# GameNet and GameSage:
# Videogame Discovery as Design Insight

## James Ryan, Eric Kaltman, Timothy Hong, Katherine Isbister, Michael Mateas, and Noah Wardrip-Fruin

Center for Games and Playable Media
University of California, Santa Cruz
`{jor, ekaltman, michaelm, nwf}@soe.ucsc.edu`
`{thon6, katherine.isbister}@ucsc.edu`

## ABSTRACT

The immense proliferation of videogames over the course of recent decades has yielded a discoverability problem that has largely been unaddressed. Though this problem affects all videogame stakeholders, we limit our concerns herein to the particular context of game designers seeking prior work that could inform their own ideas or works in progress. Specifically, we present a tool suite that solicits text about a user's idea for a game to generate an explorable listing of the existing games most related to that abstract idea. From a study in which 182 game-design students used these tools to find games related to their own, we observe a demonstrated utility exceeding that of the current state of the art, which is the coordinated usage of assorted web resources. More broadly, this paper provides the first articulation of videogame discovery as an emerging application area.

## Keywords

videogame discovery, exploratory search, visualization, design support, tools, NLP

## INTRODUCTION

Attendant to the immense proliferation of videogames over the course of recent decades is a *discoverability problem* that has gone largely unaddressed. Though it launched only seven years ago, there are over 400,000 games available on the iTunes App Store and 376 new titles are being added each day (Pocket Gamer, 2015). Google Play features a comparable number of games (VentureBeat, 2015), and there remains a considerable array of titles published for consoles and personal computers. Regarding the latter, more games were released on the digital-distribution platform Steam (Moore, 2009) in the first twenty weeks of 2014 than in the entirety of 2013 (Rose, 2014). These staggering numbers point to a simple fact: there are so many games in existence that players, scholars, critics, and designers need dedicated tools if they are to successfully encounter titles related to their particular discovery needs. There are only a handful of such tools, however, and so a discoverability problem persists.

A major symptom of this problem is a strong *bandwagon effect* (Adler, 1985) that has arguably yielded a blockbuster-driven games industry. Because a lack of resources makes games largely undiscoverable, the best-marketed titles reach the largest audiences, which in turn persuades industry prime movers that only games like those *can* reach wide audiences, which causes more such games to be well-marketed, and so forth. At its most extreme, this feedback loop produces endless instantiations of the same formulaic series, and so we find that the *Call of Duty* franchise, for instance, has generated 26 titles in twelve years.

Of course, other media also have discoverability problems, but in games the problem is compounded by the high dimensionality of their ontologies as media artifacts. Games incorporate text, sound, moving images, and nontrivial interaction, each axis of which may subsume several areas of concern. The only way to alleviate this discoverability problem is to develop dedicated tools for game discovery, but such tools must index games according to their composition along *all* these dimensions, so that a user can operate over these concerns while exploring the system. We contend that this represents a harder task than building discovery systems for other media that are lower-dimensional in their ontological makeup, which means that we cannot simply adapt to this domain successful discovery tools that were built for other domains.

While all videogame stakeholders are susceptible to the discoverability problem we have so far outlined, let us consider the particular case of how it affects game designers. Like any designer, a person who makes games may seek out prior work that could inform his or her own process, or more specifically the development of a new design. But lacking tools that meet this need, especially ones that may facilitate the discovery of novel or historically notable games, we find game designers whose practices are clearly unaware of related games, and often of the rich history of the medium. This is arguably one respect in which videogames show their immaturity relative to more established media. In this paper, we will limit our focus to this design-centric use case for a game-discovery system as a *creativity support tool* (Shneiderman, 2007). Before doing so, however, we would like to further articulate what exactly we mean by 'game discovery'.

First, game discovery is not game recommendation. Recommender systems (sometimes just called *recommenders*) are often part of larger commercial applications such as online retailers (Lam and Riedl, 2004), and a prototypical task of these systems is product recommendation (Park et al., 2012). In contrast, *discovery tools* (also called *exploratory search systems*; Marchionini, 2006) promote user *learning* above user purchasing. While in the recommendation task there is often an explicit notion of the correctness or accuracy of a recommendation (Ricci et al., 2011), the parallel concern in discovery is the *usefulness* of an item. (Of course, this distinction has made evaluating discovery tools a much trickier issue; White et al., 2008.) Finally, while recommenders are susceptible to a *popularity bias* by which a small proportion of the item space is recommended exponentially more often (Jannach et al., 2013)—a phenomenon that may actually aggravate the bandwagon effect we have just outlined—discovery tools aim to provide diverse, serendipitous offerings (White and Roth, 2009; Race et al., 2012). Neither should game discovery be considered as merely game information retrieval, primarily for the reason that a game-discovery tool will index *games*, not information about them. Moreover, in game discovery, getting back results is not a resolution—the offerings provided by a discovery tool are meant to be analyzed and explored (White and Roth, 2009). Finally, game discovery should always be user-centric, while in information retrieval ancillary concerns (namely algorithmic nuances) tend to take center stage, a truth that bears out in the offline, batch-style evaluation methods that have prevailed in the field for decades (White et al., 2008).

So, let us distill what distinguishes discovery tools from related applications into a succinct specification for what a good game-discovery tool should do: it should index games along all dimensions of their high-dimensional ontologies, afford queries that may operate over

the same dimensions, and present diverse, serendipitous, explorable offerings.

In this paper, we present a system for videogame discovery that we hope meets these criteria; it is specifically one that may allow game designers to find related prior work that could inform the development of an idea for a game, or a game that is already in development. This system comprises two interconnected components: GameNet is a network of nearly 12,000 games that are linked to one another according to how related they are, and GameSage is an interface to GameNet that solicits a description of an idea for a game (or a game that already exists) and uses natural language processing (NLP) to generate an explorable listing in GameNet of the existing games most related to the user's idea. Influenced by the notion of *task-based evaluation* of exploratory search systems (Kraaij and Post, 2006), in which a system is evaluated for its adequacy in the natural context of its user task, we conducted a user experiment in which game-design students sought out games related to their own using GameNet and GameSage and also a baseline method in which they were permitted to use any resources available online. We chose this baseline method because we believe it represents the (lack of a) state of the art in game discovery today. The primary variables of interest to our analysis are the number of games discovered using both methods, the diversity of games discovered (*i.e.*, percentage that were unique), and the proportion of discovered games that were unfamiliar to users prior to the experiment. From these, we constructed three major hypotheses, which we return to below in the discussion of our results:

—**H1.** Participants will discover more related games using our tools.
—**H2.** Participants will discover a greater diversity of games using our tools.
—**H3.** Participants will discover a greater proportion of unfamiliar games using our tools.

Our contribution here is threefold: we present the first articulation of videogame discovery as an emerging application area; we describe a significant extension to what to our knowledge is only the second game-discovery system to have appeared in the literature; we carry out the first controlled user study of such a system. Further, and more broadly, we believe that the interaction method of the system we present—in which a fully formed description of an idea for an artifact is matched against a database of real artifacts—could be adapted to different media (*e.g.*, film, music, literature) or to altogether different domains (*e.g.*, a system that takes an abstract for a prospective paper and matches it against the existing literature). It is our hope that this work will encourage the development of subsequent tools that may help to alleviate the videogame discoverability problem that we have outlined above.

## RELATED WORK

Videogame discovery is an emerging application area for which very little work has yet been done. In what would appear to be the first published effort in this domain, Lee et al. (2015) present Vizmo, a discovery tool that indexes games by visual style and mood. Influenced by earlier work in constructing browsers for other media, the tool was built using a *faceted-metadata* approach. Specifically, Vizmo is underpinned by a database of games that have been manually annotated for their visual style and mood using a subset of a larger videogame metadata schema that was created by the same group (see Lee et al., 2014, 2015). Users can browse the games in this database by setting filters for different combinations of visual style and mood, and the results are displayed in a chronologically oriented chart. From an

expert evaluation conducted with nineteen game professionals, Lee et al. found Vizmo to be an aesthetically pleasing tool with potential use for game discovery along aesthetic or historical concerns (Lee et al., 2015). Though promising, Vizmo is an early prototype that currently houses only 604 titles, many of which are platform variants of the same game. That so few games are yet included is not surprising given that each must be manually annotated in order to be indexable by the tool. Until more games are added, it appears that Vizmo will not offer extensive practical use for game discovery (a point we return to later in discussing the results of our experiment).

If Vizmo may be thought of as taking a top-down approach to game discovery, one in which humans handcraft indexable representations of games (which a discovery tool may then operate over), our own recent work in game discovery takes a decidedly bottom-up approach. Specifically, we have submitted large quantities of text describing games to a bottom-up technique from statistical NLP called *latent semantic analysis* (Dumais, 2004). By this method, we semiautomatically derived indexable representations for nearly 12,000 games, which we used to build a suite of visualizations (Ryan et al., 2015) and game-discovery tools (Ryan et al., 2015a; Edwards et al., 2015), namely GameNet and GameSage, as well as a videogame recommender system (Ryan et al., 2015). Because we describe our model and our tools in detail below, here we will only discuss the apparent trade-offs between these two approaches. Employing a semiautomated method, we could quickly build full-fledged discovery tools comprising several thousand games. But relying on a statistical procedure to derive indexable game representations, we were left with tools that operate over rather opaque notions of what games are made of—as we discuss in Ryan et al. (2015b), our tools reason about games in terms of arcane statistical features of their textual descriptions. Vizmo, on the other hand, reasons over games purely in terms of human-crafted specifications of them. As such, it will always be clear how Vizmo indexes games, because its indexable representations are simply human annotations. (One perhaps unexpected advantage of *unsupervised* reasoning, however, is that it may reveal unexpected associations among games that counter problematic human presuppositions; Ryan et al., 2015b.) But while Vizmo's game indexing is perhaps more reliable and certainly more transparent, GameNet appears to offer reliable-enough indexing for twenty times as many games (Ryan et al., 2015b).

## GAMENET AND GAMESAGE

GameNet and GameSage are components of a videogame-discovery system that solicits text concerning a user's idea for a game and, using natural language processing, generates an interactive listing of the existing games most related to that abstract idea. Both tools are underpinned by a *latent semantic analysis* model trained on Wikipedia articles describing games; the tools are hosted online as freely available web apps. In this section, we briefly explain latent semantic analysis and the construction of the model underpinning these discovery tools before proceeding to describe the tools themselves. While we describe GameSage in depth for the first time, we only outline latent semantic analysis, our corpus, and GameNet, since we have already thoroughly detailed these concerns elsewhere (Ryan et al., 2015b).

## Latent Semantic Analysis

Latent semantic analysis (LSA) is an NLP technique by which words are attributed vectorial semantic representations according to their contextual distributions across a large corpus

Figure 1: The GameNet home page, along with an example entry.

of text (Dumais, 2004). With an LSA model, one can easily calculate how semantically related any of the corpus documents are by taking the cosine between their LSA vectors—this measure is called *cosine similarity*. In corpora in which each document pertains to a specific individual concept, such as a corpus of encyclopedia articles, these relatedness scores can reasonably be utilized as a measure of the relatedness of the *concepts themselves*. As we illustrate below, GameNet and GameSage crucially rely on this notion.

## Our LSA Model

The LSA model underpinning GameNet and GameSage was trained by submitting 11,829 Wikipedia articles about videogames to latent semantic analysis (Ryan et al., 2015b). Each of these articles described an individual game and, at the time of extraction (May 2014), was at least 250 words in length and not marked by Wikipedia as being a stub. Using cosine similarity, defined above, this model affords the direct calculation of how related any two of its 11,829 games are—this is what fuels GameNet and GameSage. In Ryan et al. (2015), we argue that this computation operationalizes an intuitive notion of how games may be alike, which we call *game relatedness*. Beyond the tools under consideration here, we have also built a suite of interactive visualizations of this model (Ryan et al., 2015).

## GameNet

GameNet is a tool for game discovery in the form of a network in which related games are linked; it houses entries for the 11,829 games that are included in the LSA model we have just described. Each game's entry includes links to GameNet entries for other games that are related to that game, as well as to gameplay videos and other informative sources found elsewhere on the web. Figure 1 shows excerpts from the GameNet homepage and its entry for the Nintendo Entertainment System game *Wall Street Kid* (1989).

## GameSage

GameSage, the core technical contribution of this paper, is an interface to GameNet that solicits free-text input describing an *idea* for a game (or alternatively a game that already exists) and lists the existing games that are most related to that idea. This is done by utilizing the notion in LSA of *folding in*, whereby a new document that was not used during model training is fitted with a representation in the semantic space derived by the model. By treating the user's input text (which specifies his or her game idea) as a corpus document (on par with the videogame Wikipedia articles used to train the videogame LSA model) and folding it into the model, the tool is able to derive an LSA vector for the idea. From here, GameSage determines which existing games (from among GameNet's 11,829 games) are most related to the game idea by using cosine similarity, just as GameNet already does in determining which existing games are most related to one another. GameSage was briefly introduced in a conference demonstration (Ryan et al., 2015a), but the current paper represents its formal
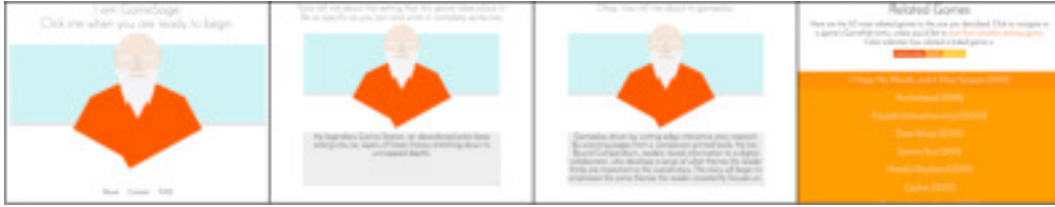
Figure 2: Excerpts from a GameSage session; the user describes the indie game *Ice-Bound* and obtains an explorable listing of related games.

presentation and evaluation, the latter of which we report below.

## *Usage*

At the GameSage home page, the system is personified as a sage character who proceeds to ask the user eight questions about his or her game idea, each pertaining to a particular aspect of the (prospective) game. The eight aspects that the sage asks about are *genre*, *setting*, *plot*, *gameplay*, *mechanics*, *aesthetics*, *unique aspects*, and *miscellaneous*. After the final question is answered, the system concatenates the responses and preprocesses this text using the same procedure that was enacted during corpus preprocessing (see Ryan et al., 2015b). From here, the preprocessed text is attributed an LSA vector by folding it into the model underpinning GameNet, whose 11,829 games are then ranked (using cosine similarity) according to how related they are to the user's idea. Finally, GameSage makes a request to GameNet to generate an entry for the game idea, to which the user is then taken. Here, the idea's most related and unrelated games are listed just as they are for all other GameNet entries. This entire process of folding in, from idea submission through generation of a GameNet entry, takes only a few seconds.

## Example Session

In this section, we will work through an example session with the tool in which a user describes the indie game *Ice-Bound* (Reed et al., 2014) using text from the game's official website; Figure 2 shows excerpts from this session. At the GameSage home page (shown in the first frame of Figure 2), the user clicks the sage character and proceeds to answer his eight questions (responses to the questions on *setting* and *gameplay* are shown in Figure 2), clicking the character to submit each:

**Genre**  First, tell me a bit about the genre of this game.

> *Ice-Bound was born from our love of interactive narratives and our frustration that hand-authored branching paths are still state of the art.*

**Setting**  Now tell me about the setting that this game takes place in.

> *the legendary Carina Station, an abandoned polar base sinking into ice, layers of frozen history stretching down to unmapped depths.*

**Plot**  Next, tell me about the narrative or plot of this game.

> *A nested-doll narrative that reveals more depth the more it's explored. A cunning publisher commissions an AI simulacrum of a long-dead author, to finish his famously incomplete masterpiece. The AI is neurologically identical to its human predecessor, but as a constructed sentience, has no human rights. And a curious book begins*

*appearing, stacked on street corners and left on busses: a real paper book, filled with glitched transmissions, contradictory drafts, distorted photos and vicious secrets. It's a book only one person was ever meant to see.*

**Gameplay**  Okay, now tell me about its gameplay.

*Gameplay driven by cutting-edge interactive story research. By scanning pages from a companion printed book, the Ice-Bound Compendium, readers reveal information to a digital collaborator, who develops a sense of what themes the reader thinks are important to the overall story. The story will begin to emphasize the same themes the reader consistently focuses on.*

**Mechanics**  Next, tell me some of the specific actions the player may take in this game.

*Dynamic conversation with an intriguing, reactive character.*

**Aesthetics**  Now, tell me about the visual and aesthetic style of this game.

*Text that shimmers and changes as you read it. A gorgeous full-color print book. A nested, recursive story inspired by writers like Borges and Nabokov and books like House of Leaves.*

**Unique aspects**  Now, what makes this game unique?

*Unique merger of physical and digital storytelling. Thousands of sculptable permutations of each story.*

**Miscellaneous**  Lastly, tell me anything else about this game that you'd like to add.

*The game's constructive narrative system is part of our ongoing PhD research. An academic paper describing our combinatorial narrative system is available to those interested in more technical details.*

After the final question is answered, the sage squints in thought for a moment before the user is delivered to a generated GameNet entry for the game (shown in the last frame of Figure 2). Here, the user finds an explorable list of fifty related games that includes other unique indie games (*Dinner Date*, *Jazzpunk*, etc.), story-driven games (*I Have No Mouth, and I Must Scream*; *Anchorhead*; *Gone Home*; etc.), and games that were developed or are discussed in narrative-technology research (*Façade*, *Dwarf Fortress*). Having this generated GameNet entry, he or she is then free to explore that tool in the way we have described above.

## EXPERIMENT

We conducted a user study to evaluate the utility of GameNet and GameSage as a system that game designers may use to discover existing games related to their own works in progress, as a way of gathering insight during the early stages of game design. For readers familiar with Csikszentmihalyi's steps in the creative process (Csikszentmihalyi, 1996), we propose that our system could be most useful during the *evaluation* phase, in which a creator seeks to evaluate whether an idea is valuable and worth pursuing. In the context of game design, this would make game discovery a means for determining whether (and in what ways) a game idea is novel or interesting relative to existing games. Given the lack of dedicated tools in existence, a problem we have outlined above, the state of the art in game discovery is best represented by the coordinated usage of assorted web resources. As such, in this experiment we assessed the utility of these tools, with regard to this specific task context, relative to that of this current (lack of a) state of the art.

## Participants

182 participants (20% women) took part in the experiment, with ages ranging between 18 and 27 ($M = 19.45$). The participants were all undergraduate students enrolled in an introductory game-design course. This course fulfills a general-education requirement and has no prerequisites; as such, the students hailed from diverse academic backgrounds encompassing 42 different degree programs. In a preliminary questionnaire, we asked the participants about their level of game-development expertise: 34% reported no game-development experience prior to enrolling in the class, 57% claimed novice-level expertise, and the remaining 9% called themselves experienced.

## Experimental Design

We employed a within-subjects design in which all participants searched for existing games related to a game they had made using both of two experimental conditions: a *baseline condition*, in which they could freely use a web browser to utilize any resources available online, and a *tools condition*, in which they described their game to GameSage and then searched GameNet starting from the entry generated for the game. All participants carried out the baseline phase first. Though we realize this configuration could have led to order effects, we clarify later in the paper that these would only have made the test of the system's efficacy a more rigorous one.

## Experimental Task

Prior to us conducting the experiment, each of the students had completed an assignment in which he or she created a game emphasizing exploration in an unusual or metaphorical space. For both experimental conditions, the task was the same: participants were provided with an online form and asked to spend fifteen minutes finding games related to their respective games, entering their titles into the online form as they did. On this form, the participant was asked to place related games, as they discovered them, into one of two categories—*familiar games*, games they were already familiar with prior to the experiment, and *unfamiliar games*, games they were not familiar with prior. We let the participants operationalize their own criteria for familiarity, but we provided a specific notion of relatedness: students had an upcoming extra credit assignment in which they would be tasked with writing about games related to their own games that were not discussed in lecture, so we suggested that they deem a prospective game to be related if they would consider discussing it in this assignment.

## Procedure

As noted above, all participants were enrolled in an introductory game-design course. Experimental sessions were held during the course's six discussion sections (one session for each section) across a two-day period that was bookended by the course's first two lectures that week. Sections were led by three teaching assistants in total (each led two sections), but all experimental sessions were led by the same experimenter (who was not affiliated with the class) in the same university computer lab, which was reserved for class sections.

### *Pre-Experiment*

Each experimental session was preceded by a short preliminary phase in which the section leader discussed class logistics, introduced the experimenter, and explained the basic purpose of the study. Participants were told that the anonymous procedure would last roughly

45 minutes (about half of the section's allotted time), was not mandatory, and would not be graded, but that it could assist in the completion of the upcoming extra-credit assignment described above. The experimenter then detailed the basic protocol of the study and asked participants to log into their respective university desktop computers and start up their pre-installed Mozilla Firefox browsers. After this prelude, the participants were asked to follow a link to a survey hosted on Google Forms and to fill out a preliminary page there with demographic questions. Once everyone had filled out this preliminary page, the participants were asked to proceed to the second page of the survey, which we describe next.

### Baseline Phase

Here, participants were presented two boxes for free-text entry, one labeled 'Familiar games' and the other 'Unfamiliar games'. At this point, they were asked to spend the next fifteen minutes using any web resources available to find games related to their respective games, entering the titles of any discovered games into the appropriate box (depending on whether the participant was familiar with the game at hand prior to beginning the experiment). After the fifteen minutes had elapsed, the experimenter asked the participants to stop searching and to proceed to the next page of the survey, which was identical to the previous one.

### Tools Phase

At this stage, the participants were given a link to a special experimental GameSage instance and were asked to follow the link, spend no more than ten minutes there describing their game, and spend the remainder of the fifteen-minute phase exploring GameNet to discover related games (entering their titles into the survey form just as they had done in the previous step). Participants were permitted to repeat titles found during the previous phase, though these were not counted during data analysis, as we discuss in the next section.

### Post-Experiment

Finally, after this second phase, the participants were asked to proceed to and complete the final page of the survey, which was a post-experiment questionnaire in which we asked participants what resources they used in the baseline phase and how likely they would be to reuse GameNet and GameSage in the future. Additionally, we invited freeform feedback.

## Measures and Instruments

This experimental procedure yields three primary variables of interest that correspond to the hypotheses we postulated above:

- **Mean total games discovered.** Mean number of games discovered across all participants in a given condition (baseline or tools).
- **Percentage of games discovered that were unique.** Percentage of games discovered across all participants in a given condition that were *unique*. That is, with *total count* being the total number of discovery instances in a condition and *unique count* being the number of unique games discovered by all participants in that condition, the percentage of unique games for the condition is simply $(unique\ count/total\ count) * 100$. We use this a measure of the *diversity* of games discovered in each condition.
- **Percentage of games discovered that were unfamiliar.** Percentage of games discovered across all participants in a condition that were unfamiliar (prior to beginning the experiment) to the participants who discovered them. As mentioned above, we let

|  | Total* | % Unique | % Unfamiliar* |
|---|---|---|---|
| **Baseline** | 6.49 (6.02) | 58.2 | 47.0 |
| **Tools** | 14.16 (20.11) | 60.1 | 80.3 |

Table 1: Number of games discovered by participants ($n = 182$) in both conditions (including repetitions across multiple participants; means with standard deviations in parentheses), and percentages of those that were unique and familiar, respectively. *Asterisks indicate statistically significant differences between conditions at $p < 0.0001$.

participants develop and operate over their own criteria for game familiarity.

Additionally, we have another variable of interest for which data was collected in a post-experiment questionnaire (discussed below):

- **Likelihood to reuse tools.** After being informed that GameSage is freely available online, we asked participants to indicate, using a four-point Likert scale, how likely it is that they would revisit the tools to use them again. Available responses were 'very unlikely', 'somewhat unlikely', 'somewhat likely', 'very likely'.

### Potential Order Effects

At this point, the reader may be concerned with potential order effects that this study design could yield, given that we had all participants first undergo the baseline condition before the tools condition. We acknowledge that this probably generated a small practice effect, in that participants would have been more primed to the task of game discovery by the time they underwent the latter condition. However, we believe that, overall, our ordering privileges the baseline condition for two reasons. First, it is likely that a *fatigue effect* is also at play here, *i.e.*, that participants who had already spent the first experimental phase searching for games may have been less diligent in their search during the tools phase. Second, because we did not allow repetition in the tools phase of games discovered during the baseline phase, the space of potential related games for each participant was greatly reduced in the tools phase. Moreover, this reduction was not constituted by the removal of arbitrary games from the search space, but games that may have immediately sprung to a participant's mind upon beginning the task. For these reasons, we contend that, all other things being equal, the tools condition would be expected to produce fewer discovered games.

### RESULTS

Our main results are shown in Table 1. Because the two conditions exhibited unequal variances on mean total games discovered, and because we do not have huge sample sizes ($n = 182$ for both samples), we tested for statistical significance between these condition means using Welch's $t$-test, which is an adaptation of the standard $t$-test that better suits data with these characteristics (Ruxton, 2006). Differences between condition proportions were tested using a standard two-proportion $z$-test. Our results relative to our variables of interest (outlined above) are as follows:

- **Mean total games discovered.** Many more games were found in the tools condition (14.16 vs. 6.49 per participant); this difference was statistically significant ($t(213) = 4.92$; $p < 0.0001$).
- **Percentage of games discovered that were unique.** A greater percentage of games discovered in the tools condition were unique (60.1 vs. 58.2), but this difference was not statistically significant ($z(362) = -1.336$; $p = 0.09$).
- **Percentage of games discovered that were unfamiliar.** A far greater percentage of games discovered in the tools condition were unfamiliar to participants (80.3 vs. 47.0); this difference was statistically significant ($z(362) = 6.6$; $p < 0.0001$).
- **Likelihood to reuse tools.** 7% of participants reported they were "very likely" to reuse GameNet and GameSage, 42% indicated they were "somewhat likely" to do so, 31% responded "somewhat unlikely", and the remaining 20% chose "very unlikely".

## DISCUSSION

In this section, we discuss these results with regard to our initial hypotheses before providing some additional analysis.

## Revisiting Our Hypotheses

Here, we will revisit each of our initial hypotheses, given above, in light of the results we have just presented.

### H1: Participants will discover more related games using our tools

Our results strongly support this hypothesis. Indeed, participants discovered more than twice as many related games using the tools (nearly one per minute) than they did using assorted web resources. This difference is especially remarkable when considering that none of the games participants discovered in the baseline condition could be counted toward their discoveries in the tools condition. For our baseline condition, we chose to allow participants to use any available web resources because we believe this best represents the (lack of a) state of the art in game discovery. This may seem curious in that we have also outlined the existence of another game-discovery system, Vizmo, which perhaps could have served as a better method to compare against. The reason we did not do this is that Vizmo currently only contains 604 games, many of which are platform variants (which we merged in this experiment during data canonicalization). In both experimental conditions, participants cumulatively discovered more titles than exist in the entire database that undergirds Vizmo, and so we believe this decision is vindicated.

### H2: Participants will discover a greater diversity of games using our tools

The results did not support this hypothesis, as the difference in condition proportions was not statistically significant. While we had anticipated that the proportion of unique titles would be significantly higher in the tools condition, the fact that it even approximates that of the baseline condition is still remarkable. Because the effective search space yielded by all resources on the web is orders of magnitude greater than that represented by the 12,000 games in GameNet, one might expect that, all other things being equal, a much larger proportion of games discovered in the baseline condition would be unique. This was indeed not the case, however, and in fact 47% of the unique titles discovered during the baseline condition were games that are included in GameNet. This suggests that, had the

Figure 3: Distribution of *clarity scores* for the GameSage queries submitted during our experiment (M = 0.5, $\sigma$ = 0.11). A clarity score for a query is its average cosine similarity relative to the fifty most related games provided for it by GameNet.

tools condition been ordered first in our study design, participants might have discovered more games with the tools, as well as a greater diversity of games (since games included in GameNet that were already discovered by a participant in the baseline condition could not be discovered by that participant in the tools condition).

### H3: Participants will discover a greater proportion of unfamiliar games using our tools

Our results strongly support this hypothesis. We made this prediction from the intuition that, in lieu of a dedicated tool for game discovery, participants would tend to seek out (as a sort of scaffolding) games they could already name as related to their own. Indeed, roughly half of the games discovered by participants in the baseline condition were already known to them. We believe that the demonstrated facility of GameNet and GameSage to provide users with extensive listings of games that were previously unfamiliar to them is perhaps the system's greatest strength.

### Tool Appraisal and Query Performance

Given that participants discovered significantly more games using the dedicated discovery tools, it was surprising to find that only 49% indicated that they would be likely to reuse the tools in the future. As such, we decided to further investigate this aspect of our data, particularly with regard to whether the participants' GameSage queries themselves were correlated with those participants' reported likelihood to reuse the tools. This notion relates to the common observation in information retrieval that, in any search task, some queries will be more likely than others to yield good results (Cronen-Townsend et al., 2002). This variance can be accounted for by three properties of a query: its well-formedness, the degree to which it accommodates the matching operations enacted by the search engine to retrieve indexed items in the search space, and the degree to which there are actually valid matches among those indexed items. In information retrieval, a *clarity score* is often used as a metric for predicting query performance that captures (by an entropy calculation) the cumulative effect of these sources of variance in query performance (Cronen-Townsend et al., 2002).

For our purposes here, we propose a variant of the clarity score calculated as the average co-

sine similarity between a GameSage submission and the fifty most related games provided in its generated GameNet entry. Figure 3 plots the distribution of clarity scores for the 182 GameSage queries submitted in our experiment, which ranged between 0.32 and 0.79, with a mean and standard deviation of 0.5 and 0.11 respectively. To test whether participants' clarity scores correlated with desire to reuse GameNet and GameSage, we placed the participants into two groups—those who indicated they were "somewhat likely" or "very likely" to reuse the tools, and those who reported otherwise—and calculated mean clarity scores for each, which were 0.51 and 0.49 respectively. We tested for whether the former group's mean was significantly higher using a standard two-sample $t$-test, but the test did not support this ($t(362) = 1.44$, $p < 0.08$).

## Web Resources Used

During the baseline condition, participants reported using a total of 21 web resources, which speaks to the lack of dedicated game-discovery systems that we have outlined herein. The most frequently cited resources were Google search (used by 96% of participants), Google Images (27%), Wikipedia (23%), YouTube (18%), GameFAQs (10%), Steam (9%), and Giant Bomb (9%); Ryan et al. (2015b) reported a similar resource array. We note that materials from Google Images, Wikipedia, and YouTube are already integrated into GameNet.

## Limitations of the Tools

In our estimation, the two major limitations of these tools are both rooted in the LSA model underpinning them, particularly that model's reliance on Wikipedia text. First, Wikipedia authoring practices, and in fact the very nature of Wikipedia's notability standards, disadvantage a class of novel games that may live at the fringes of the medium. As we have outlined above, the bandwagon effect at play in videogames today works to privilege tried-and-true designs above more adventuresome ones, which limits the visibility of games showcasing the latter. As such, novel games tend to be known to less people, which makes it less likely that such games will earn Wikipedia authors or that they would even meet the website's notability standards. This is especially troublesome for the reason that obscure, novel games are exactly the type that could lend the greatest insight to designers exploring similar concepts. Indeed, in research on recommender systems (and in work on search more generally) there is a notion that obscure recommendations from the long tail are often the most valuable to a user (Ge et al., 2010), but in the case of GameNet and GameSage, the long tail is unfortunately truncated. Second, while videogame concerns surrounding, *e.g.*, the platform, development, or critical reception of a game may be useful to the GameSage user describing a game whose title he or she has forgotten (which is one potential usage of the tool), only concerns pertaining to elements of game design (*e.g.*, gameplay, aesthetics, narrative) are likely to interest most users. But the encyclopedic nature of Wikipedia leads its articles about games to often include description pertaining to any and all notable aspects of their ontologies. As such, these extraneous concerns creep their way into the model, and while they are useful in several of GameNet's use cases, game designers using GameSage would likely prefer the model to incorporate only concerns central to the process of game design.

## Post-Experiment Usage

GameNet and GameSage were made publicly available in June 2015. At the time of writing, the system has logged 7,399 sessions by 6,488 unique users (*i.e.*, 12.3% revisited), with 3.24

GameNet entries viewed per session. This indicates that approximately thirty new users try the tools each day.

## IMMEDIATE REVISIONS

In their freeform feedback, participants broadly indicated that GameSage asks too many questions and that submitting a game idea takes too long. As an immediate revision to the tool, its default mode of interaction has already been changed to one in which the user submits a full description of his or her game idea by filling out a single text-entry field. For users that seek a more directed experience, a button labeled 'Guide Me' may be clicked to engage a mode that asks most (but not all) of the same questions that were asked in the version described in this paper.

## CONCLUSION AND FUTURE WORK

The immense proliferation of videogames over the course of recent decades has yielded a discoverability problem that has gone largely unaddressed. Though this problem affects all videogame stakeholders, in this paper we have limited our concerns to the specific context of game designers seeking prior work that could inform their own ideas or works in progress. Specifically, we have presented GameNet and GameSage, components of a game-discovery system that solicits text about a user's idea for a game to generate (using an NLP technique called *latent semantic analysis*) an explorable listing of the existing games most related to that abstract idea. In a user study in which 182 game-design students used both this system and unrestricted web resources (the apparent state of the start in game discovery today) to find games related to their own, participants discovered more than twice as many titles using GameNet and GameSage, and a far greater proportion of these games were previously unfamiliar to the participants who discovered them. This suggests that this system represents the state of the art in videogame discovery, though we acknowledge that sadly there is currently little in the way of competition. More broadly, we have provided in this paper the first articulation of videogame discovery as an emerging application area.

Looking ahead, we envision several avenues of future work, some of which we are actively planning or already carrying out. First, *archetypal analysis* is a machine learning technique whose application to games is being explored by Chong-U Lim and D. Fox Harrell (*e.g.*, Lim and Harrell, 2015). By this method, instances in a data set are represented as mixtures of *pure types*, or *archetypes*, which themselves are represented as mixtures of the instances (Cutler and Breiman, 1994). This approach is alluring because the representations that it yields are *human-interpretable* (unlike purely vectorial representations), as they simply specify mixtures of other instances (in game discovery, this would be mixtures of games). We are enticed by the prospect of an *archetypal game-discovery tool* that could afford both the semiautomated game indexing of GameNet and the interpretability of Vizmo.

While previously we evaluated GameNet in the context of game scholars seeking games related to their research topics (Ryan et al., 2015b), that study preceded the creation of GameSage. We plan to evaluate the full tool suite in the context of this use case by having scholars explore GameNet entries generated by submitting abstracts of their research topics to GameSage. Further, as we have alluded to above, we believe this system could show utility as a tool with which game historians, scholars, and enthusiasts may discover games

that are related to topics of concern in videogame history. Currently, we are analyzing data from a user study that we recently conducted to evaluate GameNet and GameSage in this context. As an additional evaluation, we might also directly compare our system for game discovery to the other such existing system, Vizmo, though this presents a challenge in that the latter has far fewer games; perhaps we could limit GameNet's library of games to a subset roughly coextensive with Vizmo's.

Lastly, we have recently completed development of a variant of GameNet and GameSage whose underlying LSA model is trained using text from videogame walkthroughs extracted from the website GameFAQs. Because walkthroughs describe gameplay exclusively and systematically, often in exhaustive detail, we believe that the second major system limitation that we have outlined above—that game designers will prefer to operate over only a subset of the concerns represented in Wikipedia encyclopedic description, specifically *gameplay* concerns—could conceivably be defeated in this revision. In a future study, we will directly compare this revised system to the version trained on Wikipedia text.

We hope that this project will encourage the development of new tools that may work to better alleviate the videogame discoverability problem that we have outlined herein.

## LINK

GameNet and GameSage are freely available web apps hosted online at `https://gamecip.soe.ucsc.edu/projects`.

## ACKNOWLEDGMENTS

## BIBLIOGRAPHY

Adler, M. (1985). Stardom and talent. *The American Economic Review*.

Cronen-Townsend, S., Y. Zhou, and W. B. Croft (2002). Predicting query performance. In *Proc. Information Retrieval*.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*.

Cutler, A. and L. Breiman (1994). Archetypal analysis. *Technometrics*.

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*.

Edwards, G., E. Kaltman, J. O. Ryan, T. Hong, and N. Wardrip-Fruin (2015). Augmented exploration of library videogame holdings by techniques from computational linguistics. In *Society of American Archivists Science, Technology, and Healthcare Roundtable*.

Ge, M., C. Delgado-Battenfeld, and D. Jannach (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proc. Recommender Systems*.

Jannach, D., L. Lerche, F. Gedikli, and G. Bonnin (2013). What recommenders recommend–an analysis of accuracy, popularity, and sales diversity effects. In *User Modeling, Adaptation, and Personalization*.

Kraaij, W. and W. Post (2006). Task based evaluation of exploratory search systems. In *Proc. Evaluating Exploratory Search Systems*.

Lam, S. K. and J. Riedl (2004). Shilling recommender systems for fun and profit. In *Proc. WWW*.

Lee, J. H., R. I. Clarke, and A. Perti (2015). Empirical evaluation of metadata for video games and interactive media. *Association for Information Science and Technology*.

Lee, J. H., S. R. Hong, H. Cho, and Y.-S. Kim (2015). Vizmo game browser: Accessing video games by visual style and mood. In *Proc. CHI*.

Lee, J. H., N. Karlova, R. I. Clarke, K. Thornton, and A. Perti (2014). Facet analysis of video game genres. *Proc. iConference*.

Lim, C.-U. and D. F. Harrell (2015). Revealing social identity phenomena in videogames with archetypal analysis. In *Proc. AISB Symposium on AI and Games*.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*.

Moore, C. L. (2009). Digital games distribution: the presence of the past and the future of obsolescence. *M/C*.

Park, D. H., H. K. Kim, I. Y. Choi, and J. K. Kim (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*.

Pocket Gamer (2015, September). App store metrics. http://www.pocketgamer.biz/metrics/app-store/.

Race, T. M., M. Popp, and D. Dallis (2012). Resource discovery tools: Supporting serendipity. *Planning and Implementing Resource Discovery Tools in Academic Libraries*.

Reed, A. A., J. Garbe, N. Wardrip-Fruin, and M. Mateas (2014). Ice-bound: Combining richly-realized story with expressive gameplay. In *Proc. Foundations of Digital Games*.

Ricci, F., L. Rokach, and B. Shapira (2011). *Introduction to recommender systems handbook*. Springer.

Rose, M. (2014). More games have released on Steam so far in 2014 than all of last year. *Gamasutra*.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*.

Ryan, J. O., E. Kaltman, A. M. Fisher, T. Hong, T. Owen-Milner, M. Mateas, and N. Wardrip-Fruin (2015). Large-scale interactive visualizations of nearly 12,000 digital games. In *Proc. Foundations of Digital Games*.

Ryan, J. O., E. Kaltman, T. Hong, M. Mateas, and N. Wardrip-Fruin (2015). People tend to like related games. In *Proc. Foundations of Digital Games*.

Ryan, J. O., E. Kaltman, M. Mateas, and N. Wardrip-Fruin (2015a). Tools for videogame discovery built using latent semantic analysis. In *Proc. Foundations of Digital Games*.

Ryan, J. O., E. Kaltman, M. Mateas, and N. Wardrip-Fruin (2015b). What we talk about when we talk about games: Bottom-up game studies using natural language processing. In *Proc. Foundations of Digital Games*.

Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*.

VentureBeat (2015, January). Comparing Apples and Googles: The App Store vs. Google Play.

White, R. W., G. Marchionini, and G. Muresan (2008). Evaluating exploratory search systems. *Information Processing & Management*.

White, R. W. and R. A. Roth (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*.