Proceedings of the ASME 2015 International Design Engineering Technical Conferences &
Computers and Information in Engineering Conference
IDETC/CIE 2015
August 2-5, 2015, Boston, Massachusetts, USA

Draft of DETC2015-47383

A NOVEL APPLICATION OF GAMIFICATION FOR COLLECTING HIGH-LEVEL DESIGN INFORMATION

Hyunmin Cheong, Wei Li, Francesco Iorio

Autodesk Research
Toronto, ON, Canada

ABSTRACT

This paper presents a novel application of gamification for collecting high-level design descriptions of objects. High-level design descriptions entail not only superficial characteristics of an object, but also function, behavior, and requirement information of the object. Such information is difficult to obtain with traditional data mining techniques. For acquisition of high-level design information, we investigated a multiplayer game, "Who is the Pretender?" in an offline context. Through a user study, we demonstrate that the game offers a more fun, enjoyable, and engaging experience for providing descriptions of objects than simply asking people to list them. We also show that the game elicits more high-level, problem-oriented requirement descriptions and less low-level, solution-oriented structure descriptions due to the unique game mechanics that encourage players to describe objects at an abstract level. Finally, we present how crowdsourcing can be used to generate game content that facilitates the gameplay. Our work contributes towards acquiring high-level design knowledge that is essential for developing knowledge-based CAD systems.

INTRODUCTION

Current computer-aided design (CAD) systems support designers in modeling and analyzing their solutions. However, they offer limited capabilities in synthesizing solutions or helping designers explore potential solutions. Knowledge-based CAD systems [1], which are intended to offer those capabilities, have been proposed and investigated for a few decades. However, they have been limited to a small number of specific domains [2]. Typically, it is difficult to construct an extensive knowledge base to support such systems, especially high-level design information, e.g., requirement, function, and behavior of a desired solution. High-level design information is essential in formulating design problems and driving design processes, but it can be hard to acquire.

Several knowledge representation frameworks have been proposed to describe a design process and the types of information considered during the process [3-5]. To define the types of knowledge that we aim to acquire, we consider Gero's function-behavior-structure (FBS) framework [3], which has been used to analyze various design studies [6]. For the current paper, we consider four types of information elements in FBS: requirement, function, behavior, and structure [7].

Figure 1 provides an example of how descriptions of a cane could be classified using the FBS framework. Requirements identify problem goals, desired or undesired properties, users, use context, etc. of an object to be designed. The function of an object describes its intended purpose or action that will satisfy the requirements. The behavior of an object describes how its structure achieves its function. The structure consists of "the elements of an object and their relationships" [7]. In our current work, we consider the structure information as "low-level", which is often needed during the detail/part design stage, and the requirement, function, and behavior information as "high-level", which are often needed during the conceptual design stage. The high-level information is also not tied to the physical form of a particular object, which means that it describes the design problem rather than the solution.

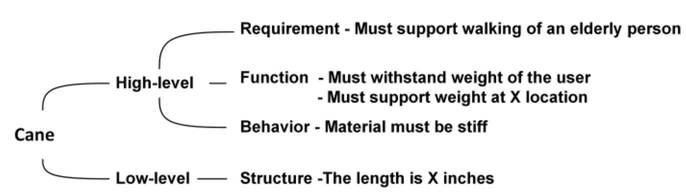

Figure 1. Descriptions of a cane classified using the FBS framework [7].

Acquiring high-level design information is a significant challenge. Knowledge acquisition efforts have been made to extract various elements of design knowledge, e.g., [8-18]. However, most of the readily available sources of design data are the structure information of objects that have already been designed, e.g., product specifications, catalogues, and CAD models. Therefore, the knowledge that can be extracted from such sources is often limited to low-level design information, i.e., solutions. Data sources that contain high-level design information, i.e., problems, are rarely available.

To acquire high-level design information, we investigated leveraging gamification and crowdsourcing. Both techniques have been shown to facilitate knowledge acquisition at a large scale. Crowdsourcing can leverage the massive parallelization offered by a crowd performing intelligent tasks that machines cannot perform. Gamification can make the tasks enjoyable and competitive, and therefore could provide intrinsic motivation for people to perform the tasks.

We investigated a game that we named as "Who is the Pretender?" It is a multiplayer game that mainly involves players to describe simple objects without revealing too much information about the objects, and guess each other's objects. We used this game because it encourages players to provide high-level descriptions of objects instead of providing superficial characteristics of objects that could easily reveal the objects. The game also features both cooperative and competitive game mechanisms. In this work, we focus on evaluating the game interactions in an offline context and whether they can elicit high-level design descriptions from players.

As part of the current work, we also investigated the use of crowdsourcing to generate game content. The objects used in the game are prepared in functionally similar pairs to ensure enjoyable gameplay. We used Amazon Mechanical Turk (MTurk)[1] to generate the necessary game content and show that it is an effective method for generating game content.

This paper makes the following contributions. We present a novel approach to collect high-level design descriptions of physical objects using gamification. Our method is different from existing gamification systems in that it could elicit high-level semantics data from players and features rich multiplayer game dynamics. In addition, we demonstrate that this gamified method suffices to collect high-level design information. We show the benefits of the method over simply asking people to provide data in terms of both the differences in the collected data and the measures that indicate the quality of the activity. We also demonstrate that MTurk can be used to generate game content that are essential in facilitating enjoyable gameplay.

RELATED WORK

We first review various data mining efforts for engineering design, followed by discussion of using gamification and crowdsourcing for data collection.

[1] http://www.mturk.com/

Data Mining for Engineering Design

Various data mining efforts have been performed. Most related to our current work is using text mining or classification to support design. Much of the work involves mining retail e-commerce data [9-12]. Types of information analyzed typically involve product catalogue data [11] or customer reviews [9,12]. The information can be used to benchmark competitive products or identify customer preferences that can assist in developing new products. Examples of other text corpora mined include biology textbooks [13] or patent documents [14] to support design-by-analogy, design reports to identify design requirements [15] or product knowledge [16,17], and bug reports to identify design rationale [18].

Most of the work described above aims to extract knowledge from documentation of design solutions. CAD models, product catalogues, and patent documents all describe the final artifacts produced from design processes. Such information unfortunately does not give much insight into the design processes and how those solutions were produced.

Researchers have proposed mining resources such as product requirements documents created in the early stages of design to extract high-level design information [8,12]. However, such documents are hard to obtain because they contain sensitive information related to intellectual properties. Therefore, we need to consider other knowledge acquisition techniques.

Gamification and Crowdsourcing for Data Collection

Gamification uses game elements in non-game contexts [19,20], such as in design [21], marketing [22] and educations [23]. Based on this concept, Von Ahn and Dabbish introduced a technique called Games-with-a-Purpose (GWAP) [24]. In addition to providing a fun game experience to its players, an important motivation of GWAP is to leverage large-scale human intelligence and solve challenging problems for machines. Many successful GWAP systems have been deployed, e.g., to collect tags for music [25] or images [26] and to identify objects in images [27]. For instance, the work of [26] was used for Google Image Labeler[2], which helped Google acquire tags for their cached images. For all these systems, the efforts focus on large-scale labeling of web content.

Similar to our purpose, gamification has been used to collect metadata for creating Semantic Web [28,29]. In addition, Arlitt et al. [30] developed a game called "Biology Phenomenon Categorizer", which enables collection of assertions about a given biological phenomenon. By playing the game, players contribute toward collecting a set of specific relations found in biological phenomena, which could be used to populate a knowledge base for biologically inspired design. Arlitt et al.'s approach is novel in that relation knowledge can be collected, rather than simple labels as with the most GWAP systems.

Our game, "Who is the Pretender?" is a multi-player game played by more than three players. This contrasts from most

[2] The system is now offline.

existing gamification methods that involve two players. Also, we focus on collection of high-level design information that may be difficult to achieve with games based on labeling tasks.

Amazon Mechanical Turk (MTurk) has been widely used to source paid micro-tasks that require human intelligence, e.g., to collect user experience feedbacks [31], process documents [32] or videos [33], find relevant information to questions [34], or even solve design problems [35]. For our work, we use MTurk to generate game content, which will be explained in a later section.

GAME MECHANICS

The game was originally inspired from a game played in a popular Chinese TV show, translated as "Who is the Spy?" in English[3]. Although the overall game mechanics are similar to the game featured in the TV show, we made several modifications for the purpose of recording data, controlling the game content to always involve designed objects, and refining the game content using MTurk.

The following subsections describe the game rules, as well as the game dynamics that enable us to collect high-level design information.

Game Rules

1. The game is played amongst six players as a default configuration. However, we believe that the exact player number, n, can be flexible ($n \geq 3$).
2. At the start of the game, each player receives an object card. The object card contains an object name. Five players, named as "masters," receive identical cards and one player, named as a "pretender," receives a different card. Players must hide the card from each other and therefore do not know whether they are a master or the pretender.
3. The game is played in one or more rounds. At each round, each player must provide a description of his/her object. The description could be of any length, but cannot contain the name of the object. Each player writes down a description secretly and then shows the description at the same time. Players are not allowed to repeat the same description used in previous rounds including their own or others'.
4. After seeing each other's descriptions, players try to guess whether they are one of the masters, or the pretender. Each player then votes who he/she thinks is the pretender. The player who receives the most votes is eliminated from the group. If there is a tie, the next round proceeds without anyone being eliminated. The goal of all players is to survive as long as they can.
5. If the pretender is eliminated (identified by a moderator), the remaining masters win and collect a point each. If one of the masters is eliminated, the game continues into the next round. If the pretender survives until he/she is one of the two remaining players in the group, the pretender wins and collects four points (in general, n-2 points).
6. During the gameplay, players are not allowed to communicate with each other besides sharing their written descriptions.

Game Features

Several interesting game dynamics are present in this game. Most importantly, the game discourages players from providing specific descriptions of objects, especially in the early rounds. If a player is the pretender and provides a description that is likely to be only specific to his/her object, the chance of other players recognizing that player as the pretender would increase. For the masters, they also would not want to provide specific descriptions that reveal their objects. If the pretender survives that round, he/she would have likely recognized what the masters' object is and could provide descriptions relevant to that object in subsequent rounds, successfully pretending to be a master. This is the key dynamic that enables elicitation of high-level descriptions of objects.

Another important feature for data collection is that the game discourages players from providing descriptions that are too vague or irrelevant to their objects. Such behavior would lead other players to suspect that the player is the pretender and eliminate the player. Hence, most of the players' descriptions would be relevant to the objects assigned to them. One exception is the descriptions provided by the pretender, who could try to describe the masters' object once the player learns that he/she is the pretender. For this case, we can simply ignore the data provided by the pretender, which would only be a minor portion ($1/n$) of the entire data collected from the game.

As suggested above, the game features both competitive and cooperative elements, which we have not seen in previous game systems used for knowledge acquisition. The competitive element comes from the fact that the ultimate goal of each player, whether a pretender or a master, is to survive the elimination. The cooperative element comes from the fact that the masters as a group must implicitly agree not to provide descriptions that are too specific. If one master decides to give a very specific description to maximize his/her chance of survival, the chance of the group winning the game would be reduced because the pretender would identify the masters' object.

GENERATING GAME CONTENT USING AMAZON MECHANICAL TURK

An important requirement for this game is that the object pairs used as the masters' and the pretender's objects should have some similarities. If objects are dissimilar, the chance of masters identifying the pretender and ending the game in early rounds is very high. If objects are similar, there is a greater chance that the descriptions given by the pretender and masters are relevant to both objects, making it harder for the group to identify the pretender.

Therefore, in contrast to other gamification efforts that aim to collect knowledge about existing web content, we need a

[3] http://www.hunantv.com/v/2012/happycamp/

method that can generate a collection of refined game content. The method should also be scalable to support future gameplay modes, e.g. if the game is made available to be played online.

To generate game content, we leveraged crowdsourcing, in particular Amazon Mechanical Turk (MTurk), to identify similarity between pairs of objects.

Object Dictionary

We first constructed a dictionary of objects to be used in the game. For the actual knowledge acquisition task in the future, this dictionary would consist of design objects that we want to collect knowledge about, and therefore would be manually created.

For the current study, we selected 50 objects. The main heuristic used for listing these objects was based on Functional Basis [36]. Functional Basis contains taxonomy of verbs used to model functions of electro-mechanical products. We tried to pick objects such that they would be evenly distributed when classified by the top-level function categories of Functional Basis. Table 1 shows the objects classified by function categories.

Table 1. Object dictionary used for the game and corresponding function categories for objects.

Function Category	Object Dictionary
Branch (Separate, Remove, etc.)	Scissors, Knife, Axe, Saw, Drill (tool)
Channel A (Transfer, Transmit, etc.)	Gear (toothed wheel), Linkage (mechanical), Chain drive, Wheel, Axle
Channel B (Translate, Rotate, etc.)	Electric motor, Crank, Spinning top, Door knob, Propeller
Connect (Couple, Join, Link, etc.)	Screws, Nail (fastener), Knot (tied rope), Staple (fastener), Zipper
Control Magnitude A (Actuate, Increase, etc.)	Crowbar, Bottle opener, Crane (machine), Pulley, Jack (lifting device)
Control Magnitude B (Decrease, Prevent, etc.)	Bubble wrap, Helmet, Goggles, Shin guard, Bumper (automobile)
Convert	Windmill, Turbine, Flywheel, Battery, Fuel cell
Provision (Store, Contain, Collect, etc.)	Vase, Mug, Crate (shipping container), Cabinet, Computer case
Signal (Display, Indicate, etc.)	Billboard, Banner, Traffic sign, Traffic light, Lighthouse
Support (Stabilize, Secure, Position)	Cane (walking stick), Crutch (mobility aid), Bracket (supporting structure), Shelf (storage), Stool (seat)

Task Design

We collected similarity ratings for all possible object pairs in the object dictionary using MTurk. The task (named as Human Intelligent Task or HIT on the MTurk website) involved MTurk workers to assign functional similarity scores for pairs of objects based on a 5-point Likert scale. Figure 2 shows a screenshot of a HIT posted on MTurk. We wanted to measure the functional similarity between objects because such pairs would lead to longer-lasting games if the players were to describe the objects with high-level design information, e.g., functions. Longer-lasting games would then allow us to collect a greater number of high-level design descriptions of the objects used in the game.

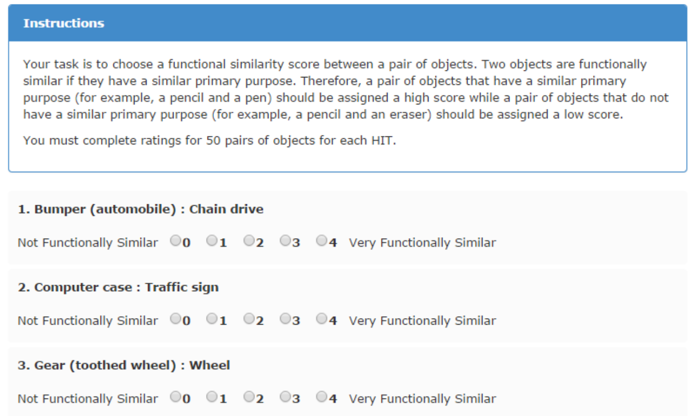

Figure 2. Screenshot of MTurk Human Intelligence Task requested to generate game content.

We collected data in two batches. The first batch involved all possible comparisons based on a group of 25 objects. This requires 300 unique comparisons ((25x25-25)/2). The reason for first collecting data on a small set was that we wanted to test whether MTurk was appropriate for collecting this kind of data. The first batch involved creating six HITs, each HIT consisting of 50 comparisons, with ten assignments per HIT (i.e., ten observations for each comparison). For the final similarity score assigned to each object pair, we calculated the mean of the ten observations. We paid $0.75 per HIT.

After confirming the quality of the collected data (explained in the forthcoming sections), we ran a second batch involving all possible comparisons based on 25 additional objects, including the new comparisons between the objects in the first batch and the objects in the second batch. This required 925 new unique comparisons ((50x50-50)/2–300). We randomly added 25 comparisons from the first batch to make the total comparisons in the second batch divisible by 50. Therefore, the second batch involved 19 HITs, each HIT consisting of 50 comparisons (950 comparisons in total), requesting ten assignments per HIT, and paying $0.75 per HIT.

Expert Evaluations

We evaluated the quality of the first data batch by comparing them to the data collected from experts. We recruited the following group of experts: Two industrial designers, both working for design consulting firms for 5+ years; and two engineers, both with MASc in mechanical engineering, one with 1+ year of experience as a Medical Device Engineer, and another with 3+ years of experience in Product Development for a large retailer and currently pursuing a PhD in mechanical engineering.

We gave the experts the same task instructions as the MTurk tasks, but on a screened set of 50 comparisons. The authors picked 25 of these pairs formed with objects within the same function categories (Table 1), while the other 25 pairs were randomly picked with objects from different function categories. Simple random sampling would have resulted in data with several very low or zero similarity ratings, which would have made identifying correlations between the expert data and the MTurk data difficult.

Analysis of Similarity Data

First, we calculated the intra-class correlation amongst the four experts to ensure that functional similarities could be objectively measured. We found $ICC(2,4)=.85$, $p<.005$ and confirmed agreement within expert ratings.

We then compared the expert ratings to the MTurk ratings for the same set of 50 object pairs. Between the expert group and the MTurk group, we found $ICC(2,2)=.80$, $p<.05$ (Figure 3). Therefore, we confirmed that MTurk is a reliable method to identify function similarities between pairs of objects.

Figure 4 shows the histogram of similarity data. A large number of object pairs resulted in very low similarity scores. The results indicate that randomly picking any two objects for the game likely would not work because players will mostly receive two very dissimilar objects. Instead, we can use crowdsourcing to identify a few pairs of functionally similar objects that could be used for the game. Although the number of object pairs might be low, the scalability of crowdsourcing should enable us to generate the necessary game content.

EXPERIMENTAL METHOD

After generating the game content, a user study was conducted to evaluate the feasibility and benefits of the game in collecting high-level design descriptions.

Participants

We recruited four groups (A, B, C, D) of six participants, totaling 24 participants for the user study. Participants' ages ranged from 20 to 50. Twenty males and four females participated. Educational backgrounds were biased towards computer science. One of the co-authors acted as the game moderator. Figure 5 shows a scene of the user study.

Experimental Setup

We used a within-subjects design with two independent variables. The first variable was the degree of similarity of object pairs used. We created two conditions based on the MTurk similarity scores, named as "High Similarity" and "Low Similarity." We sampled 10 objects pairs each for the High Similarity and Low Similarity groups based on the following criteria. We first ranked all object pairs by their similarity scores. For the High Similarity group, we picked the top-ten object pairs with similarity scores higher than two (the median of possible similarity scores), subject to a constraint that all objects must be unique. For the Low Similarity group, we randomly sampled object pairs with similarity scores lower than two, again subject to a constraint that the objects must be unique from previously picked objects. This procedure resulted in 20 object pairs with all 40 objects being unique. The object pairs were further randomly divided into four sets (two four-pair sets and two six-pair sets), subject to a constraint that each set contained an equal number of High and Low Similarity pairs. Table 2 shows the sets of object pairs selected.

Table 3 shows object pair sets given to each group for the Listing and Game activities. Because each set contains pairs with both high and low similarities, each group performed activities with both similarity conditions. However, the objects participants used for each activity were different. Participants were not aware of the similarity conditions.

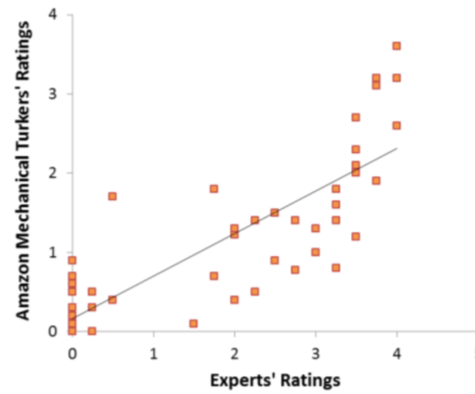

Figure 3. Correlation between the MTurkers' ratings and the experts' ratings.

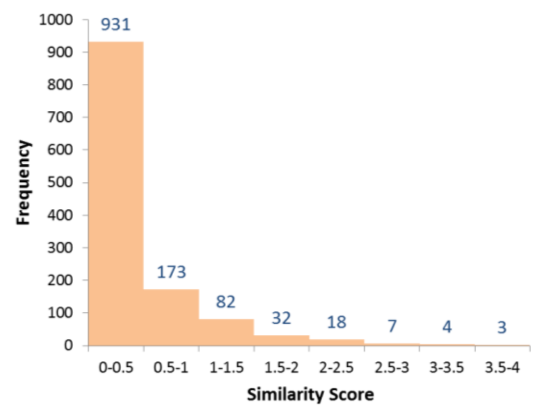

Figure 4. Distribution of similarity scores assigned by MTurk workers to object pairs.

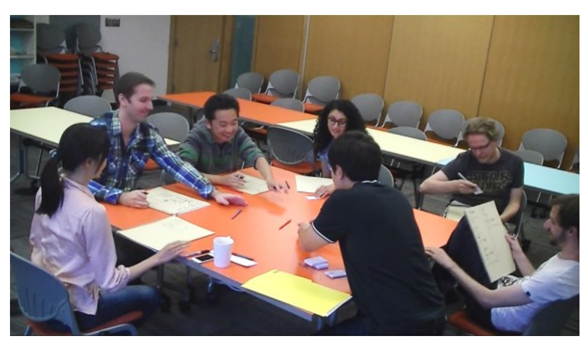

Figure 5. Participants of Group C playing the game.

Table 2. Object pair sets created.

	Object Pairs		Similarity
Set 1	Bracket (supporting structure)	Linkage (mechanical)	Low
	Vase	Mug	High
	Bubble wrap	Bumper (automobile)	Low
	Scissors	Knife	High
Set 2	Screw	Nail (fastener)	High
	Windmill	Wheel	Low
	Knot (tied rope)	Crowbar	Low
	Helmet	Shin guard	High
Set 3	Billboard	Traffic sign	High
	Zipper	Chain drive	Low
	Flywheel	Gear (toothed wheel)	High
	Electric motor	Door knob	Low
	Cane (walking stick)	Crutch (mobility aid)	High
	Crate (shipping container)	Stool (seat)	Low
Set 4	Jack (lifting device)	Turbine	Low
	Axe	Saw	High
	Battery	Fuel cell	High
	Goggles	Banner	Low
	Cabinet	Computer case	High
	Crank	Propeller	Low

Table 3. Assignment of object pair sets to Listing/Game conditions for each participant group.

Group	Sets used for Listing	Sets used for Game
A	Set 1	Set 2
B	Set 2	Set 1
C	Set 3	Set 4
D	Set 4	Set 3

The second experimental variable was the activities that participants performed. For all groups, participants were first asked to simply list three descriptions of given objects. We name this condition as "Listing." After the Listing activity, participants played multiple games, each with a different object pair. We name this condition as "Game." We chose the requirement of three descriptions for the Listing activity because our pilot studies revealed that on average the game resulted in three rounds, which is equivalent to three descriptions provided per participant for each object.

Each activity consisted of the number of Listing/Game tasks corresponding to the number of object pairs used. For example, Groups A and B played the game four times while Groups C and D played the game six times. This was due to the time constraints for the first two groups. The first groups spent about 60 minutes and the second group spent about 90 minutes to finish the activities. There was no specific time limit for each of the Game and Listing task.

The order of the object pairs within each set was initially randomized and this same order was used for each group/activity condition. Table 3 shows the order. For both activities, the first object in the pair was used as the masters' object cards (x5) while the second object was used as the pretender's object card (x1). The moderator randomly distributed the object cards to each player for each activity.

Participants were informed that the purpose of both the Game and Listing activities was to collect descriptions of common objects. However, they were not given any guidance on the type of descriptions that they should provide and they were not asked to provide high-level design descriptions.

Materials

We printed each object name on a single heavyweight business card, as shown in Figure 6. Participants received one of these cards at the start of each game. We did not use images because they would bias participants towards describing visual or geometric characteristics of objects. We also gave each participant a sketchbook and markers to write down object descriptions and votes during the game.

Figure 6. Object cards used for the study.

Survey Questions

After participants performed both activities, they were asked to complete the survey questions shown in Table 4. The first seven questions (1-7) asked participants to rate their agreements to the questions for each activity, in a 5-point Likert scale with zero and four indicating strong disagreement and strong agreement, respectively. The next three questions (8-10) repeated the first three questions, but asking participants to rate which activity they preferred, in a 5-point Likert scale with zero and four indicating their preferences towards Listing or Game, respectively. The purpose of this data collection is to validate that the game is enjoyable, fun, and engaging so that players would have intrinsic motivations to play the game, which is an essential requirement for successful gamification.

1/8)	I enjoyed doing the activity
2/9)	The activity makes describing objects fun
3/10)	I was fully engaged in the activity
4)	Describing objects was frustrating
5)	Describing objects was difficult
6)	I had to work hard to describe objects
7)	I felt rushed when describing objects

Table 4. Survey questions asked to participants.

Data Analysis

We used the FBS coding scheme [7] to categorize the object descriptions provided by participants. The coding scheme has been used to analyze a number of design studies [6]. We used four elements in the FBS for our analysis: requirement, function, behavior, and structure. The definitions and example descriptions of each element was presented in Introduction (Figure 1). We also included a category, "others", to classify descriptions that did not belong to the FBS coding scheme [7], which consisted of 4% of the data and was not considered in the analysis.

One of the authors classified all the descriptions collected and a second independent rater, familiar with the FBS coding, classified 30% of the descriptions to check inter-rater reliability. We obtained 81.5% of agreement with Cohen's kappa of 0.73, which is considered as "substantial" agreement [37]. Hence, we considered the author's classification data reliable and used them for the analysis.

We hypothesized that the Game condition would produce a higher percentage of high-level descriptions, e.g., requirement, function, and behavior, while the Listing condition would produce a higher percentage of low-level descriptions, e.g., structure.

We also hypothesized that the descriptions produced in the Game condition would have a greater lexical diversity than the Listing condition, because players have a greater incentive to be creative with their descriptions. The lexical diversity can be defined as [38]:

$$\mathit{Lexical\ diversity} = \frac{\mathit{Number\ of\ unique\ set\ of\ words}}{\mathit{Number\ of\ words\ used}}$$

Hence, the score ranges from zero to one, with a higher score indicating a more lexically diverse corpus.

Finally, we recorded the average number of rounds taken to finish each game played by participants. We assumed that the greater number of rounds corresponds to more challenging gameplay, i.e., the games that lasted longer indicate that finding the pretender was more challenging. In addition, the games that lasted longer would help us collect a greater number of object descriptions. We hypothesized that the High Similarity object pairs would result in a greater number of rounds played than the Low Similarity object pairs.

RESULTS

We present results that demonstrate the benefits of: 1) the gamified method for collecting high-level design descriptions and 2) generating game content via MTurk. We also present an example of gameplay and object descriptions collected.

Benefits of Gamification

Figure 7 compares the percentages of object descriptions belonging to each of the FBS information category between the Listing and Game conditions. We highlight two significant differences found in two of the four information categories. We found that 26.6% and 16.6% of object descriptions from the Game and Listing conditions, respectively, consisted of requirement information. This difference was statistically significant (Fisher's exact test, p=.0032). We also found that 27.8% and 38.5% of object descriptions from the Game and Listing conditions, respectively, consisted of structure information. Again, this difference was statistically significant (Fisher's exact test, p=.0070). These results confirm our hypotheses that object descriptions produced from the Game activity contain more high-level design information and less low-level design information than object descriptions produced from the Listing activity. Interestingly, the significant difference in high-level descriptions was only found for requirement information, which is the most abstract level of design information considered in our analysis. The results indicate how the game motivated participants to use more abstract information to describe their objects.

For the lexical diversity, no statistically significant difference was found (Listing: M=.85, SD=.10; Game: M=.89, SD=.11; t(38)=.89, p=.38). Figure 8 shows the results.

We also compared the survey responses between the activity conditions using the Mann-Whitney U test. We found that participants found the Game activity more enjoyable, fun, and engaging than the Listing activity (Q1: U=510, p=.0000; Q2: U=531, p=.0000; Q3: U=491, p=.0000). For other rated agreement questions (Q4-Q7), we did not find any statistically significant difference in their answers. These results are shown in Figure 9. For the comparison questions (Q8-10), we observed that participants in general preferred the Game activity than the Listing activity, shown in Figure 10.

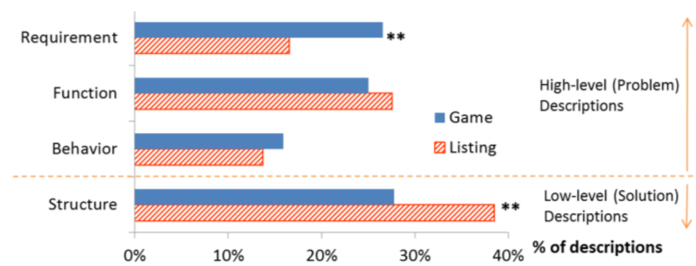

Figure 7. Comparison of percentages of object descriptions belonging to each of the FBS information category (p<.01).**

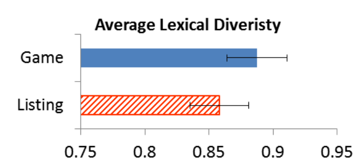

Figure 8. Comparison of average lexical diversity scores.

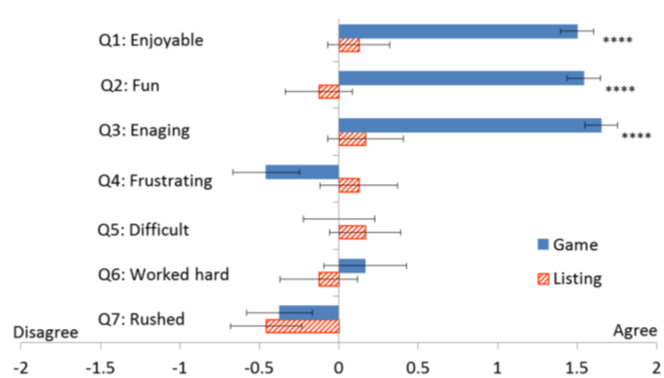

Figure 9. Comparison of responses to survey questions (**p<.0001).**

Figure 10. Participant responses to survey questions 8-10, comparing their preferences between Game and Listing.

Benefits of Game Content Generation with MTurk

We found that the games played using object pairs with high similarity lasted for a greater number of rounds than the games played using object pairs with low similarity. The difference was statistically significant (High Similarity: M=3.2, SD=1.75; Low Similarity: M=1.6, SD=.97; $t(18)=2.53$, p=.021). Figure 11 shows the results. Again, the high/low similarity pairs were sampled based on the MTurk ratings, which demonstrates the benefit and feasibility of generating game content through crowdsourcing.

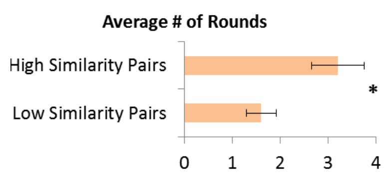

Figure 11. Comparison of average number of rounds played between low similarity and high similarity pairs (*p<.05).

Table 5 shows an example of a game played by the Group C participants and object descriptions collected. Each column corresponds to the players involved and each row corresponds to the game rounds played. Cells contain object descriptions given by each player at each round. A crossed-out (X) cell indicates that the player was eliminated at the end of the previous round. An interesting outcome for this game is that the pretender (Player 4) did not even know that he was the pretender and eventually provided the distinguishing structure information, "Thin," that led to his elimination in the sixth round. This scenario would only be likely if a pair of highly similar objects was used. Also, this example demonstrates a scenario in which a player (Player 6) providing a description not relevant to the object and being eliminated. This highlights how the game encourages players to provide descriptions relevant to their objects.

DISCUSISON AND FUTURE WORK

The results indicate that the game provides an enjoyable, fun, and engaging experience for participants to provide object descriptions than simply listing them. More importantly, the descriptions collected through the game consist of a greater proportion of requirement information and a smaller proportion of structure information than the listing task. Hence, we demonstrated the potential of gamification in collecting high-level design information.

Several techniques have been developed in the artificial intelligence community to extract common knowledge from various sources [39-43]. Typically, these techniques use self-supervision [41] or crowdsourced input [42] to classify the knowledge extracted from the sources as true or false. Our method is advantageous because the object descriptions will be true in most cases. Therefore, the data collected could be used as is or as training data for the machine learning techniques used in knowledge extraction.

As for an immediate application of the data collected from the game, we developed a proof-of-concept information retrieval system with the objects indexed by their high-level design descriptions. The preliminary usage test of this system shows that designers could search for objects using requirement, function, or behavior descriptions as keywords. For example, the query of "cut" returns both "axe" and "scissors". The returned objects could be used as stimuli for concept generation.

To make better use of the data collected, we need to develop techniques to process and classify the data. First, because we did not put any restriction on the linguistic structure of object descriptions, we may have to treat the data as a "bag-of-words." In the future, we may require players to provide descriptions in full sentences, which could be parsed to aid extraction of syntactic and semantic information. In addition, we manually classified the object descriptions according to the FBS framework for the current work, which is not ideal for large-scale data collection. We plan to investigate machine learning techniques that can automatically classify the collected data into the FBS categories. Features used to classify the data could come from the different elements of the game. For example, if players are more likely to provide requirement descriptions at the beginning of the game, the number of rounds at which a particular description was collected could be an important feature in the classification model.

Round	**Player 1** **Object: Axe**	**Player 2** **Object: Axe**	**Player 3** **Object: Axe**	**Player 4*** **Object: Saw**	**Player 5** **Object: Axe**	**Player 6** **Object: Axe**
1.	"Tool"	"Tool"	"Tool"	"Tool"	"Sharp blade"	"Honest"
2. (tie)	"Useful"	"Metal"	"Hand-held"	"Metal"	"Swung"	X
3.	"Cuts"	"Ancient"	"Tedious"	"Cut"	"Chops"	
4.	"Top heavy"	X	"Digging tool"	"Wood"	"Used by firemen"	
5. (tie)	"Handle"		X	"Require strength"	"Used in battle"	
6.	"Used by loggers"			**"Thin"**	"Single or double-edged"	
Results	**Winner**			X	**Winner**	

Table 5. Example gameplay and object descriptions collected. *Player 4 is the pretender. "X" indicates a player being eliminated. Highlighted in gray are low-level (structure) descriptions.

In terms of the game content generation, we could explore applying techniques from natural language processing to replace or augment the crowdsourcing approach. For instance, we could apply Latent Semantic Analysis [43] or other vector semantics methods developed for design knowledge retrieval, e.g., [44], to compute functional similarity between a pair of objects. We would need to refine these methods and validate them against the human assessment of functional similarity.

Another important future work will involve implementing the game on social network or smart phone platforms. One challenge is that the game requires multiple players and the game state is dependent on the action of players [45]. In other words, the game can only proceed if each player contributes an action. In addition, if the game is played online, some of the positive experience that participants reported as coming from face-to-face social interactions may be lost. We believe that implementing the game on an existing social network such as Facebook or Twitter would be a good strategy. Playing the game with familiar people within a social network would augment existing social interactions as part of the gameplay. In addition, each game in our experiment lasted only about 10 minutes, so this short timeframe is ideal for the game to be played online. Regardless of the implementation method, we need further evaluation to ensure that the game can attract and retain players over time.

We also plan to experiment with altering the game rules or format to collect different distributions of design descriptions based on the FBS categorization. For example, could we adjust or introduce rules to further discourage describing structures, but encourage providing descriptions of function and behavior information? In addition, one limitation of the game is that players must be familiar with the objects used. We could perhaps create different "themes" of the game that involve objects from a specific domain, e.g., mechanical components, and only players who are familiar with the domain could be recruited to play the game. This strategy would be possible once the game is featured online and players could be recruited from a large group of people.

CONCLUSIONS

The current work explored novel application of a game, "Who is the Pretender?" to collect high-level design descriptions of objects. The game features complex yet engaging game mechanics, which elicits more high-level problem-oriented descriptions, e.g., requirements, and less low-level solution-oriented descriptions, e.g., structure, of designed objects than simply asking people to list object descriptions. We also showed an application of crowdsourcing, in generating game content. Our work demonstrated the feasibility and effectiveness of gamification to acquire high-level design information. Such information would be useful in constructing design ontologies, supporting the retrieval of design solutions based on their functions and requirements, and eventually developing knowledge-based CAD systems in the future.

REFERENCES

[1] Gero, J. S. (1985). Knowledge Engineering in Computer-Aided Design: Proceedings of the IFIP WG 5.2 Working Conference on Knowledge Engineering in Computer-Aided Design. Elsevier Science Inc.

[2] Tomiyama, T. (2007). Intelligent computer-aided design systems: Past 20 years and future 20 years. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 21(01), 27-29.

[3] Gero, J. (1990). Design prototypes: A knowledge representation schema for design. AI Mag. 11(4), 26-36.

[4] Umeda, Y., Takeda, H., Tomiyama, T., & Yoshikawa, H. (1990). Function, behaviour, and structure. Applications of artificial intelligence in engineering V, 1, 177-194.

[5] Chandrasekaran, B., Goel, A. K., & Iwasaki, Y. (1993). Functional representation as design rationale. Computer, 26(1), 48-56.

[6] Gero, J.S., Kannengiesser, U., and Pourmohamadi, M. (2012) Commonalities across designing: Empirical results. DCC'12, 285-302.

[7] Kan, J.W.T., and Gero, J.S. (2009). Using the FBS ontology to capture semantic design information in design protocol studies. In J. McDonnell, and P. Lloyd (eds), About: Designing. Analysing Design Meeting, Taylor & Francis, pp. 213-229.

[8] Braha, D., eds. (2002) Data Mining for Design and Manufacturing. Springer.

[9] Hu, M., and Liu, B. (2004). Mining and summarizing customer reviews. In Proc. ACM KDD, 168-177.

[10] Kohavi, R., Mason, L., Parekh, R., and Zheng, Z. (2004). Lessons and challenges from mining retail e-commerce data. Machine Learning 57, 1-2, 83-113.

[11] Labský, M., Praks, P., Svátek, V., and Sváb, O. (2005). Multimedia information extraction from HTML product catalogues. In K. Richta, V. Snaesel, J. Pokorny (eds), DATESO, pp. 84-93.

[12] Rai, R. (2012). Identifying key product attributes and their importance levels from online customer reviews. In ASME 2012 IDETC/CIE, Chicago, IL, 12-15 August.

[13] Shu, L.H., Ueda, K., Chiu, I., and Cheong, H. (2011). Biologically inspired design. CIRP Annals - Manufacturing Technology 60, 2, 673-693.

[14] Verhaegen, P.A., D'hondt, J., Vandevenne, D., Dewulf, S., and Duflou, J.R. (2011). Identifying candidates for design-by-analogy. Computers in Industry 62, 4, 446-459.

[15] Zeng, Y. (2008). Recursive object model (ROM) - Modelling of linguistic information in engineering design. Computers in Industry, 59(6), 612-625.

[16] Li, Z., and Ramani, K. (2007). Ontology-based design information extraction and retrieval. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing 21, 02, 137-154.

[17] Li, Z., Yang, M. C., & Ramani, K. (2009). A methodology for engineering ontology acquisition and validation. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 23(1), 37-51.

[18] Rogers, B., Qiao, Y., James, G., Mathur, T., and Burge, J. (2014). Using text mining techniques to extract rationale from existing documentation. DCC'14.

[19] Webster, J. (1988). Making computer tasks at work more playful: Implications for systems analysts and designers. Proc. ACM SIGCPR Conf. Management of Information Systems Personnel.

[20] Deterding, S., and Dixon, D. (2011). From game design elements to gamefulness: Defining "gamification". ACM MindTrek, 9-15.

[21] Tsang, M., Fitzmaurice, G., Kurtenbach, G., and Khan, A. (2003). Game-like navigation and responsiveness in non-game applications. Commun. ACM 46, 7, 57–61.

[22] Zichermann, G. and Linder, J. (2010). Game-Based Marketing: Inspire Customer Loyalty Through Rewards, Challenges, and Contests. Wiley.

[23] Gee, J.P. (2003). What Video Games Have to Teach Us About Learning and Literacy. Palgrave Macmillan: New York, NY, USA.

[24] Von Ahn, L., and Dabbish, L. (2008). Designing games with a purpose. Commun. ACM, 51, 8, 58-67.

[25] Law, E., and Von Ahn, L. (2009). Input-agreement: a new mechanism for collecting data using human computation games. In Proc. ACM CHI.

[26] Von Ahn, L., and Dabbish, L. (2004). Labeling images with a computer game. In Proc. ACM CHI.

[27] Von Ahn, L., Liu, R., and Blum, M. (2006). Peekaboom: A Game for locating objects in images. In Proc. ACM CHI, 55–64.

[28] Simko, J., Tvarozek, T., and Bielikova, M. (2013). Semantics discovery via human computation games. Int. J. on Semantic Web & Information Systems 7, 286, 23-45.

[29] Siorpaes, K. and Hepp, M. (2008). Games with a purpose for the semantic Web. IEEE Intelligent Systems 23, 3, 50-60.

[30] Arlitt, R., Berthelsdorf, F., Immel, S., and Stone, R. (2014). Using human computation to assist biologically inspired design: Evaluating a Game-With-A-Purpose. In ASME 2014 IDETC/CIE, Buffalo, NY, 17-20 August.

[31] Kittur, A., Chi, E., and Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In Proc. ACM CHI.

[32] Bernstein, M.S., Little, G., Miller, R., Hartmann, B., Ackerman, M., Karger, D., Crowell, D., and Panovich, K. (2010). Soylent: A word processor with a crowd inside. In Proc. ACM UIST.

[33] Bernstein, M.S., Brandt, J., Miller, R.C., and Karger, D.R. (2011) Crowds in two seconds: Enabling realtime crowd-powered interfaces. In Proc. ACM UIST.

[34] Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Tatarowicz, A., White, B., White, S., and Yeh, T. (2010) VizWiz: Nearly real-time answers to visual questions. In Proc. ACM UIST.

[35] Yu, L., Kittur, A., and Kraut, R.E. (2014). Distributed analogical idea generation: Inventing with crowds. In Proc. ACM CHI.

[36] Hirtz, J., Stone, R.B., Szykman, S., McAdams, D.A., and Wood, K.L. (2002). A functional basis for engineering design: Reconciling and evolving previous efforts. NIST Technical Note 1447.

[37] Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics 33, 1, 159-174.

[38] Bird, S., Klein, E., and Loper, E. (2014). Chapter 2: Assessing Text Corpora and Lexical Resources. Natural Language Processing with Python. http://www.nltk.org/book/ch02.html. 19 Sep. 2014.

[39] Agichtein, E., & Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. In the Fifth ACM Conference on Digital Libraries (pp. 85-94). ACM.

[40] Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H., & Shadbolt, N. R. (2003). Automatic ontology-based knowledge extraction from web documents. Intelligent Systems, IEEE, 18(1), 14-21.

[41] Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007). Open information extraction for the web. In IJCAI (Vol. 7, pp. 2670-2676).

[42] Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E. R., & Mitchell, T. M. (2010). Toward an Architecture for Never-Ending Language Learning. In AAAI (Vol. 5, p. 3).

[43] Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. JASIS, 41(6), 391-407.

[44] Murphy, J., Fu, K., Otto, K., Yang, M., Jensen, D., & Wood, K. (2014). Facilitating design-by-analogy: Development of a complete functional vocabulary and functional vector approach to analogical search. In Proc. of ASME 2014 IDETC/CIE.

[45] Bjork, S. (2010). Principles and patterns of social games. Where is the difference compared to other games? GCC Games Convention Online.