Data-driven Design: A Case for Maximalist Game Design
Paper type: Position paper

Gabriella A. B. Barros[1]**, Michael Cerny Green**[1]**, Antonios Liapis**[2] **and Julian Togelius**[1]

1: Tandon School of Engineering, New York University, New York, USA
2: Institute of Digital Games, University of Malta, Msida, Malta
gabbbarros@gmail.com, mcgreentn@gmail.com, antonios.liapis@um.edu.mt, julian@togelius.com

Abstract

Maximalism in art refers to drawing on and combining multiple different sources for art creation, embracing the resulting collisions and heterogeneity. This paper discusses the use of maximalism in game design and particularly in data games, which are games that are generated partly based on open data. Using Data Adventures, a series of generators that create adventure games from data sources such as Wikipedia and Open-StreetMap, as a lens we explore several tradeoffs and issues in maximalist game design. This includes the tension between transformation and fidelity, between decorative and functional content, and legal and ethical issues resulting from this type of generativity. This paper sketches out the design space of maximalist data-driven games, a design space that is mostly unexplored.

Introduction

The unprecedented availability of digital data impacts most human endeavors, including game design. In particular, freely available data can be combined with procedural content generation (PCG) and computation creativity to create systems that can generate games (or game content) based on open data. We have previously identified such games as "data games" (Gustafsson Friberger et al. 2013).

This paper explores some of the aesthetic challenges, particularities and concerns associated with games that are created from data. We start from the idea that the use of data games is in many ways similar to notions in art such as collage, sampling, and remixing. We draw on content from many different sources, causing creative collisions between them. This lets us apply some of the same conceptual apparatus to study data games as has been applied to these types of art. We also start from a series of game generators we have created, collectively referred to as "Data Adventures". These generators create adventure games, such as murder mysteries, from open data from e.g. Wikipedia and Open-StreetMap. Our ongoing struggle with getting these generators to produce playable and interesting content from something as varied and occasionally unreliable as Wikipedia has illuminated both possibilities and pitfalls of this approach.

This paper is an attempt to explore the design space of maximalist data-driven games (and other data games) in order to form an initial understanding of it. It is also an attempt to systematize reflections from our own and others' attempts at creating such games. We address the following questions:

- What does it mean for games designed from/for data to be maximalist?
- What is the tradeoff between transforming data and staying true to the source in terms of generating games?
- What are the characteristics of game content that can be generated from data?
- For what purposes can data-driven maximalist games be designed and how does that affect their character?
- What new legal and ethical issues, including copyright issues and the potential for generating offensive, misinforming and biased content, are raised by this type of game design?

Data-driven design and data games

This age of data sharing (whether sharing is free or not) has certainly been advantageous to research in computational creativity. While computational creativity does not necessarily need to emulate human creativity (Pease and Colton 2011), freely available human-annotated data can be exploited as an inspiring set (Ritchie 2007) to any creative software. In natural language generation, Google N-grams have been exploited to identify analogies and similes (Veale 2014), corpora of phonetic information for all words have been exploited to generate jokes (Ritchie and Masthoff 2011), and books of a specific author have been used to generate stories typical of the genre (Khalifa, Barros, and Togelius 2017). In visual generation, crowd-sourced annotations of data were used to create image filters (Heath and Ventura 2016), while object recognition models based on deep learning of Google images was used to choose how generated 3D shapes would represent an object (Lehman, Risi, and Clune 2016). Similarly, deep learning from massive musical corpora was used to create new music (Hawthorne et al. 2017).

In the creative domain of games, on the other hand, similar approaches have been used to create different game components. Google's autocomplete function (which uses a form of N-grams) was used to discover names for enemies and abilities of a game character whose name was provided by the player (Cook and Colton 2014). In the same game,

Google image search used discovered names to select images for these enemies' sprites. In other work, Guzdial and Riedl (2016) used Youtube playthroughs to find associations in the placement of level elements (e.g. platforms, enemies) which were used to generate levels for *Super Mario Bros* (Nintendo 1985). Patterns in *Starcraft II* maps (Blizzard Entertainment 2010) were learned through deep learning (Lee et al. 2016); these encodings were used to change the frequency of minerals in the map without the usual exploratory process of e.g. an evolutionary algorithm. To better coordinate the learning process of level patterns, a corpus of diverse games has been collected (Summerville et al. 2016).

While using existing game data —often annotated with human notions of quality— has been explored in computational game creativity (Liapis, Yannakakis, and Togelius 2014), most efforts perform minor adjustments to existing games. Game generators such as Angelina (Cook, Colton, and Pease 2012), A Rogue Dream (Cook and Colton 2014), and Game-O-Matic (Treanor et al. 2012) use data outside the game domain, enhancing their outcomes with human-provided associations (and content such as images). Even so, the core gameplay loop is simple: in Angelina, for example, the player performs the basic actions of a platformer game (e.g. jump, run); in A Rogue Dream the player moves along 4 directions and perhaps uses one more action. Gameplay in all these games is mechanics-heavy, relying on fast reactions to immediate threats rather than on high-level planning or cognitive ability. Many *data games* take an existing game mechanic and generate new content for that game from open data (Gustafsson Friberger et al. 2013; Gustafsson Friberger and Togelius 2013; Cardona et al. 2014). In some cases, such as the game *Bar Chart Ball*, a new game mechanic is added to an existing data visualization (Togelius and Gustafsson Friberger 2013). To play even simple data games, the player must have some understanding of the underlying data. Playing data games requires some mental effort, deduction or memory; not only dexterity.

While most data-driven game generation software focus on a simple and tight gameplay loop, there is considerable potential in using and re-using information outside of games to create more complex game systems and more involved experiences. We argue that data-driven game generation can allow for a new gameplay experience. Using the Data Adventures series of game generators as a concrete example, we articulate the tenets of maximalism in game design inspired by the art movement of the same name. Moreover, we discuss two possible dimensions of maximalist game design, and how it can start from the raw data on one end or from the gameplay experience on the other. Finally, we envision the potential uses and issues of maximalist game design.

Maximalism in data-driven design

We are inspired by the notion of maximalism in the arts, rather than in the game design sphere. In music, for example, maximalism "embraces heterogeneity and allows for complex systems of juxtapositions and collisions, in which all outside influences are viewed as potential raw material" (Jaffe 1995). We similarly embrace the use of heterogeneous data sources as notes (i.e. the individual components) and melody (i.e. the overarching game or narrative structure) to produce a game as an orchestration of dissimilar instruments (Liapis 2015). In that sense, maximalism in data-driven design is likened with mixed media in art, where more than one medium is used. De facto, the heterogeneity of the data, its sources, and the people who contribute to its creation and curation will insert juxtapositions and collisions. This may not always be desired, and several catastrophic, inconsequential or seemingly random associations should be redacted. However, the "grain" of data-driven design (Khaled, Nelson, and Barr 2013) is built on the collision and absurdity of different elements that find their way into the game.

It should be noted that maximalism in the artistic sphere refers to materials or identities of elements within an image, song, or novel. We refer to maximalist game design in that sense, focusing on how game elements originating from different data sources (or transformed in different ways) are visualized, combined and made to interact together, thus not directly opposed to minimalist game design. Nealen, Saltsman, and Boxerman's (2011) minimalist game design encourages removing the unnecessary parts of the design, highlighting the important bits. Sicart's approach (2015) refers to the game loop; minimalist games have a simple core game loop which is largely unchanged throughout the game. Sicart uses *Minecraft* (Mojang, 2011) as an example where the simple core loop gather→craft→build that remains relevant and unchanged (except from the specific materials worked) throughout the game.

A data-driven game, maximalist in the artistic sense, can also be minimal in the gameplay loop sense. Data Adventures (Barros, Liapis, and Togelius 2015) has a simple core gameplay loop of traveling to a new location, talking to a non-player character in that location, learning the clue for the next location. Games that we would define as maximalist on the design sense, on the other hand, have the broadest mechanics of options for solving a problem — e.g. killing a dragon with stealth, magic, followers, swords, fists, poison, etc. in *Skyrim* (Bethesda, 2011) — or subsystems that are so elaborate or numerous that the player becomes unable to distinguish a core game loop — e.g. the diverse driving, shooting, spraying, running, etc. minigames in *Saints Row IV* (Deep Silver, 2013) which are the main ways to progress in the game. While certainly data-driven design can offer the latter form of maximalism, e.g. with individual minigames where different forms or sources or data are presented and interacted with in each, not all data-driven games need to have maximalist game loops.

Case Study: Data Adventure Games

The Data Adventures series of game generators exemplify the use of a high volume of data to procedurally generate content (Barros, Liapis, and Togelius 2015). The generated adventure games use information gathered from Wikipedia, DBpedia, Wikimedia Commons and OpenStreetMap (OSM) to automatically create an adventure, complete with plot, characters, items and in-game locations. The series consists of three games: Data Adventures (Barros, Liapis, and Togelius 2015; 2016b), WikiMystery (Barros, Liapis, and To-

(a) Data Adventures map screen (b) Data Adventures NPC screen

(c) WikiMystery location screen (d) WikiMystery accusation screen

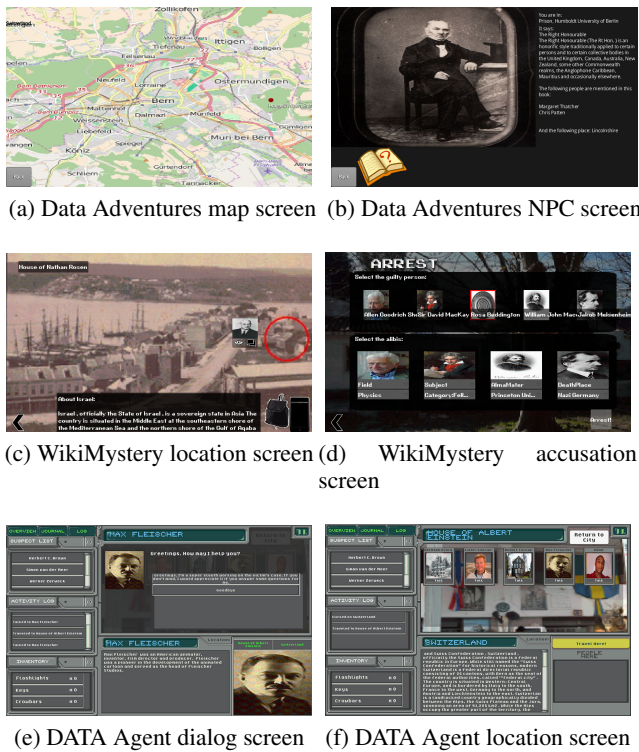(e) DATA Agent dialog screen (f) DATA Agent location screen

Figure 1: Screenshots from the different games in the Data Adventures series. Sources: (Barros, Liapis, and Togelius 2016a; 2016b).

gelius 2016a; Barros et al. 2018b) and DATA Agent (Barros et al. 2018a). Each evolved from the previous one, with DATA Agent being the most recent, complex and powerful. Most of the gameplay, however, is the same: a point-and-click interface inspired by "Where In The World is Carmen Sandiego?" (Brøderbund Software 1985).

The series' first installment is Data Adventures, an exploration game created from the connections between two Wikipedia articles about specific people. Two Non-Playable Characters (NPCs) are generated representing each of these people. The player receives a quest from the first NPC, asking them to find the second one. To do so, the player has to travel through cities, talking to other generated NPCs and reading books. All information is created from a path linking one article (of the starting NPC) to the other article (of the goal NPC). Figures 1a and 1b show a map screen generated using OSM and a location showing a NPC and a book.

The second game, WikiMystery, plays differently from Data Adventures. On one hand, the game has an arguably more interesting plot, where the player is a detective trying to solve a murder. Additionally, it is generated using only one input: the victim's name. The system finds people related to the victim, forming a pool of possible suspects, and evolves a small list of suspects that are somehow related to each other. It also provides evidence of innocence to any suspect that is, as the name implies, innocent. The player's goal is to find the one suspect which has no evidence of in-

nocence, and arrest him or her. It thus requires that all four pieces of evidence (one per innocent NPC) are collected before the game can be completed. Figures 1c and 1d show a location screen and the accusation screen, where the player identifies the culprit and provides evidences of innocence.

In DATA Agent, the player acts as a time-traveler in charge of finding a murder suspect, who went back in time and killed an important person. The game provides a list of suspects, and the player must travel through locations and uncover clues by talking to NPCs or interacting with items, in order to identify which among the suspects is the culprit. Similar to WkiMystery, DATA Agent's generator is capable of creating a full adventure when given a real person's name. This person becomes the center of the story, as the victim of a murder. Using artificial intelligence techniques over Wikipedia and DBpedia content, the system finds articles related to the person's article, and fleshes out links between suspects and the victim. Every in-game NPC, object, location, dialog or image is created from real information. Unlike WikiMystery, there is no evidence of innocence. The game finishes when an a suspect NPC is interrogated by the player and answers wrongly on personal information; the player must have collected the real information during gameplay. NPCs in the game have a much more involved dialog system, and can give information about suspects or about themselves, such as their birth day and occupation, or the reason they were chosen as suspects by the system. Figures 1e and 1f show a dialog screen and an in-game location.

Designing Games for Maximalism

A major challenge of maximalist game design is deciding what to prioritize. One can shape data in order to fit the game, or modify the game design to better showcase the original data. One can also have data ingrained in the game mechanics, directly affecting gameplay, or show the data in a decorative manner. Maintaining a balance between data transformation to fit other data and the game itself, or staying faithful to the original data while providing an engaging experience is challenging. This section describes two dimensions of maximalist game design: *Data Transformation versus Data Fidelity* and *Functionality versus Decoration*.

Data Transformation versus Data Fidelity

The tension between data fidelity and data transformation is rooted in the priorities of a maximalist designer: the original game design or the original data. When using open data, designers may wish to adapt that data to the game, or to keep the data as it is and mold the game around it. Extensive data transformations may improve the game experience, but are also susceptible to loss of information or inaccuracies.

Transforming data gives designers more freedom and might be preferred if they have an inflexible idea, or if the data itself is malleable. DATA Agent is such an example of data transformation. In the game, the engine transforms individual facts about separate people into a murder mystery. The facts are also transformed into dialog lines, used by NPCs when prompted by the player. Some facts are altered purposely, in order to "lie": the culprit's dialog differs

from reality in order to point to the time-agent (and thus, the player) that he or she is guilty. WikiMystery, on the other hand, uses proof of innocence in a similar manner but never misrepresents the actual data: all proofs given by NPCs are true, and the player must memorize them in order to use them in the game's accusation sequence.

On the other hand, designers may instead wish to stay faithful to the original data, molding the game to the data instead. This way, information present in the data is more likely to be clearly presented within game content. The rigidity to data restricts what kind of game elements can be used, or forces designers to be creative in their implementations. While less time might be spent cleaning and translating data, more time will likely be spent on raw game and mechanic design. An example can be found in Data Adventures, where data instantiated in the game is sourced from OSM and Wikipedia articles about people, places, and concepts. The designers built a game that could involve all four of those elements: a game where the player travels around the world searching for links to the goal NPC. It introduces some alterations from the original material, but most of it remains unchanged in the game. However, the game lacks a convincing narrative and theme, such as a murder mystery in later installments of the data adventures series. Another example is WikiRace[1], where the game uses Wikipedia to navigate the game.

Functional versus Decorative

Another dimension pertaining to maximalist game design is functionality versus decoration. We define data being *functional* when it has a strong impact on gameplay. If the player does not have to interact with the data, or the data does not impact gameplay in a significant way, then it is *decorative*.

In order to be functional, data can be incorporated in a variety of ways. In DATA Agent, dialog and character names heavily rely on open data. To progress in the game, you must interact with these characters and talk to them. The data is functional, as remembering which NPC (by name) has a certain fact is necessary to identify the culprit. In OpenTrumps (Cardona et al. 2014), the raw mechanics of the game come from open data, as the cards themselves are created from it. The maps in *FreeCiv* generated by Barros and Togelius (2015) are based on real-world terrain data, and impact gameplay as terrain affects players' city production.

On the other hand, any data that does not serve a functional purpose is decorative. Data can serve a decorative purpose in many ways. A Rogue Dream (Cook and Colton 2014) uses open data to name player abilities, however the in-game effects of the abilities are not affected by their names or the underlying data. In DATA Agent, city maps and NPC profile images are used as visual stimuli and play no mechanical role in game. *World of Warcraft* (Blizzard, 2004) uses real-world time to create an aesthetic day-night cycle in-game, which has no affect on actual gameplay.
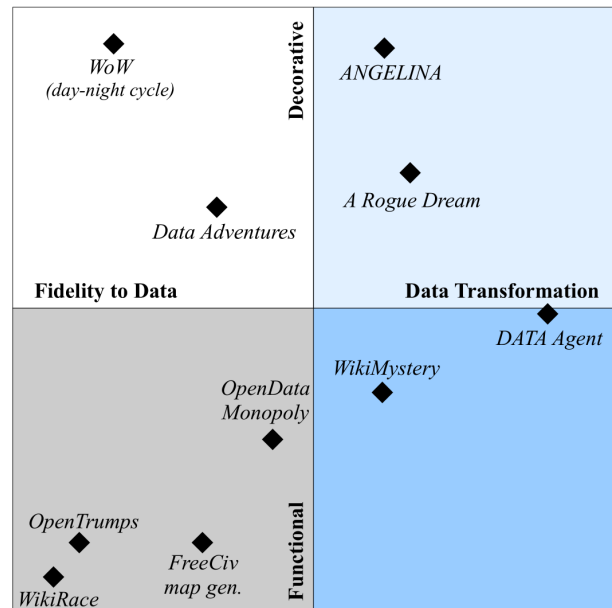
---
[1]http://2pages.net/wikirace.php

Figure 2: Examples of games within the two dimensions.

Instances of Data-driven Design

Figure 2 shows the two dimensions described above. The X-axis represents Data Transformation versus Data Fidelity, while the Y-axis represents Functional versus Decorative. Games where the goal is to preserve the original data, adapting the game to do so, are on the leftmost side of the figure: WikiRace and OpenTrumps exemplify this. Data in these games is also extremely functional: all mechanics in these games rely on a direct interaction with and understanding of the data (in WikiRace through reading the articles, and OpenTrumps through the values that affect deck superiority). Less faithful to the data, but similarly functional, is the FreeCiv map generation where geographic data is used as-is; resource placement is based on the original data but also adapted (i.e. transformed) for playability.

Moving upwards along the Y-axis, we find Open Data Monopoly (Gustafsson Friberger and Togelius 2013) and WikiMystery. Both use data in a decorative manner, the former as names for lots on the board and the latter as images, but some of the original data is also functionally translated (e.g. lots prices in Open Data Monopoly and proof of evidence in WikiMystery). On the far end of the Y-axis we have ANGELINA (Cook, Colton, and Pease 2012), which uses visuals as framing devices but without affecting the core platformer gameplay. The visuals are based on text of a newspaper article: the source data is transformed via natural language processing, tone extraction, and image queries.

Purposes of Maximalist Games

While we attempt to highlight the principles and directions for designing maximalist game experiences, it is important to consider the purpose of such a game. Maximalist game design is desirable for many different reasons, highlighted in this section. Depending on the purpose, moreover, the

design priorities could shift in the spectrum of data fidelity or decoration versus function.

Learning

Modern-day practice sees students of all levels refer to Wikipedia for definitions, historical information, and tutorials. When browsing such knowledge repositories, it is common to access linked articles that may not have been part of the topic of inquiry. Maximalist game design that exploits open sources of knowledge, such as Wikipedia, can be used as a tool for learning in a playful context. One strength found in games is their ability to motivate a player for long periods of time. They also allow several ways to engage players, which can vary based on decisions and learning goals of game designers. Furthermore, games present failure as a necessary event for learning (Plass, Homer, and Kinzer 2015), causing players to explore and experiment more, since failing in game is less consequential than in the real world (Hoffman and Nadelson 2010). Studies have also shown that when games immerse the player in a digital environment, they enhance the player's education of the content within the game (Dede 2009). Games, unlike raw open data, are adaptive to players' skill level and are the most fun and engaging when they operate on the edge a player's *zone of proximal development* (Vygotsky 1978). Thus, we believe players can learn facts within open data during gameplay.

Data-driven maximalist games intended for learning can highlight and allow the player to interact with the data playfully. DATA Agent, to a degree, builds on this concept by creating NPCs out of articles about people, whether they are historical or fictional. These NPCs can then answer questions about their birth date or life's work. More relevant to the game progression, each NPC leads to an object (NPC, item or location) about another article, which can be interacted with and is associated to the current NPC somehow. The Data Adventures series was not designed with the explicit purpose of learning in mind; possibly, alternative design priorities could build on more "educational" principles. It would be possible, for example, to check for understanding by asking the player questions relating to the data in a diegetic manner — not unlike DATA agent, where the player must interrogate suspects and cross-check with their own obtained knowledge to detect falsehoods.

Information used to instantiate data should be fairly transparent to the player-learner. Therefore, transformation of data should not be convoluted, and perhaps even textual elements from original articles can be used as "flavor text". The veneer between encyclopedic content and game content does not need to be thick, in order to ensure that the right information is provided. In terms of function versus decoration, maximalist games for learning tend to edge closer towards data that influences the outcome of a game session in order to motivate learners to understand and remember the data. Such checks for understanding, however gamified they may be, will have an impact on the success or failure of the game.

Data exploration

Data transformed into interactive game content, forming a consistent whole that goes far beyond the sum of its parts, can allow human users to explore the data in a more engaging way. Data visualization has been used extensively with a broad variety of purposes — far beyond the ones listed here — to take advantage of how most humans can more easily think through diagrams (Vile and Polovina 1998). In that vein, gameplay content originating from data can act as a form of highly interactive data visualization. The fact that data from different sources is combined together based on associations imagined by an automated game designer allows players to reflect on the data and make new discoveries or associations of their own. Due to the potential of emotional engagement that games have beyond mere 2D bar plots, the potential for lateral thinking either through visual, semantic, gameplay or emotional associations (Scaltsas and Alexopoulos 2013) is extraordinary. In order for a game to offer an understanding of the data that is used to instantiate it and allow for that data to be re-imagined, the transformation into game content should be minimal. Examples of data games which already perform such a highly interactive data visualization are BarChartBall (Togelius and Gustafsson Friberger 2013) or OpenTrumps (Cardona et al. 2014). However, a more maximalist approach could benefit games like the above by providing a more consistent storyline and progression, as well as a stronger emotional investment in the data.

Contemporaneity

Automated game design has been always motivated, to a degree, by the desire to create perpetually fresh content. With data-driven design, this can be taken a step further by generating a new game every day. Such a game could be contextually relevant based on the day itself, e.g. building around historical events which happened on this day (such an extensive list can be found on onthisday.com and Wikipedia) or people who had important personal events on that day (e.g. date of birth, date of death, graduation day). Moreover, the social context can be mined and used to drive the automated design process by including for instance trending topics of Twitter or headlines of today's newspapers. Early examples of such data-driven process have been explored for example by ANGELINA (Cook, Colton, and Pease 2012) which used titles and articles from The Guardian website and connected them with relevant visuals and the appropriate mood. It is expected that a more maximalist data-driven design process would strengthen the feeling of contemporaneity by including more data sources (i.e. more data to transform) or stronger gameplay implications (i.e. broader transformations and functional impact).

Contemporaneity can make games generated on a specific day appealing to people who wish to get a "feel" for current issues but not necessarily dig deeply. On the other hand, the plethora of games (30 games per month alone) and the fact that each game is relevant to that day only could make each game itself less relevant. Contemporaneity and the fleeting nature of daily events could be emphasized if each game was playable only during the day that it was produced, deleting all its files when the next game is generated. This would enhance the perceived value of each game, similarly to *permadeath* in rogue-like games as it enhances nostalgia and

the feeling of loss when a favorite gameworld is forever lost.

Any maximalist game could satisfy a contemporaneity goal, but such games can be more amenable to data transformation. For example, data could be transformed to more closely fit the theme of the day, e.g. query only female NPCs on International Women's Day. Contemporaneous data can be functional (to more strongly raise awareness of issues) but can also easily be decorative, e.g. giving a snowy appearance to locations during the Christmas holidays.

Personalization

When game content is generated from data, it is possible to highlight certain bits of information. When the game takes player input as part of the data selection process, it personalizes their experience. If player information is available in the form of interests, important personal dates such as birthdays, or even social networks, the potential data sources that can be selected to form the game can be narrowed down. Presenting game content which is personally relevant (e.g. adventures with NPCs based on people living before Christ for an archeology student), or contextually relevant (such as solving the murder of an NPC born on the player's birthday) could contribute to a more engaging experience. It might also be possible to tailor the game's source repositories based on such personal interests. There are numerous online wikis, most of which follow a common format; therefore a user can implicitly (via personal interests) or explicitly (by providing a direct URL) switch search queries of a data-driven maximalist game to a specific wiki of choice.

Opinion & Critique

Often designers want to make a statement through their games. For instance, Game-o-matic (Treanor et al. 2012) creates games from manually defined associations (as *microrhetorics*). *September 12th: A Toy World* (Newsgaming 2003) makes a political statement about the futility of America's War on Terror. Open data could similarly be used in a game to critique some aspect of culture by adding a weight of relevance and realism. For instance, a game such as *September 12th* could use the real map or skyline of Baghdad, or data on daily deaths in Iraq, to instantiate the challenge of the game. Similarly, if designers wish to critique the unprofessional use of social media in the White House, one could use real tweets to form dialog lines rather than generating them as in DATA Agent (Barros et al. 2018a).

Entertainment

Ostensibly, all games have entertainment as a (primary or secondary) purpose. This includes maximalist games, even if they have an additional purpose as listed in this paper. It is meaningful therefore to investigate what data-driven maximalist design has to offer to the entertainment dimension of any such game. Since maximalism —as we define it— does not necessarily apply to the mechanics of a game, a more relevant dimension is the end-user aesthetic that such games facilitate, following the mechanics-dynamics-aesthetics framework of Hunicke, Leblanc, and Zubek (2004). Data-driven maximalist games primarily enhance the aesthetic of *discovery*, similarly to data exploration via such a game, and *expression* if it can be personalized to a user based on provided input such as birthday, hometown or interests. In many ways, data-driven games can enhance the aesthetic of *fantasy* by using and transforming real-world information. DATA agent, for example, describes an alternate history setting where a famous historical figure has been murdered (often by colleagues). The fantasy aesthetic is further enhanced by having a player take the role of a detective traveling through time and space to interrogate suspects. Other possible aesthetics that can be enhanced through data are *sensation* if the data comes from sources of high quality video, audio, or visuals (e.g. paintings of the National Gallery of London), or *fellowship* if the data comes from other users (e.g. anonymous users' trending tweets or social media postings of the player's friends). Evidently, games geared primarily towards entertainment can be fairly flexible in terms of data transformation, and can adapt the data to the intended game mechanics and game flow. While data can act as a decoration in such games (if intended to enhance the sensation aesthetic), in general games intended primarily for entertainment are fairly focused in the mechanics and feedback loops, and thus data would primarily be transformed into functional elements.

Human Computation

Presenting hand-picked results from a vast database in an engaging, playful way is not only relevant for humans to consume. The human-computer interaction loop can be closed if human users provide feedback on the quality of the data itself. This human feedback can be used internally by the game, adapting its criteria in order to avoid unwanted data repositories, queries, associations or transformations made to the data. For instance, a future DATA agent version could re-compute the set of suspects for the next games (removing one or more suspects from the pool of possible suspects) if a player provides negative feedback explicitly (e.g. via a 'report' button) or implicitly (e.g. by being unable to solve the mystery). More ambitiously, the positive or negative feedback of players engaging with the playable —transformed— data can be fed back to the source repositories which instantiated the game. This can allow for instance misinformation found in Wikipedia to be flagged, alerting moderators that either a human error (e.g. a wrong date or a misquote) or malformed data (e.g. unreadable titles) exists and must be corrected. Whether these corrections should be made by an expert human curator, or directly based on player interactions with the game could be a direction for future research.

Issues with Data-Driven Game Design

Accomplishing *good* data-driven maximalist game design is a challenge. While the previous sections presented ways of doing so, there are still many implementation- or game-specific details which affect the design process. Beyond the core challenge of a good game design, there are several peripheral challenges to the design task itself which however spring from the practice of data-driven design. We elaborate on those peripheral challenges here.

Legal & Ethical Issues

Any software which relies on external data that it cannot control may be prone to legal or ethical violations. Privacy of personal information may be a concern for a game generated from the social media profile of a user, especially if that game can then be played by a broader set of people. Using results from Google Images may lead to direct infringements of copyrights; using results from models built from text mining, on the other hand, may or may not result in such copyright infringements depending on whether the model returns actual copyrighted material. The issue of copyright becomes more complex when the data is transformed: relevant to data mining, a judge has ruled for fair use for Google Books as "Google Books is also transformative in the sense that it has transformed book text into data for purposes of substantive research, including data mining and text mining in new areas" (Sookman 2013). One can only assume that transformations of data into game content, depending on the fidelity to the original data and the purpose (e.g. data exploration and education), would make for a clearer case of fair use.

Game content built on fair use or open data combined into an interactive experience may lead to unexpected issues. This is especially true in cases where the player has sufficient agency to interpret or act upon content of high fidelity with the original data in an open-ended fashion: consider, for example, a violent shooter game where opponents' visual depictions (3D models or faces) are those of Hollywood celebrities. Even in Data Adventures, where player interaction is fairly "curated", a generated game featured solving the murder of Justin Bieber (Barros, Liapis, and Togelius 2016a). Apart from the fictional narrative of a popular celebrity's death, the game identifies another celebrity as the murderer: both of these decisions may cause concern to highly visible people (be they depicted murdered, murderers, or suspects). A disclaimer that the game characters are fictional can only alleviate that much of the ethical responsibility of game designers for such data-driven games.

Misinformation & Bias

Connected to the concerns of misrepresenting contemporary or historical celebrities are the inherent issues of error in the source data. Before data is transformed into game content, open repositories that can be edited by anyone can be saturated by personal opinion and perhaps deliberate misinformation. As noted previously, not all data provided by different stakeholders in the information age are factual; this may be more pronounced in certain repositories than others. Beyond deliberate misinformation, an inherent bias is also present even in "objective" data. For example, algorithms for Google query results or image results are based on machine learned models that may favor stereotypes (based on what most people think of a topic). Even though WikiMystery uses what we arguably consider "objective" repositories such as Wikipedia, the 8 most popular locations in 100 generated games were in North America (Barros et al. 2018b), pointing to a bias of the articles or the DBpedia entries chosen to be digitized. Other cases where misinformation may arise is when different content is combined inaccurately: examples from the Data Adventures series include cases where

an image search for a character named Margaret Thatcher resulted in an image of Aung San Suu Kyi (Barros, Liapis, and Togelius 2016b). When data-driven design uses social network data such as trending topics on Twitter, then the potential for sensitive or provocative topics to be paired with inappropriate content or combined in an insensitive way becomes a real possibility. If data-driven maximalist games are intended towards critique or opinion, the misinformation or misappropriation could be deliberately inserted by a designer (by pairing different repositories) or accidentally introduce a message that runs contrary to the intended one.

Outlook

Maximalist game design encourages creation through reuse and combination. If one imagines its most powerful form, it would likely involve taking any mixture of information, pouring it into any game content cast, and reveling in its results. It would provide a freedom to interact with any data in the best, most personalized way possible.

Current PCG techniques allow for unlimited playability for a large variety of games. However, they can lack a level of contemporaneity and relevance that could be provided by open data. Additionally, research has suggested that concepts can be effectively learned through gameplay (Dede 2009). Using games as a method of interacting with open data may create a novel way for learning about the data in a fun way. Rather than use Wikipedia to learn about specific people and places for the first time, players could play games where they can talk to these people and visit these places.

Open data is available to all, to create as well as consume. Sometimes the data is inaccurate. The idea of visualizing this information in any form can provide means to "debug" the original data, in a more engaging way than just browsing Wikipedia or poring through a massive database.

Conclusion

This paper discussed an approach to game design inspired by the notion of maximalism in the arts. It encourages the reuse and combination of heterogeneous data sources in the creative design process. Maximalist game design embraces the generation of game content using different data sources, re-mixing them in order to achieve something new.

We drew from our experience with the Data Adventures series to propose a mapping of the maximalist game design space along two dimensions, *data transformation versus data fidelity* and *functionality versus decoration*. The former focuses on the extent that the data is transformed from its original form, while the latter refers to the actual role of the data in the game. Additionally, we described how maximalist game design can serve different purposes in the design process and which tradeoffs emerge from each purpose. Finally, we highlight issues and ethical concerns that may arise from and in maximalist games.

Acknowledgements

References

Barros, G. A. B., and Togelius, J. 2015. Balanced Civilization map generation based on open data. In *Proceedings of the IEEE Congress on Evolutionary Computation*.

Barros, G. A. B.; Green, M. C.; Liapis, A.; and Togelius, J. 2018a. DATA Agent. In *Proceedings of the Foundations of Digital Games Conference*.

Barros, G. A. B.; Green, M. C.; Liapis, A.; and Togelius, J. 2018b. Who killed albert einstein? From open data to murder mystery games. *IEEE Transactions on Games*. Accepted for publication.

Barros, G. A. B.; Liapis, A.; and Togelius, J. 2015. Data adventures. In *Proceedings of the FDG workshop on Procedural Content Generation in Games*.

Barros, G. A. B.; Liapis, A.; and Togelius, J. 2016a. Murder mystery generation from open data. In *Proceedings of the International Conference on Computational Creativity*.

Barros, G. A. B.; Liapis, A.; and Togelius, J. 2016b. Playing with data: Procedural generation of adventures from open data. In *Proceedings of the International Joint Conference of DiGRA and FDG*.

Cardona, A. B.; Hansen, A. W.; Togelius, J.; and Gustafsson Friberger, M. 2014. Open trumps, a data game. In *Proceedings of the Foundations of Digital Games Conference*.

Cook, M., and Colton, S. 2014. A rogue dream: Automatically generating meaningful content for games. In *Proceedings of the AIIDE Workshop on Experimental AI in Games*.

Cook, M.; Colton, S.; and Pease, A. 2012. Aesthetic considerations for automated platformer design. In *Proceedings of the Artificial Intelligence for Interactive Digital Entertainment Conference*.

Dede, C. 2009. Immersive interfaces for engagement and learning. *Science* 323(5910):66–69.

Gustafsson Friberger, M., and Togelius, J. 2013. Generating interesting monopoly boards from open data. In *Proceedings of the IEEE Conference on Computational Intelligence and Games*.

Gustafsson Friberger, M.; Togelius, J.; Cardona, A. B.; Ermacora, M.; Mousten, A.; Jensen, M. M.; Tanase, V.; and Brøndsted, U. 2013. Data games. In *Proceedings of the FDG Workshop on Procedural Content Generation*, 1–8. ACM.

Guzdial, M., and Riedl, M. 2016. Toward game level generation from gameplay videos. In *Proceedings of the Foundations on Digital Games Conference*.

Hawthorne, C.; Elsen, E.; Song, J.; Roberts, A.; Simon, I.; Raffel, C.; Engel, J.; Oore, S.; and Eck, D. 2017. Onsets and frames: Dual-objective piano transcription. *CoRR* abs/1710.11153.

Heath, D., and Ventura, D. 2016. Before a computer can draw, it must first learn to see. In *Proceedings of the International Conference on Computational Creativity*.

Hoffman, B., and Nadelson, L. 2010. Motivational engagement and video gaming: a mixed methods study. *Educational Technology Research and Development* 58(3).

Hunicke, R.; Leblanc, M.; and Zubek, R. 2004. MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on the Challenges in Games AI*.

Jaffe, D. 1995. Orchestrating the chimera-musical hybrids, technology, and the development of a 'maximalist' musical style. *Leonardo Music Journal* 5.

Khaled, R.; Nelson, M. J.; and Barr, P. 2013. Design metaphors for procedural content generation in games. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 1509–1518.

Khalifa, A.; Barros, G. A. B.; and Togelius, J. 2017. DeepTingle. In *Proceedings of the International Conference on Computational Creativity*.

Lee, S.; Isaksen, A.; Holmgård, C.; and Togelius, J. 2016. Predicting resource locations in game maps using deep convolutional neural networks. In *Proceedings of the AIIDE workshop on Experimental AI in Games*.

Lehman, J.; Risi, S.; and Clune, J. 2016. Creative generation of 3d objects with deep learning and innovation engines. In *Proceedings of the International Conference on Computational Creativity*.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2014. Computational game creativity. In *Proceedings of the International Conference on Computational Creativity*, 46–53.

Liapis, A. 2015. Creativity facet orchestration: the whys and the hows. In *Dagstuhl Reports*.

Nealen, A.; Saltsman, A.; and Boxerman, E. 2011. Towards minimalist game design. In *Proceedings of the Foundations of Digital Games Conference*, 38–45. ACM.

Pease, A., and Colton, S. 2011. On impact and evaluation in computational creativity: A discussion of the turing test and an alternative proposal. In *Proceedings of the AISB symposium on AI and Philosophy*.

Plass, J. L.; Homer, B. D.; and Kinzer, C. K. 2015. Foundations of game-based learning. *Educational Psychologist* 50(4):258–283.

Ritchie, G., and Masthoff, J. 2011. The STANDUP 2 interactive riddle builder. In *Proceedings of the International Conference on Computational Creativity*.

Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17.

Scaltsas, T., and Alexopoulos, C. 2013. Creating creativity through emotive thinking. In *Proceedings of the World Congress of Philosophy*.

Sicart, M. 2015. Loops and metagames: Understanding game design structures. In *Proceedings of the Foundations of Digital Games Conference*.

Sookman, B. 2013. The Google Book project: Is it fair use? *Journal of the Copyright Society of the USA* 61:485–516.

Summerville, A. J.; Snodgrass, S.; Mateas, M.; and Ontañón, S. 2016. The VGLC: The Video Game Level Corpus. In *Proceedings of the FDG Workshop on Procedural Content Generation*.

Togelius, J., and Gustafsson Friberger, M. 2013. Bar chart ball, a data game. In *Proceedings of the Foundations of Digital Games Conference*.

Treanor, M.; Blackford, B.; Mateas, M.; and Bogost, I. 2012. Game-o-matic: Generating videogames that represent ideas. In *Proceedings of the FDG Workshop on Procedural Content Generation*.

Veale, T. 2014. Coming good and breaking bad: Generating transformative character arcs for use in compelling stories. In *Proceedings of the International Conference on Computational Creativity*.

Vile, A., and Polovina, S. 1998. Thinking of or thinking through diagrams? The case of conceptual graphs. In *Thinking with Diagrams Conference*.

Vygotsky, L. 1978. Interaction between learning and development. *Readings on the development of children* 23(3).