

Open Trumps, a Data Game

Andrew Borg Cardona
Center for Computer
Games Research
IT University of Copenhagen
Rued Langaards Vej 7,
2300 Copenhagen, Denmark
aboc@itu.dk

Julian Togelius
Center for Computer
Games Research
IT University of Copenhagen
Rued Langaards Vej 7,
2300 Copenhagen, Denmark
julian@togelius.com

Aske Walther Hansen
Center for Computer
Games Research
IT University of Copenhagen
Rued Langaards Vej 7,
2300 Copenhagen, Denmark
aswa@itu.dk

Marie Gustafsson
Friberger
Department of
Computer Science
Malmö University
Östra Varvsgatan 11A,
211 19 Malmö, Sweden
marie.friberger@mah.se

ABSTRACT

We present *Open Trumps*, a version of the popular card game *Top Trumps* with decks that are procedurally generated based on open data. The game is played among multiple players through drawing cards and selecting the feature that is most likely to trump the same feature on the other players' cards. Players can generate their own decks through choosing a suitable dataset and setting certain attributes; the generator then generates a balanced and playable deck using evolutionary computation. In the example dataset, each card represents a country and the features represent such entities as GDP per capita, mortality rate or tomato production, but in principle any dataset organised as instances with numerical features could be used. We also report the results of an evaluation intended to investigate both player experience and the hypothesis that players learn about the data underlying the deck they play with, since understanding the data is key to playing well. The results show that players enjoy playing the game, are enthusiastic about its potential and answer questions related to decks they have played significantly better than questions related to decks they have not played.

1. INTRODUCTION

The increasing amount of data means that there is an increased need to understand and make sense of data. One category of data that is becoming increasingly available is that of open government data (OGD), where governmental and other organisations release their data in a format and license so that others can understand it and potentially

reuse it. The main reasons for doing this are to increase transparency and enable innovation. Government data is commonly released using data portals such as that of the US¹ and UK² governments. Other organisations include the United Nations³ and the World Bank⁴, as well as cities, such as London⁵. Visualisations, mashups, and dedicated applications are means of making use of open data. Examples include the GapMinder tool [7] for visualising world development, mobile apps for comparing general practitioners, and mashups for seeing where bike accidents happen.

It is evident that the release of such quantities of important data in a structured form represents a major opportunity, but the opportunity comes with its own set of challenges relating to how to access, select, visualise and interpret the data. Common methods to enable understanding and overview of data are through tables, diagrams, and interactive visualizations.

There is clearly a need for more engaging and usable methods for exploring and visualising such data. It has previously been proposed that another means of making sense of data is through games, an approach called *data games*. As proposed in [5] "Data games are games where gameplay and/or game content is based on real-world data external to the game, and where gameplay supports the exploration of and learning from this data."

In most cases, the raw data in itself does not make for good game content, but it needs to be transformed in some way. Procedural content generation (PCG) refers to the algorithmic creation of game content (levels, rules, items etc) with limited human input [8]. Part of the data games concept is that PCG techniques are used to create the content of games

¹data.gov

²data.gov.uk

³data.un.org

⁴data.worldbank.org

⁵data.london.gov.uk

based on real-world data.

Data games can be seen as a form of serious games [1], as the primary purpose is not entertainment but rather assisting the player in understanding and engaging with data [5]. However, they should be distinguished from educational games that aim to teach the player a particular curriculum; instead, data games focus on enabling player-guided exploration of data.

Building games on real world data is both an opportunity and a challenge. We identify four themes of research questions in [5]: (1) Exploration, learning and playability, (2) Data selection and access, (3) Game design and development, and (4) evaluation, necessary to inform the previous three. In this paper, we examine research questions related to the first, third and fourth themes:

1. Is it possible to automatically create a balanced, playable Trumps deck based on open data?
2. Do participants enjoy playing Open Trumps?
3. Do participants score quiz questions significantly better when having been exposed to related data through Open Trumps than when they have not?

In [5], we also provide several examples of data games, such as: Open Data Monopoly (where the player generates a Monopoly board by selecting and weighing indicators for UK municipalities, which then form the “streets” of the game board), Bar Chart Ball (where a bar chart makes a platform game, and game movement is done through changing the indicator visualized), and Open Street Racer (racing games generated based on open street maps).

In this paper, we describe and evaluate Open Trumps, which is based on the game of Top Trumps (Section 2) and data sets from the UN and the World Bank describing different nations. Top Trumps has the benefit of being mappable to these types of data sets, as well as being a game that is highly accessible to many types of players. The first part of this paper is description of the Top Trumps game design (Section 3) and the card generator (Section 4). The second part of the paper describes an evaluation of Open Trumps through a focus group and user tests, methodology (Section 5) and results (Section 6). Finally, discuss the game and its evaluation (Section 7) as well as conclude and look forward (Section 8).

2. TOP TRUMPS

Top Trumps is a card game where the deck of cards is based on a theme, such as cars, aircrafts, or characters from a film. Each card in the deck, such as a specific car model, aircraft or film character includes an image and description, and most importantly, specific values for a handful of attributes (such as speed, size, mental ability, as relevant to the theme). For example, cards may represent car models (example attributes: top speed, engine capacity, price) or football players (example attributes: height, years of international experience, caps). See Figure 1 for an example card.

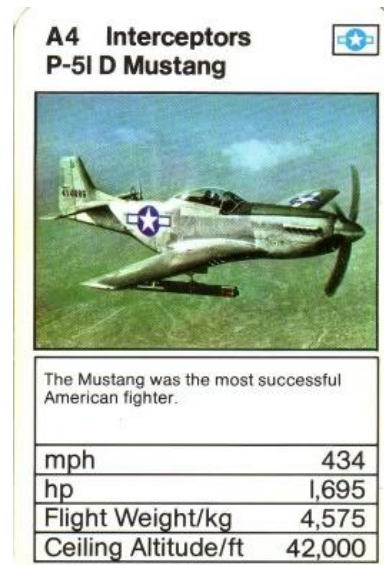


Figure 1: An example of a Top Trumps card from an aircraft-themed edition of the game.

The game is played with at least two players. All cards are dealt among the players, each receiving an equal number of cards. The starting player (the one next to the dealer) draws a card from their deck, selects an attribute and reads its value. The other players read out their values for this attribute, and the player with the ‘best’ (usually the highest) value wins all of the cards, which are placed at the bottom of the winner’s stack. Due to the way in which the card decks for Open Trumps are generated, the highest value is always taken as the winner. The next attribute is selected by the winner of the previous round. When a player has no more cards, they drop out, until the winner has collected all cards.

3. DESIGNING OPEN TRUMPS

The design task we set ourselves was to create a generator for Top Trumps-style games, that would generate card decks (games) based on a data set chosen and according to further criteria (such as deck size) set by the users. This means that we need to extract data in such form that an instance of the data could correspond to a potential card, and an attribute of the data set could correspond to a potential card attribute. It also means that we should provide an easy way for the user to specify their criteria for deck creation, for example a graphical user interface.

Furthermore, we need the card sets to be playable, and preferably also enjoyable to play. To do this we need to understand the Top Trumps game dynamics, in particular what constitutes a good and a bad card deck. Arguably, a good deck rewards skill, so that a player that knows which feature to play does better than a player that picks random features. Therefore, a good deck does not contain cards that beat all (or most) other cards regardless of which feature is chosen, or cards that lose to most cards regardless of chosen feature. (In the language of optimisation, as many cards as possible should be close to the *non-dominated front*.) There-

fore, we formulated the following criteria for a good Open Trumps deck: *As many cards as possible should be able to win against as many cards as possible.* Note that this implicitly means that each card will also lose to as many cards as possible, if the wrong feature is played. This definition was directly used in our generator, as described below.

4. THE OPEN TRUMPS GENERATOR

The Open Trumps generator software takes one or several data sources and generates a complete, playable card deck in the form of a table where the rows represent the individual cards and the columns represent the attribute. The current version does not generate a graphical layout of the cards or any images, but this can easily be done manually and would be straightforward to do mechanically. The software is written in C# and controlled from a graphical user interface that allows for user control of key steps of the process.

The software has two parts: the preprocessor and the generator. The preprocessor aggregates and preprocesses data into a format which can be used by the generator, namely rows which represent potential cards and columns containing potential features. The generator selects a subset of cards and features that is interesting and playable. Through this modular approach, the generator can always assume the expected format of the data and the procedural content generation can proceed without taking data formatting issues into account.

4.1 Preprocessor

The preprocessor is a simple but extensible module that caters for varying data sources as long as an appropriate library is provided. To demonstrate this feature, two libraries were developed, one handling United Nations published open data (United Nations Statistics Division) and another handling the World Development Indicators from the World Bank Group (The World Bank Group).

The general program flow is as follows. First, the user browses for the library file used for the data format of the particular data source, and then manually selects the location of the input and output files. The interface for performing this can be seen in Figure 2. Depending on what is specified in each library, the preprocessor can read e.g., XML or CSV files, and handle common preprocessing tasks such as normalisation. The output of the preprocessor is a single CSV file where each line represents a possible card (e.g., a country) and each column represents a possible attribute (e.g., GDP). It is not guaranteed that values are specified for all features on all cards in the current implementation of the preprocessor.

4.2 Generator

The more interesting part of the framework is the generator, which generates decks of cards. Simply put, the generator selects a subset of all rows (cards) and columns (attributes) in the CSV file, and generates a new, smaller file from these. What is interesting is how it does this.

The interface to the generator is shown in Figure 3. The user can select how many cards to generate (50 in the example) and how many attributes to use per card (4 in the example).

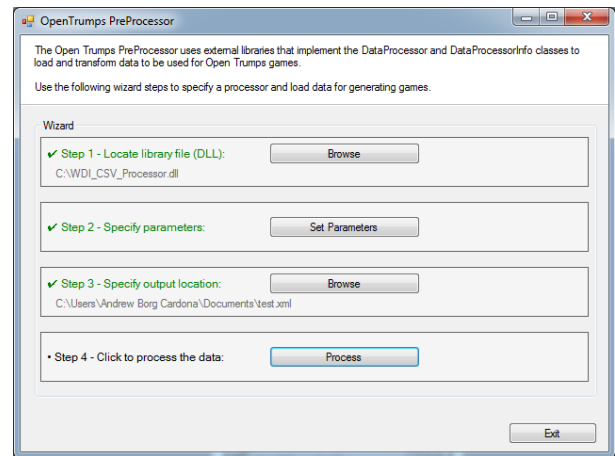


Figure 2: The preprocessor interface.

If the file generated by the preprocessor divides the possible attributes into categories, the user can also select which attribute categories to use.

Deck generation is done by a very simple evolutionary algorithm: a 1 + 1 Evolution Strategy, also known as a stochastic hillclimber. The genome consists of two fixed-size lists representing a full deck of cards: one of which instances (countries) to use, and one of which attributes to use. The only genetic operator is mutation, which works through replacing between 5 and 15 countries with other countries that have defined values for the chosen attributes. With a small probability it can also replace the deck with a completely new deck.

Initialisation of the first individual is done in the following manner: First a single country is picked randomly. Then a number of attributes are picked, such that the single selected country has defined values for these attributes. (Some countries lack values for some attributes.) Then, the rest of the countries are chosen randomly, with the only constraints that (1) they should have defined values for all the chosen attributes and (2) they should be unique.

As described in Section 3, the goal of the generator is to create decks where each card could potentially win against as many cards as possible. This was implemented as a fitness function by simply summing the number of cards each card could potentially win against, if the best attribute was chosen. The fitness for each card ranges between 0 (if it can never win over any other cards) and 1 (if it would win over all other cards given the right attribute), and the fitness of the whole deck is simply the sum of fitness of all cards.

Initial testing suggested that fitness growth had stagnated by the 200th generation. By this time, the average winning rate of a deck was over 0.95; therefore the length of an evolutionary run was fixed at 200 generations.

5. EVALUATION METHODOLOGY

In evaluating Open Trumps, we began with a focus group followed by user tests. Before describing the methods used for these, we describe the decks used.

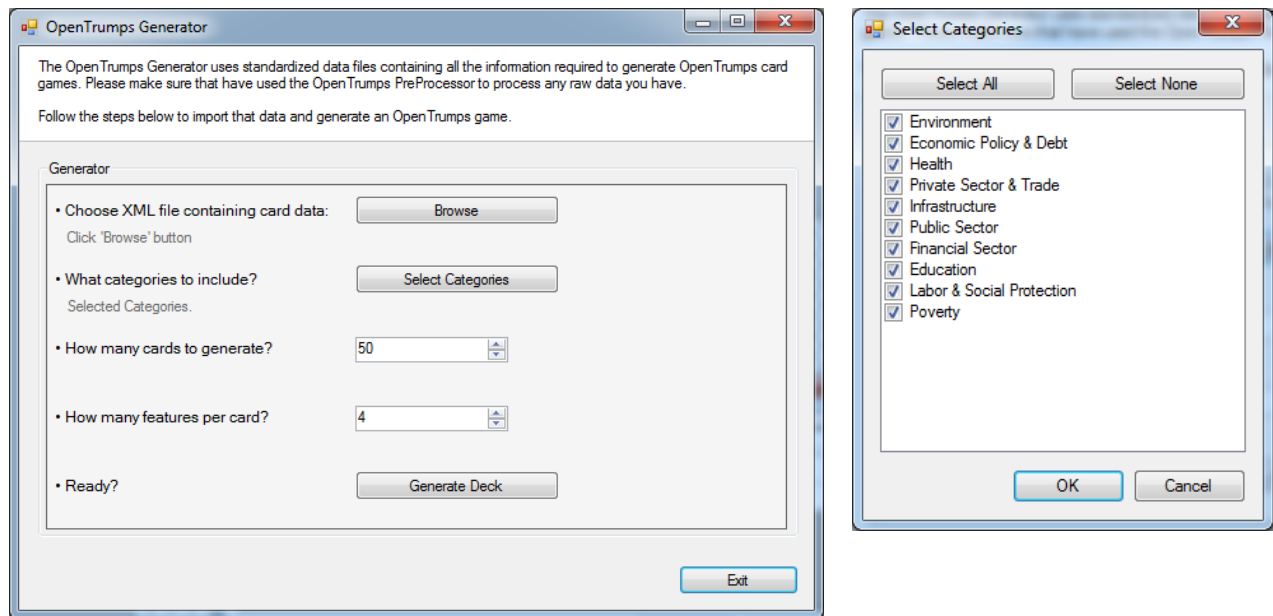


Figure 3: The generator interface.

5.1 Decks

In order to test Open Trumps, a physical prototype based on decks generated by the Open Trumps software was created. Figure 4 shows an example of a card with the attributes Tomato Production, Adult Mortality Rate, Life Expectancy and Mortality Rate from Cancer. Flags were used as illustrations for the countries. The decks were printed and glued to standard decks of cards. Four decks were created: one for the focus group and three for the user tests. For each user test the participants played two of the three decks. Which decks were used followed a repeated pattern of the six possible permutations. The rules for Top Trumps as outlined in Section 2 were used.

5.2 Focus group

Before conducting the user study, a focus group of five participants (three men and two women) was assembled to study what issues might be important to investigate further. The format is a group discussion where the participants can engage with physical artefacts and discuss open-ended questions [6]. The result of a focus group is shaped by group dynamics and one cannot use the results the same way as individual interviews or user tests.

The structure of the focus groups was as follows:

1. Introduction, where the participants fill out a small questionnaire.
2. Introduction to OGD and the current UI.
3. Prototype game play. One game with three players and one with two players. All participants played the game.
4. Three step group discussion of, e.g., strengths, weaknesses, potentials, beginning with initial, unstructured

thoughts. This was followed by facilitation through cards with value-loaded words inspired by Product Reaction Cards [9]. Finally, various positive or negative statements about the game were used as starting points for discussion.

5. Final comments, e.g., summing up or about issues that had not yet been covered.
6. Final questionnaire.

The questionnaires mainly had the purpose to test quiz questions about the played deck. Some questions were the same on the pre- and post-session questionnaire and others were unique. This way we got initial indications on whether it would make sense to test learning this way.

The focus group session was recorded and transcribed, after which the data was sorted and categorized [6]. To make sense of the results we used Severity Rating from 1-4 (irritant, moderate, severe, unusable) of the identified issues combined with how many who agreed that the given was an issue [9]. Based on this we identified issues to proceed with.

5.3 User test

The goal of this test was to collect enough quantitative data on both enjoyment and learning to make it possible to identify significant results. With this in mind we designed a relatively quick test consisting of two games of each 10 minutes followed by one questionnaire. The average length of a test including introduction was about 35 minutes. Participants in all tests were students recruited from the corridors of the university where the study took place.

To evaluate the test design, a pilot test was conducted with two students. The test resulted in some changes in the test



Figure 4: Example of a prototype card based on open data (in Danish).

guide to facilitate a smoother test. The general test design showed to be satisfying.

18 students took part in the user test. A test guide was used in order to streamline the tests as much as possible. To facilitate understanding and to emphasize that the underlying software actually was implemented and used in the creation of the prototype we gave the participants a printout of the UI as shown in Figure 3. We asked participants to keep in mind the underlying software when answering the questionnaire.

To capture the complete picture we kept notes of observations during each test, we recorded which attributes were played how many times, how often the one to select attribute switched, which countries were mentioned and/or discussed, how they reacted on the attributes and if they laughed a lot or were especially silent.

5.3.1 Self-reported metrics: Test of enjoyment

To test the enjoyment of Open Trumps we used self-reported metrics in the questionnaire inspired by the USE questionnaire [9] but with several questions relating specifically to the qualities of Open Trumps. Table 1 shows the statements used in the questionnaire.

Each question was rated on an analogue scale where the user puts a mark on a line to score how much they agreed with the statements from “Strongly disagree” to “Strongly agree”. This results in a value between 0-100 where the extremes 0 equals “Strongly disagree” and 100 equals “Strongly agree”. A score right between the two extremes (50) can be understood as a neutral stand. To test the whether ratings differ significantly the Students’ t-test is appropriate since it is valid when you have a small sample size and the population standard deviation is unknown [4].

Table 1: Statements used in the questionnaire.

A	In general I like to play games
B	Open Trumps is fun
C	Open Trumps is flexible
D	It is effortless to play Open Trumps
E	Open Trumps holds a greater value than regular trump cards
F	I would get Open Trumps if it was fully developed as a mobile or computer game
G	I would recommend Open Trumps to a friend if it was fully developed as a mobile or computer game
H	I would like to play the game again with the same attributes
I	I would like to play the game again with new attributes
J	I find the attributes we have used satisfying

5.3.2 Quiz: Test of learning potential

To test the learning potential of Open Trumps we used 12 quiz questions of three different types. The questions were designed so that there were four similar questions about content from each deck. A correct answer was given a 1 and otherwise 0, so that a participant receives a score from 0-4 for each of the three sets of questions related to the different decks. We do not take into account the order in which the two decks were played when scoring learning potential. Thus, any difference in effect from playing, e.g., deck 1 before deck 2 versus playing deck 2 before deck 1 are not considered.

In this case it is not proper to test for significance with the t-test because the sample comes from a non-normal distribution. Instead, a Mann-Whitney U-test was used to test whether the two samples are from different distributions (Meery, et al., 2011 pp. 452-459). With $n > 5$ the U-test returns a p-value corresponding to the p-value returned from a typical t-test. To do calculations, vassarstats.net was used, for both the self-reported metrics and the quiz.

6. EVALUATION RESULTS

6.1 Focus group results

6.1.1 General dynamics

The immediate expectations expressed when introduced to Open Trumps was that it might hold a noticeable learning potential. When playing the game the participants often discussed the countries, especially when they knew the country or when there were attributes with what they perceive as extreme values. Further, they shared knowledge with the group, when they thought they had some knowledge explaining a specific value. For instance explaining a low “total female enrollment” with civil war. Concerning enjoyment, the game often provoked laughter and surprised looks and statements about strong or weak countries.

6.1.2 Usability issues

Issues that came up were the general understanding of attributes, whether a high or low value should win, the relevance of some attributes, and if the data was reliable.

Throughout the discussion, it became evident that the participants did not understand the different attributes in the same way and that they in general found it difficult to identify the actual meaning of the attributes. Some attributes were abstract and ungraspable to some of the participants. Ease of doing business index was at first hard to understand but throughout the game participants realized that high values were associated with difficulties doing business. Short-term debt was confusing and the participants never landed on a clear understanding of the attribute. Based on the problems they all agreed that it would help to have a point of reference like Denmark or England. The deck mostly consisted of non-Western countries.

Despite the problems of understanding the attributes, the participants quickly grasped what constituted high values for most attributes. In the beginning of the game, there was a lot of discussion about how to decide what constitutes a good value. For example, a large “annual freshwater withdrawal” was at first identified as good, but after discussing environmental and sustainability issues it was agreed on that low values should be best. When Open Trumps generates a deck, it always assumes that high values beat low values. The discussion showed that this evidently did not go with the way the participants understood the attributes.

Some pointed out that they were in doubt whether they could trust in the values given on the cards. Are they up to date and are there false values? For instance, one participant wondered if a value of zero really meant a hole in the data set.

6.1.3 Value of Open Data

The participants all agreed that Open Trumps holds potential by using OGD. They wanted a digital multiplayer application for smart devices and several times, they referred to the app “Quiz battle” and emphasized the strength of short intervals of gaming at a time. One participant said that even though he did not find the specific attributes of interest he enjoyed the game and thereby he found the learning potential great. Some also pointed out that the game could hold big value for niche groups of different kinds. For instance, young people studying medicine, golf players, WoW enthusiasts etc.

6.1.4 Learning potential

The participants answered the same two questions both before and after the session. Pre-session they answered 1 out of the 10 questions correct. Post-session they answered 7 out of the 10 questions correct. This indicated that a test of learning potential might be fruitful.

6.1.5 Severity rating and the prototype

Based on a Severity Rating and the number of participants who agreed with the different issues we chose to handle the top three issues in the final test. The top three issues were:

- Incomprehensible attributes, Severe, 80% agreement
- Irrelevant attributes, Severe, 100% agreement
- High or low wins?, Moderate, 100% agreement

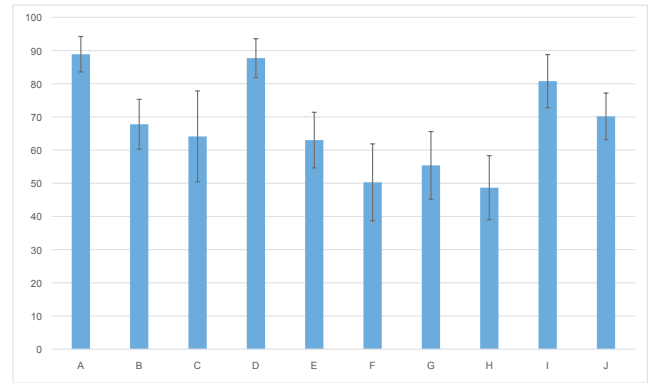


Figure 5: Overall average on each statement in Table 1.

For the user test we tried to eliminate these issues by using attributes that are understandable and have a general relevance, as well as by emphasizing that highest values always win. The identified issues challenge our definition of an interesting deck and shows that a strict focus on a balanced deck might be the wrong priority.

6.2 User tests

6.2.1 Enjoyment

Figure 5 shows the average score by all participants (n = 18) on each statement in Table 1.

They find the used attributes satisfying (J) and there is no negative attitude towards the fun of the game (B). The participants like in general to play games (A). Whether Open Trumps holds a greater value than regular trump cards (E) scores above neutral meaning that there in general is some agreement with the statement. Whether the participants would like to play the game again with respectively the same (H) or new (I) attributes are significantly different from each other (H = 48.7, I = 80.8, p = 0.00003) which tells us that the possibility to choose new attributes are very valuable. This is an indication that the use of OGD enhances the enjoyment compared to regular trump cards.

No one combination of decks differs significantly from the two others in the way each statement is rated.

We observed a combination of occasional laughter and surprised reactions on attributes together with intervals where the participants played the game in silence. The typical pattern was that each new game started with a relative high energy level that later in the game mellowed down. This was typically the case with both the first and the second deck the participants played, and it can indicate that the change of attributes enhances the durability of the game compared to playing the same deck several times.

In the generation of cards, it is assumed that the highest value always wins, and this was the instruction to the players. However, they found it illogical that negative attributes like mortality won by having the largest value. Instead they found it more meaningful to win by having the lowest mortality.

Table 2: Scores grouped by whether related deck has been played or not

n Question related	n Question not related	\bar{x} Played related	\bar{x} Not played related	p-value
36	18	1.67	0.83	0.0045

6.2.2 Learning potential

As described in Section 5.3.2, each participant gets three scores between 0-4, one for each deck.

Table 2 shows that the scores derived from participants who have played the related decks are significantly higher than the scores derived from participants who have not played the related decks.

In our observations of game play, we found that extreme values made an impression. The test subjects often mentioned and even discussed when they met values they experienced as extreme. For instance, Chad was mentioned in most games with the two decks that included Chad. Similarly the tomato production in the USA was mentioned in most games with the deck where it was included. Several times the USA was called a “Joker” or similar names indicating its extremely high tomato production. This also showed in the questionnaire, where 5 out of 12 commented on the large tomato production in the USA.

The participants only rarely showed any sign that they did not understand the attributes though some verbal comments showed that they misunderstood some of them without realizing it. This is probably due to the fairly complex attributes explained by only one line. For instance, “Adult Mortality Rate (Number of 15-year olds out of 1.000 that die before they reach the age of 60)”. In order to understand the data the explanation in the parenthesis will generally be needed, but at the same time some players might never take the time to read the text in the parenthesis. This was also indicated in the way the attributes often were mentioned by the shortest possible label. For instance, Adult mortality rate might be shortened to mortality rate, even though the same deck had a cancer mortality rate as well. This several times led to misunderstandings during the games.

Another example of difficulties understanding the attributes was how a participant understood the HIV-percentage attribute on one of the decks as the percentage of women in a given country who had HIV. In fact the card explained that the percentage was in relation to the total HIV infected meaning that a value of 40% meant that the “last 60%” was HIV infected men and not healthy women.

Finally the attributes are not chosen by how they relate to each other and do not make any exhaustive whole, and so the player is at risk of drawing false conclusions. This is exemplified in comments from the questionnaire like: “Poor countries have a smaller occurrence of cancer than do the rich countries?” and “Mortality rate and mortality rate by cancer does not necessarily cohere”. The first comment might not take into account that the average lifetime is generally lower in poor countries which could be part of the expla-

nation of a smaller occurrence of cancer, and similarly with the second comment. The point is that data are taken from a larger context which might make it hard to really understand the presented data.

7. DISCUSSION

The data sets themselves can pose problems both to developers and players. This issue has been somewhat discussed by others. Ding et al. [3] discuss the issues of handling incomplete information, special codes or acronyms, as well as lack of metadata, problems encountered when developing the game, some of which affected also the game play. Dawes [2] points out that the potential promising value of OGD comes with risks for validity, relevance, and trust. Data quality is a cause of this, and Dawes suggests that the open data community can learn from the work done with data quality the nineties, looking towards the work of e.g., [10], where four data quality aspects are identified:

- Intrinsic quality, e.g., accuracy, objectivity, believability and the reputation of the data source.
- Contextual quality, e.g., timeliness, relevancy, completeness and value to the user.
- Representational quality, e.g., formats and whether data is understandable and easy to interpret.
- Accessibility, e.g., ease of access and access security.

The evaluation of Open Trumps points specifically to two of these issues: the contextual and representational quality. Both enjoyment and learning potential were hampered especially by problems of relevancy, completeness, and the degree to which the data is understandable and easy to interpret. How to enhance the quality of the data and the meaningfulness from the players’ perspective needs further investigation.

These aspects might be informed by initiatives where people engage as laymen with large data sets, such as crowdsourcing activities. It is also relevant to analyze similarities and differences between data games and crowdsourcing, in terms of, for example, learning and engagement. Also, what is the potential of using data games as a crowdsourcing platform?

It is possible that the fitness function could be refined. The current fitness function does not explicitly take the balance between cards into account, so in theory a deck where most cards win over many other cards but some cards lose against all others could get good summed fitness. We have not investigated whether this actually occurs. The design of the fitness function is clearly a key problem to address for future work. Further, a more sophisticated evolutionary algorithm could be used, and the mutation operator could probably be improved.

It is also worth questioning the basic principle for building these decks that was laid out in Section 3 and forms bases for the fitness function of the generation: that each card should be able to win against as many cards as possible. While a deck built according to this principle indeed rewards skill in the sense that a player that can choose the best feature

will play better, it can also lead to situations where the same player keeps winning every round because every card will almost always win given that the best feature is chosen. This might be counteracted by setting as a target that each card should win often, but not every time, when a feature is chosen. For example, the target might be a 75% win rate given that the best feature was chosen, rather than more than 95% as is commonly the case with decks generated using the current fitness function. Alternatively, one could look into adding some extra mechanic to the game to give it more depth.

As found out in the user test, players had problems with understanding some of the attributes, either because they were too complicated or because they were simply unknown. One research challenge is to automatically identify which features would be known to the target group, and how they could be succinctly presented to the player. To provide more background on the attributes, the generator could have an option to include the range of values for each attribute (within the deck) on each card, which could both improve gameplay and learning. It would be also useful to be able to automatically identify whether lower or higher is better for a given feature. Alternatively, this could be made part of the game.

Another issue is rooted balancing the selection of countries. Decks with both poor and rich countries typically consist of a mix of negative and positive attributes, so that the rich countries win on the positive and the poor ones win on the negative. This can potentially result in decks where countries with specific characteristics will be left out of a deck and, for instance, only the richest and poorest countries are used.

The current implementation of Open Trumps is somewhat cumbersome to use, as it only runs in Windows and produces a CSV file as an output, from which the player has to make the actual cards. It would be very desirable to make a web-based version of the framework. This version should allow the user to select data sources from anywhere on the web, and also automatically find suitable pictures to make the actual cards, which should be delivered in a form ready for the user to print. These improvements would be quite straightforward.

Finally, we would like to perform a deeper evaluation of Open Trumps, trying to evaluate in detail what sort of learning is facilitated by the game and over how long time periods. One interesting study could be whether testing over time exposing players to the same countries many times but with different attributes would give them a fuller picture of the countries. There is also a need to investigate problems with players misunderstanding or simplifying attributes used in the game.

8. CONCLUSION

This paper has presented the game Open Trumps, where trumps cards representing nations are automatically generated based on open government data. The enjoyability and learning potential of the game was evaluated through user tests involving 18 university students. In the user satisfaction survey, users agreed that the game was enjoyable and had great potential. The results of a quiz confirmed that

there is a learning potential by playing Open Trumps. The test shows that when participants immediately after playing Open Trumps answer the questions related to the decks they played their answers are significantly more often correct than their answers on the questions related to the deck they did not play.

Even though these are promising results, the user tests also reveals a number of outstanding challenges. These include how to automatically select attributes that the users can understand and perceive as important, meaningful as well as trustworthy, understanding whether higher or lower is better, and presenting them in a succinct form. More work should also be done to make the system user-friendly and available.

Acknowledgements

Thanks to Richard Bartle for suggesting the idea of looking into Top Trumps as a data game.

9. REFERENCES

- [1] I. Bogost. *Persuasive Games: the Expressive Power of Videogames*. MIT Press, Cambridge, MA, USA, 2007.
- [2] S. S. Dawes. A realistic look at open data. *Center for Technology in Government, University at Albany/SUNY Available at http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf*, 2012.
- [3] L. Ding, J. Michaelis, D. L. McGuinness, and J. Hendler. Making sense of open government data. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.
- [4] C.-. Foundation. *CK-12 Probability and Statistics-Advanced , Volume 1 Of 2*. CK-12 Foundation, 2010.
- [5] M. G. Friberger, J. Togelius, A. B. Cardona, M. Ermacora, A. Mousten, M. Jensen, V. Tanase, and U. Brøndsted. Data games. In *Proceedings of the Workshop on Procedural Content Generation, colocated with Foundations for Digital Games*, 2013.
- [6] J. Kitzinger. Qualitative research. introducing focus groups. *BMJ: British medical journal*, 311(7000):299, 1995.
- [7] H. Rosling. Visual technology unveils the beauty of statistics and swaps policy from dissemination to access. *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics*, 2007.
- [8] J. Togelius, G. N. Yannakakis, K. O. Stanley, and C. Browne. Search-based procedural content generation: a taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3:172–186, 2011.
- [9] T. Tullis and W. Albert. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann, 2010.
- [10] R. Y. Wang, D. M. Strong, and L. M. Guarascio. Beyond accuracy: What data quality means to data consumers. *J. of Management Information Systems*, 12(4):5–33, 1996.