Proceedings of the 2013 International Conference*, at http://www.digra.org/wp-content/uploads/digital-library/paper_230_formattingfixed.pdf, accessed 9 January 2016.

John Funk, 2013. "MOBA, DOTA, ARTS: A brief introduction to gaming's biggest, most impenetrable genre," *Polygon* (2 September), at <http://www.polygon.com/2013/9/2/4672920/moba-dota-arts-a-brief-introduction-to-gamings-biggest-most>, accessed 9 January 2016.

Gérard Genette, 1997. *Paratexts: Thresholds of interpretation*. Translated by Jane Lewin. Cambridge: Cambridge University Press.

Seth Giddings and Helen Kennedy, 2008. "Little Jesuses and fuck-off robots: On aesthetics, cybernetics, and not being very good at Lego Star Wars," In: Melanie Swalwell and Jason Wilson (editors). *The pleasures of computer gaming: Essays on cultural history, theory and aesthetics*. Jefferson, N.C.: McFarland, pp. 13–32.

Dan Golding, 2014. "The end of gamers" (28 August), at <http://dangolding.tumblr.com/post/95985875943/the-end-of-gamers>, accessed 2 January 2016.

Owen Good, 2015. "Dota 2 is Steam's first game with 1 million users playing at the same time," *Polygon* (15 February), at <http://www.polygon.com/2015/2/15/8042171/dota-2-concurrent-users-million-steam-pc>, accessed 2 February 2016.

Catherine Goodfellow, 2015. "Russian overlords: vodka and logoffski: Russian-and English- language discourse about anti-Russian xenophobia in the *EVE Online* community," *Games and Culture*, volume 10, number 4, pp. 343–364. doi: <http://dx.doi.org/10.1177/1555412014560193>, accessed 1 March 2016.

Lawrence Grossberg, Cary Nelson and Paula Treichler (editors), 1992. *Cultural studies*. New York: Routledge.

Martin Heidegger, 1962. *Being and time*. London: SCM Press.

Larissa Hjorth, 2010. "The game of being social: Web 2.0, social media and online games," *Iowa Journal of Communication*, volume 42, number 1, pp. 73–92.

Susan Herring, Daniel Kutz, John Paolilo and Asta Zelenkauskaitė, 2009. "Fast talking, fast shooting: Text chat in an online first-person game," *HICSS '09: Proceedings of the 42nd Hawaii International Conference on System Sciences*, pp. 1–10. doi: <http://dx.doi.org/10.1109/HICSS.2009.215>, accessed 15 April 2016.

Brett Hutchins, 2015. "Tales of the digital sublime: Tracing the relationship between big data and professional sport," *Convergence*. doi: <http://dx.doi.org/10.1177/1354856515587163>, accessed 14 March 2016.

Harold Innes, 1950. *Empire and communications*. Oxford: Clarendon Press.

Henry Jenkins, 1992. *Textual poachers: Television fans & participatory culture*. New York: Routledge.

Leif Johnson, 2016. "Blizzard erases gaming history by axing a fan-made 'World of Warcraft' server," *Motherboard* (10 April), at <http://motherboard.vice.com/read/world-of-warcraft-legacy-server-shut-down>, accessed 15 April 2016.

Brendan Keogh, 2014. "Across worlds and bodies: Criticism in the age of video games," *Journal of Games Criticism*, volume 1, number 1, at <http://gamescriticism.org/articles/keogh-1-1/>, accessed 10 January 2016.

Graeme Kirkpatrick, 2012. "Constitutive tensions in gaming's field: UK gaming magazines and the formation of gaming culture 1981–1995," *Game Studies*, volume 12, number 1, at <http://gamestudies.org/1201/articles/kirkpatrick>, accessed 14 January 2016.

Graeme Kirkpatrick, 2009. "Controller, hand, screen: Aesthetic form in the computer game," *Games and Culture*, volume 4, number 2, pp. 127–143. doi: <http://dx.doi.org/10.1177/1555412008325484>, accessed 14 March 2015.

Rob Kitchin, 2014. "Big data, new epistemologies and paradigm shifts," *Big Data & Society*, volume 1, number 1.

doi: <http://dx.doi.org/10.1177/1555412008325484>, accessed 10 January 2016.

Friedrich Kittler, 1990. "The mechanized philosopher," In: Laurence Rickels (editor). *Looking after Nietzsche*. Albany: State University of New York Press, pp. 195–207.

Chong-U Lim and D. Fox Harrell, 2015. "Toward telemetry-driven analytics for understanding players and their avatars in videogames," *CHI EA '15: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1,175–1,180.
doi: <http://dx.doi.org/10.1145/2702613.2732783>, accessed 20 March 2016.

Stine Lomborg and Anja Bechmann, 2014. "Using APIs for data collection on social media," *Information Society*, volume 30, number 4, pp. 256–265.
doi: <http://dx.doi.org/10.1080/01972243.2014.915276>, accessed 30 March 2016.

Lev Manovich, 2012. "Trending: The promises and the challenges of big social data," In: Matthew Gold (editor). *Debates in the digital humanities*. Minneapolis: University of Minnesota Press, pp. 460–475.
doi: <http://dx.doi.org/10.5749/minnesota/9780816677948.003.0047>, accessed 20 March 2016.

Annette Markham and Elizabeth Buchanan, 2012. "Ethical decision-making and Internet Research: Recommendations from the AoIR Ethics Working Committee (version 2.0)," at <http://aoir.org/reports/ethics2.pdf>, accessed 11 August 2015.

Adrienne Massanari, 2015. "#Gamergate and the fapping: How Reddit's algorithm, governance and culture support toxic technocultures," *New Media & Society*.
doi: <http://dx.doi.org/10.1177/1461444815608807>, accessed 20 March 2016.

Marshall McLuhan, 1974. *Understanding media: The extensions of man*. London: Abacus.

Maurice Merleau-Ponty, 2002. *Phenomenology of perception*. Translated by Colin Smith. London: Routledge.

Nick Montfort and Ian Bogost, 2009. *Racing the beam: The Atari Video computer system*. Cambridge, Mass.: MIT Press.

Willie Osterweil, 2016. "Discipline and pleasure," *New Inquiry*, at <http://thenewinquiry.com/essays/discipline-and-pleasure/>, accessed 1 June 2016.

Stuart Reeves, Barry Brown and Eric Laurier, 2015. "Experts at play: Understanding skilled expertise," *Games and Culture*, volume 4, number 3, pp. 205–227.
doi: <http://dx.doi.org/10.1177/1555412009339730>, accessed 10 January 2016.

Magy Seif El-Nasr, Anders Drachen and Alessandro Canossa (editors), 2013. *Game analytics: Maximizing the value of player data*. London: Springer-Verlag.
doi: <http://dx.doi.org/10.1007/978-1-4471-4769-5>, accessed 21 June 2016.

Jane Singer, 2014. "User-generated Visibility: Secondary gatekeeping in a shared media space," *New Media & Society*, number 16, volume 1, pp. 55–73.
doi: <http://dx.doi.org/10.1177/1461444813477833>, accessed 21 June 2016.

Adrienne Shaw, 2011. "Do you identify as a gamer? Gender, race, sexuality, and gamer identity," *New Media & Society*, volume 14, number 1, pp. 28–44.
doi: <http://dx.doi.org/10.1177/1461444811410394>, accessed 14 March 2015.

David Sudnow, 1983. *Pilgrim in the microworld*. New York: Warner Books.

Melanie Swalwell, 2013. "Moving on from the Original Experience: Games history, preservation and presentation," *DiGRA '13: Proceedings of the 2013 DiGRA International Conference*, at http://www.digra.org/wp-content/uploads/digital-library/paper_454.pdf, accessed 6 January 2016.

Steve Swink, 2009. *Game feel: A game designer's guide to virtual sensation*. Boston: Morgan Kaufmann.

T.L. Taylor, 2012. *Raising the stakes: E-sports and the professionalization of computer gaming*. Cambridge, Mass.: MIT Press.

T.L. Taylor, 2009. "The assemblage of play," *Games and Culture*, volume 4, number 4, pp. 331–339.
doi: <http://dx.doi.org/10.1177/1555412009343576>, accessed 26 December 2015.

T.L. Taylor, 2003. "Power gamers just want to have fun: Instrumental play in a MMOG," *DiGRA '03: Proceedings of the 2003 DiGRA International Conference*, at <http://www.digra.org/digital-library/publications/power-games-just-want-to-have-fun-instrumental-play-in-a-mmog/>, accessed 21 June 2016.

Ruth Tringham and Michael Ashley, 2015, "Becoming archaeological," *Journal of Contemporary Archaeology*, volume 2, number 1, at <https://journals.equinoxpub.com/index.php/JCA/article/view/27089>, accessed 29 December 2015.
doi: <http://dx.doi.org/10.1558/jca.v2i1.27089>, accessed 29 December 2015.

Valve, 2015. "Dota 2," at <http://store.steampowered.com/app/570/>, accessed 18 September 2015.

waterandshade 2015, "Introducing YASP: A free, open-source stats Website with rfeplay parsing," at http://www.reddit.com/r/DotA2/comments/2sp595/introducing_yasp_a_free_opensource_stats_website/, accessed 11 January 2016.

Dmitri Williams, Nick Yee and Scott Caplan, 2008. "Who plays, how much, and why? Debunking the stereotypical gamer profile," *Journal of Computer-Mediated Communication*, volume 13, number 4, pp. 993–1,018.

doi: <http://dx.doi.org/10.1111/j.1083-6101.2008.00428.x>, accessed 11 June 2016.

Asta Zelenkauskaite and Eric Bucy, 2016. "A scholarly divide: Social media, Big Data, and unattainable scholarship," *First Monday*, volume 21, number 5, at <http://firstmonday.org/article/view/6358/5511>, accessed 11 June 2016.

doi: <http://dx.doi.org/10.5210/fm.v21i5.6358>, accessed 11 June 2016.

Appendices

datdota The International 2015

Hero	Name	Game s	Levels PM	Kills PM	Deaths PM	Assists PM	LH PM	Denies PM	Dmg PM	Tower Dmg PM	Healing PM
anti mag e	Anti Mage	25	0.54	0.16	0.08	0.11	10.4	0.48	250	92.6	2.5
spec tre	Spectre	1	0.35	0.23	0.11	0.35	8.6	0.35	412	57.6	0.0
neve rmor e	Shadow Fiend	47	0.50	0.17	0.11	0.18	8.2	0.36	323	72.2	22.2
luna	Luna	9	0.51	0.12	0.08	0.24	7.7	0.49	251	84.3	0.0
gyro copt er	Gyrocopter	91	0.49	0.19	0.10	0.24	7.4	0.40	374	69.4	0.0
morp hling	Morphling	1	0.55	0.32	0.05	0.12	7.2	0.97	405	207.8	0.0
phan tom_ lanc er	Phantom Lancer	26	0.51	0.14	0.06	0.17	7.2	0.39	241	64.7	0.3

Appendix 1: Tabulation of player Actions Per Minute, from matches of a LAN tournament in 2015.
Source: datdota.com.

Appendix 2: Tabulation of API calls and functions.		
String	Function	Potential indexical marker or application
hero_id=<id>	Search for matches with a specific hero (Character) being played	Playstyle; how people play the game
game_mode=<mode>	Search for matches of a given mode (see below)	Game mode (e.g., Ranked mode, serious play versus Normal mode)

		casual play; could be applied as quantitative marker of 'serious play'
skill=<skill>	0 for any, 1 for normal, 2 for high, 3 for very high skill	Marks player ability, used to stratify data, player ability
tournament_games_only= <string>	Show tournament (professional) only games	Can be used as a marker of player skill; professional versus amateur level play
start_at_match_id=<id>	Start the search at the indicated match ID, descending	Search all games from a designated start point

Editorial history

Received 21 April 2016; revised 13 June 2016; revised 21 June 2016; accepted 22 June 2016.



This paper is in the [Public Domain](#).

Big playerbase, big data: On data analytics methodologies and their applicability to studying multiplayer games and culture
by Ben Egliston.

First Monday, Volume 21, Number 7 - 4 July 2016

<http://journals.uic.edu/ojs/index.php/fm/rt/prINTERfriendly/6718/5531>

doi: <http://dx.doi.org/10.5210/fm.v21i7.6718>