



Article

A Mirror to Human Question Asking: Analyzing the Akinator Online Question Game

Gal Sasson * and Yoed N. Kenett *

Faculty of Data and Decision Sciences, Technion—Israel Institute of Technology, 3200003 Haifa, Israel

* Correspondence: galsasson@campus.technion.ac.il (G.S.); yoedk@technion.ac.il (Y.N.K.)

Abstract: Question-asking is a critical aspect of human communications. Yet, little is known about the reasons that lead people to ask questions, which questions are considered better than others, or what cognitive mechanisms allow the ability to ask informative questions. Here, we take a first step towards investigating human question-asking. We do so by an exploratory data-driven analysis of the questions asked by Akinator, a popular online game of a genie who asks questions to guess the character that the user is thinking of. We propose that the Akinator's question-asking process may be viewed as a reflection of how humans ask questions. We conduct an exploratory data analysis to examine different strategies for the Akinator's question-asking process, ranging from mathematical algorithms to gamification-based considerations, by analyzing complete games and individual questions. Furthermore, we use topic-modelling techniques to explore the topics of the Akinator's inquiries and map similar questions into clusters. Overall, we find surprising aspects of the specificity and types of questions generated by the Akinator game, that may be driven by the gamification characteristics of the game. In addition, we find coherent topics that the Akinator retrieves from when generating questions. Our results highlight commonalities in the strategies for question-asking used by people and by the Akinator.

Keywords: question asking; topic-modelling; Akinator; LDA; BERT; text analysis



Citation: Sasson, G.; Kenett, Y.N. A Mirror to Human Question Asking: Analyzing the Akinator Online Question Game. *Big Data Cogn. Comput.* **2023**, *7*, 26. <https://doi.org/10.3390/bdcc7010026>

Academic Editor: Carson K. Leung

Received: 6 December 2022

Revised: 22 January 2023

Accepted: 25 January 2023

Published: 29 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Questions are essential in human interactions, from children to adults [1–3]. Questions are used in a majority of the conversations we have, efforts to acquire knowledge, and attempts to solve problems [4–6]. Thus, one may ask: Why do we ask questions? What is the functional significance of asking questions in human communication and what are the cognitive mechanisms that realize this ability? Choosing which questions to ask and at what moment to ask these questions may have significant effects on the information gained [7]. Surprisingly, despite its ubiquity in our daily communication, scarce research has been conducted on the cognitive capacity of asking questions. Currently, only a few formal models have been proposed regarding question-asking and question evaluation (e.g., [4,7]). However, these models are challenging to apply to the rich variety of questions asked in everyday life.

One potential reason for asking questions is to reduce entropy and increase information gain [6,8]. In information theory, entropy measures the amount of uncertainty in an environment of observations and thus in many cases it is a value that we wish to minimize [9]. On the other hand, information gain measures the amount of information added by a single observation, and thus it can be referred to as the difference in entropy before and after adding the new observation [10].

Another influential account of how people ask good questions is based on the optimal experiment design (OED) framework [10–12]. According to the OED framework, people choose questions to ask by aiming to maximize the knowledge they expect to gain from knowing the answer to that question [13,14]. However, the OED framework relies on

multiple theoretical assumptions concerning the question askers' prior knowledge and cognitive capacities, as well as being unsuccessful in capturing the full richness of human question-asking [10].

Another proposal for the role of questions is in the resolution of different alternatives—e.g., choosing between competing alternatives based on a question-asking process [4]. Rothe et al. [4] conducted a series of behavioral and computational studies on this perspective, via various scenarios of the battleship game, to study ambiguous situations. The authors demonstrate how participants consider alternatives via the questions that they asked across these various scenarios, albeit not in an optimal fashion. Overall, the authors acknowledge that further computational research is needed to elucidate the human capacity to ask questions [4].

The research in question-asking is making significant advances through the application of computational models [8]. While early such research mostly explored rule-based methods that strongly depend on a priori rules [14,15], current research adapts methods from neural network models. However, such current models are still far from being able to model the full complexity of human question-asking [8]. For example, Damassino [15] proposed a questions Turing test, where the player needs to accomplish a yes/no enquiry in a humanlike and strategic way, i.e., with as few questions as possible. Yet, to do so, knowing the types of questions and the process of asking questions aimed at information seeking is necessary.

Bloom et al. [16] proposed a taxonomy of question types, with increasing abstraction, ranging from basic-level questions for information retrieval to higher-level questions for synthesis and integration. Given the role of questions in problem finding—the initial stage of the creative process which is critical in formalizing and identifying the problem space [17]—it is surprising how little is known about why humans ask questions and what types of questions are asked. Here, we aim to take a step forward in analyzing the process of question-asking, by gaining insight from online question-asking games. Such insights may help us move forward in characterizing and elucidating the question-asking process.

Among numerous tasks that require question-asking, many social games use questions as the key to proceeding and accomplishing certain goals of the game [18]. These types of games may reflect aspects of the question-asking process, and therefore in this paper, we chose such a game as a case study. We focus on Akinator—a worldwide popular online game (<https://en.akinator.com/>, accessed on 1 November 2021), developed by the French company Elokence around 2007. This game has gained much interest regarding its process due to its impressive performance and success rate [18]. In the game, a magical genie—the Akinator—asks users to think of a character (real or fictional), object or animal. Akinator will ask the user questions until it guesses who or what the user has thought of. This is a version of the “20 questions game”, where one player can ask the other player up to 20 questions to guess what they thought of. The main spotlight of this game is directed at characters, rather than on objects or animals, so this research focuses on characters as well. The questions posed to the player are yes or no questions, but the user can also answer with “don't know”, “probably” or “probably not” if they are unsure of the right answer. Since it was first published, the game was played billions of times and is available on multiple platforms.

An example of gameplay with the Akinator might be: the user thinks about Mickey Mouse and chooses it as its character for the game. Then, the Akinator begins to ask questions regarding the character, and the user provides answers based on their knowledge.

Q: Is your character real?

A: No.

Q: Is your character a male?

A: Yes.

Q: Does your character wear shoes?

A: Yes.

Q: Does your character have special powers?

A: No.

Q: Is your character from a cartoon?

A: Yes.

This process continues with additional questions until the Akinator finally announces “I think of Mickey Mouse” and asks the user to indicate whether it’s right. In this case, the Akinator is indeed right. If the Akinator was wrong, the user can ask the Akinator to continue asking questions until it makes another guess.

Obviously, the Akinator needs to use some strategy for the selection of the questions to be asked, otherwise, it would need to ask an enormous number of questions to guess the right answer. What are these possible strategies, and what types of questions facilitate its ability to achieve its goal of “guessing” the character the player is thinking of? This is the focus of our study. Our interest in examining the Akinator question-asking process is to gain insight into a potential human-like question-asking process. Thus, we aim to capitalize on the extant computational research conducted on question-asking (e.g., [4,7]), while moving towards more empirical research to further elucidate this critical human capacity.

One potential strategy utilized by the Akinator is based on a binary search—mathematically, if each question can eliminate half the objects, 20 questions are enough to allow the question asker to distinguish between $2^{20} = 1,048,576$ objects. Accordingly, since the Akinator is based on the concept of twenty questions games, an effective strategy would be to ask questions that split the field of remaining possibilities roughly in half each time. This can be used with decision tree learning algorithms—methods that are used to evaluate discrete-valued target functions [18]. A decision tree is a graph with nodes and branches, where each internal node and branch represent an alternative and each leaf node represents a decision. In the problem of the Akinator, the questions are used as the internal nodes and the characters as the leaves nodes. Many algorithms have been developed for building and traversing decision trees in the most efficient way and maximizing the information gain, which we will not cover here. Such techniques need to take into account outliers—users might give different answers for the same question and character, either because they presume it’s the right answer or because they get confused, didn’t know the answer or tried to trick the Akinator. To deal with such cases, it is possible to use methods like Random sample consensus [19]. The Random sample consensus is an iterative method for fitting a model when the observed data contains a significant number of outliers. In this case, it can mean iterating over small subsets of randomly chosen questions which are enough to give a single answer, rather than all questions which have been answered, and then taking the answer that was repeated for most subsets. Another possible strategy is gamification—asking questions that are funny, less popular, or seem unrelated to the information the application would want to gain. Such questions may make the users wonder about the reasons for the question or surprise the user when the Akinator eventually succeeds at guessing the correct character. In other words, these types of questions may be meant to amuse and engage the user, rather than optimally maximize information gain to help the application correctly guess the character.

We conduct an exploratory data analysis of the Akinator’s question-asking process. Specifically, we aim to examine the different possible question-asking strategies and gain insights into the content domains that facilitate success in the question-asking process. As the program is patent protected, we were limited in our ability to access its question-asking algorithm or the database that it is based on. Thus, we aim to “reverse engineer” the game, by creating a large dataset of questions asked by the program and analyzing its question-asking process and content.

2. Materials and Methods

2.1. Dataset

To track and analyze the questions posed by the Akinator, we used the `akinator.py` API wrapper for the online game (<https://pypi.org/project/akinator.py/>, accessed on

1 November 2021). It allows writing simple programs that connect to the Akinator's servers, play games and save the results. The experiments were run on a computer with an Intel i5 processor running at 2.50 GHz, using 8.00 GB of RAM. Using this API through Python 3.8, we collected data from November 2021 to January 2022 and created multiple datasets to be analyzed:

Questions asked by the Akinator—4000 games were played in two runs, two months apart from each other, with constant answers, to collect as many questions as possible. The games were played without predetermining a character, as the goal was only to document the questions. Each of the five possible answers ('yes', 'no', 'I don't know', 'probably' and 'probably not') was used in 800 of the games as the answer to all questions throughout that game, regardless of the questions asked. 7365 distinct questions were collected in total in the first run, and 108 additional questions were collected in the second run, bringing the dataset to a total of 7473 distinct questions.

Questions by games—1500 additional games were played with constant answers (always selecting the same type of answer to the Akinator's questions), divided equally between each of the five possible answers. Again, we did not predetermine any character for these games. For this set, we grouped the questions for each game. In other words, we produced a list of games, each containing all the questions that were asked in that game. In all games, the Akinator asked exactly 79 questions before ending the game.

First questions—10,000 games were initiated in two different runs and the first question of each game was recorded. The runs were completed a month apart from each other.

Real games on specific characters—we manually played multiple games about specific known characters (e.g., Mickey Mouse, Mark Zuckerberg) while making sure that questions that are repeated between games get the same answer.

2.2. Methods

2.2.1. Question Analysis

We begin by conducting exploratory data analysis on the datasets we created. We analyze the characteristics of the games and the questions that were collected, the most prominent questions and the first questions chosen by the Akinator.

2.2.2. Topic Modelling

A topic is the theme of a text, the subject that is being discussed in it. Topic modelling can be used to characterize texts through only a few words, allowing for the identification of shared themes between documents. Topic modelling is an unsupervised-learning method for discovering the topics in a collection of documents. A topic is defined by a set of weighted words that are assumed to be the most informative to describe the documents within the mutual subject. Previous research has shown how topic modelling can be used to study complex behavior [20]. We model the topics of the Akinator's questions, via two different approaches.

LDA Topic Modelling

Latent Dirichlet Analysis (LDA) is the most frequently used method for performing topic modelling [21] and is based on probabilistic assumptions. We use Gensim's LDA model [22] with two different approaches regarding the dataset: In the first approach, we look for topics representing each individual question, containing 7473 distinct questions. Each question is up to 17 words long, which means that the documents used for this approach are relatively short. Also, we can assume that each question (document) is related to only one topic. In the second approach, we examine topics in entire games comprising multiple questions as a document—in this case, we concatenated all the questions that were asked in a single game into one document (79 questions), to a total of 1500 long documents.

Preprocessing: First, we removed punctuation, removed stop-words (common words in the language with little value to the sentence, e.g., and, are, etc.), and created bigrams (pairs of adjacent words in the sentence). We then chose only words that are nouns and

performed lemmatization. As a final step, we removed unindicative words—words that don't contribute to the meaning of the topics and may get in the way of creating meaningful topics (e.g., character, link, involve, etc.). For example, the question 'Does your character make cooking videos?' was processed to ['cooking', 'video']. After preprocessing, a document corpus was created, as well as a term-frequency list.

Finding the number of topics: LDA models require a pre-defined number of topics. To choose the value for this parameter, we estimated the performance of a model using a coherence score. The topic coherence scores a single topic by assessing the degree of semantic similarity between high-weighted words in the topic, as a measure of how interpretable the topic is for human beings. This allows for distinguishing between topics that are semantically interpretable and topics that were produced due to statistical inference. We set all other hyperparameters (which will be described in the next section) to their default values and chose the full model based on the number of topics with the highest coherence score.

Finding optimal hyperparameters: After choosing the number of topics, we optimized some of the hyperparameters, specifically *alpha* (the a-priori belief about the mixture of topics in documents), *passes* (the total number of iterations of training) and *chunksize* (the number of documents that are loaded together in every chunk in an iteration). We tried all the combinations of *alpha* = ['auto', 0.1, 0.5, 0.9], *passes* = [10, 20], *chunksize* = [20, 50, 100, 200] and chose the combination that led to the highest coherence score.

BERT Topic Modelling

Current alternatives to LDA models are methods based on distributed representations of documents and words, which have shown improvements in topic coherence [23]. Our implementation is based on Top2Vec [24] and specifically uses the pre-trained transformer model BERT (Bidirectional Encoder Representations from Transformers), as implemented in BERTopic [25]. Importantly, the BERT model has been successfully implemented in question-answering systems [26–28]. LDA has many weaknesses that Top2Vec overcomes [24]. First, LDA assumes that the number of topics k is known, but this is usually not the case, especially for very large datasets. In addition, LDA requires a lot of preprocessing of data and the creation of a well-defined list of stop words as well as hyperparameter tuning. These are all limitations which we faced when we tried to create a meaningful LDA model for our data.

Top2Vec learns a continuous representation of topics by jointly embedding word, document and topic vectors such that the distance between the vectors represents the semantic similarity between them [24]. This method does not require determining the number of topics in advance, nor removing stop-words or performing lemmatization or stemming to learn good topic vectors. Angelov [24] shows that Top2Vec consistently finds topics that are significantly more informative and representative of the corpus than the ones produced by traditional models such as LDA. Using this approach, we performed the following steps:

Dataset: We use the dataset of 7473 distinct questions. As the questions are relatively short, it is reasonable to assume that every question belongs to only one topic (or at least one major topic).

Embeddings: We convert the documents (questions) to vectors using the pre-trained BERT model, which extracts different embeddings based on the context of every word. Reducing the dimensions of the embeddings allows a more efficient and accurate way to identify dense clusters of documents. UMAP (Uniform Manifold Approximation and Projection) is a dimension reduction algorithm that is based on manifold learning techniques, which has been shown to scale well to large datasets while preserving both the local and global structure [24]. Based on the best results shown in this paper, we used UMAP to reduce the dimensionality of the vectors to 5 with the size of the neighbourhood set to 15. The number of nearest neighbors controls the balance between the local and global

structure in the new embedding, and this value gave the best results in preserving the local structure.

Clustering: We clustered the UMAP reduced document embeddings using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), which is a density-based algorithm that fits UMAP well. HDBSCAN recognizes dense clusters of documents as prominent topics, so it assigns a label to each of the documents in the cluster. For the rest of the documents, it assumes they are not descriptive enough to be included in any of the topics and assigns them the noise label. The parameter for minimal cluster size was set to 30 to prevent too many clusters to be formed.

Topic creation: BERTopic uses a class-based variant of TF-IDF (Term Frequency–Inverse Document Frequency), referred to as c-TF-IDF. TF-IDF is a measure that quantifies how relevant a word is to a document, in a collection of documents (corpus). This measure considers not only a single document, but the appearance of each term across all documents, which helps to adjust the weighting of very common words (like stop-words or words that are specific to the subject of the corpus). This score increases in relation to the relevance of the term to the document. The c-TF-IDF allows us to also look at each set of clustered documents and differentiate between them. For this measure, all the documents of a single cluster are joined into one long document. Then, the following score can be given to each term in each cluster:

$$c\text{-TF-IDF}_i = \frac{t_i}{w_i} \times \log \frac{m}{\sum_j^n t_j} \quad (1)$$

where t_i is the frequency of term t in cluster i and it is divided by the total number of words in the cluster, w_i , as a regularization to adjust the weighting to frequent words in the cluster. This is multiplied by the logarithmically scaled fraction of the total number of documents across all clusters m divided by the sum of occurrences of term t in all those documents. The final score for each word in each cluster expresses the importance of that word to the topic, and thus each topic can be defined by the highest-scoring words in the cluster. For this part we used a common list of stop words and extended it with common words in the data that are not informative enough for the topics (e.g., character, link, etc.).

3. Results

3.1. Question Analysis

In all games played with constant answers, the Akinator asked exactly 79 questions before stopping. The number of new questions constantly drops as more games are pursued, indicating that the dataset of questions converges towards the full dataset used by the program (Figure 1; top panel). We also show the top 20 most frequent questions that repeat in the analyzed games (Figure 1; bottom panel). Some questions seemed to repeat more often than others for specific characters, indicating that they were important in determining the exact character. For example, when the chosen character was Mickey Mouse the question ‘Is your character an animal?’ appeared in all 30 played games, and the question ‘Is your character a mouse?’ appeared in 29 of them. The longest question was 17 words long, while the shortest question has only 3 words (‘Is he hot?’). We found some examples of similar questions that are being asked in different variations: ‘Does your character really exist?’ and ‘Is your character real?’; ‘Is your character in an anime?’ and ‘Is your character from an anime?’; ‘Is your character a TV show host?’ and ‘Is your character a TV host?’.

We next examined the number of times each question appeared as a first question in a game, in each of the runs (Table 1). The runs were completed a month apart from each other and show a difference in the topics, which might indicate a change that happened in the program during that time, as the Akinator keeps learning from games that are played.

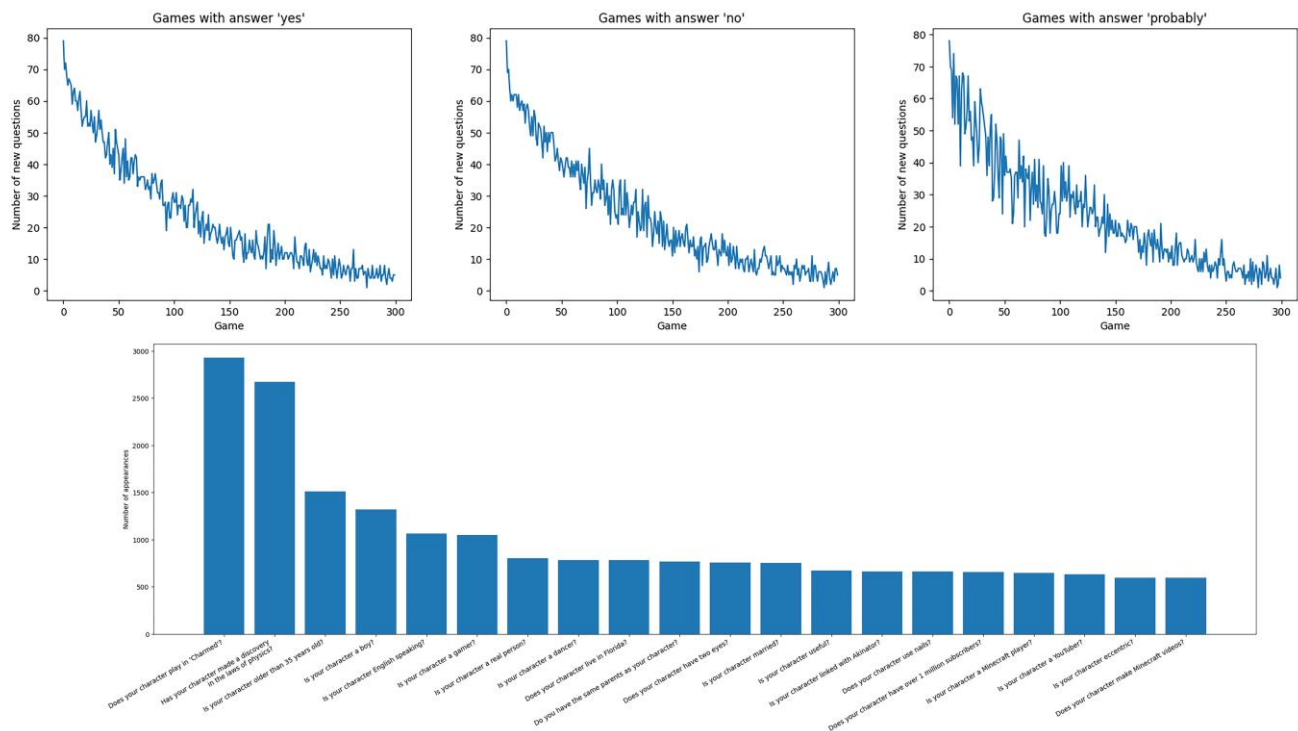


Figure 1. (Top panel) The number of new distinct questions found through 300 games played with a constant answer (‘yes’, ‘no’ or ‘probably’). This number consistently decreases as more games are played, until only a few new questions are found with every new game. (Bottom panel) The top 20 most frequent questions in 4000 games. The games were played with constant answers throughout the whole game, so that each of the five possible answers (‘yes’, ‘no’, ‘I don’t know’, ‘probably’ and ‘probably not’) was used in 800 of the games.

Table 1. Questions that were asked as first questions in games and the number of times they appeared in each of the two runs. Clearly, the set of questions changed over time.

Question	Run 1—November 2021	Run 2—December 2021	Total
Is your character a girl?	1132	2164	3296
Is your character a Youtuber?	1603	8	1611
Does your character make Youtube videos?	1115	0	1115
Is your character known for making Youtube videos?	1007	0	1007
Is your character a woman?	143	42	185
Is your character real?	0	268	268
Is your character a real person?	0	413	413
Does your character know your name?	0	337	337
Is your character a female?	0	920	920
Does your character personally know you?	0	651	651
Have you ever talked to your character?	0	8	8
Is your character a make?	0	189	189
	5000	5000	10,000

3.2. LDA Topic Modelling Analysis

To find the best number of topics k we set $chunksize = 100$, $passes = 10$, $alpha = 'auto'$, and changed the value of k in the range $[1, 30]$ with steps of 2. We measured the resulting

the topics still have a few words that seem unrelated, they are interpretable. For example, topic #2 can be titled Death, topic #4 can be titled Music, topic #5 can be titled Appearance, and topic #7 can be titled Animals.



Figure 4. The 20 highest weighted words in each of the largest topics created by the BERT model. In total, 44 topics were created.

4. Discussion

A fundamental aspect of human communication is asking questions [6,8]. Currently, only a few frameworks have been proposed to explain why people ask questions and what characterizes good questions (e.g., [4,7]). These frameworks focus on information gain [6,8], optimal experiment design [10], or choosing between alternatives [4]. Yet, much is still unknown regarding the rich, flexible capacity of human question-asking. Critically, empirical research on question-asking is lacking. Here, we argue that such a vacuum can be filled by capitalizing on online question-asking games [29]. Such games can help us empirically delve into what questions people ask and how their question-asking process facilitates various goals, such as figuring out which character one is thinking of.

The goal of this exploratory data analysis study was to use the Akinator question-asking game to explore the basis of question-asking. In many games, players are required to ask questions to proceed in the game, and therefore such games can be used to analyze the way people ask questions. We aimed to reverse engineer the Akinator, a popular and amusing online game where a genie guesses the character the player is thinking about by asking questions about the character [18].

As a first step, we created multiple datasets of questions which were used throughout the research. We found that some questions are very popular and appear in most of the games, while others are very rare. Although it can be assumed that the most popular questions are the ones that contribute the most to lowering the uncertainty, some of them were very surprising questions, possibly indicating that there are other considerations in the choice of the questions. For the first question of the game, it seems that the Akinator has a narrower set of questions that it chooses from, mostly about the gender of the character, its realness, and its connection to Youtube. Looking at complete games also revealed

interesting aspects about the ordering of the questions, as in some cases general questions were asked after more specific ones, instead of the other way around.

These datasets allowed us to examine the topics of the questions. We used two approaches to topic modelling—LDA and BERTopic. The LDA model indicated there are only a few question topics. Although the measurements showed that the topics are coherent, they were almost impossible to interpret as the groups of words seemed meaningless. The BERTopic model resulted in over 40 topics which were easily interpretable, as the highest-scored words in them showed very clear subjects. Some of the most popular topics were death, family, music, appearance, animals and more. Overall, these findings highlight the importance of structured knowledgebases of diverse topics that allow asking questions and utilizing such knowledgebases to make inferences. In humans, such knowledge is stored in semantic memory [30,31], where richer semantic memory facilitates creativity via the connection of remote ideas together [32–34]. Thus, our findings further demonstrate the role of a diverse set of knowledge domains in facilitating knowledge integration via a question-asking process, in line with the Bloom et al. taxonomy of questions [16].

The results highlight commonalities between the way people ask questions and the way the Akinator does. Both combine multiple strategies when asking questions—not only choosing the question that will lead to the highest information gain and minimize the entropy but addressing other considerations as well. For the Akinator, such considerations can be related to gamification, getting the user to be more engaged with the program and adding elements of surprise and amusement.

One potentially obvious reason to ask questions is related to information-seeking behaviour. Recently, Kenett, Humphries, and Chatterjee [35] argued that what links curiosity, creativity, and aesthetic perception is information-seeking behavior, facilitated by semantic memory and the personality trait Openness to Experience. Such cognitive and personality factors may determine what characterizes one's ability to ask good questions [4]. Importantly, to come up with creative solutions, an initial stage where the problem is elucidated (or constructed or found), is needed [36]. Such a problem construction stage is likely realized by a strategic question-asking process. Thus, advancing the development of computational models of question-asking, along line moving towards empirical research on question-asking—such as the current study, albeit being a small exploratory data analysis—can greatly elucidate how humans harness questions when aiming to achieve varied information-seeking goals.

This research was limited mostly on the technical side, as the Akinator is patent protected. The program is not publicly accessible, and we did not have full knowledge about its logic and considerations which we could test against our hypotheses, nor did we have its databases. We overcome this by using the API to create new datasets, but it is restricted in the functionalities it provides. The datasets may be expanded by playing more games through the API, to extend and reproduce the results of this research. In addition, the Akinator database may be noisy and inaccurate. New questions can be suggested by players, questions that are validated by the user who suggests them and through a moderator. Yet, it seems that the system fails in finding similarities between questions, weakening the strength of our findings. Future work is needed to replicate and expand our findings, to advance theoretical insight and empirical elucidation of human question-asking. Such work should include exploring more applications that are based on question-asking, as well as building research-specific question-asking games to allow for experimental control over the knowledge base and question-asking strategies utilized by the game [29,37]. Another limitation is that we only examined a minimal set of explicit answers to the Akinator questions, e.g., yes, no, probably, while natural human responses to questions are not discrete and vary from concrete responses (e.g., yes, no) to ambiguous responses, even non-verbal responses (such as gestures). However, limiting the possible states in machine learning studies on question-asking is a common practice [4]. Ruggeri et al. analyzed the question quality of unconstrained yes/no questions generated by children and adults [3]. Therefore, future studies should move towards less constrained

question-asking & response types, to better capture the rich, unconstrained and flexible range of human question-asking.

5. Conclusions

Why do humans ask questions? Several, yet to be explored reasons include knowledge acquisition, demonstrating one's own knowledge or shifting the topic of the conversation. In addition, asking questions that cover a wide range of subjects, sometimes unrelated to each other, is important in reducing uncertainty. Despite the ubiquity of question-asking in human communication, much is unknown about this capacity nor how it is utilized in information-seeking behavior.

Here, we tackle this challenge through an initial empirical exploratory data analysis that attempted to reverse engineer the question-asking process of the online game Akinator. Our effort is significantly constrained by the patent constraints of the software. Yet, it illustrates the potential of analyzing such online question-asking games via text-based methods, deriving further insight into the time-unfolding process of asking questions. Thus, the current study provides a proof-of-concept for empirically analyzing question-asking games, when pursuing to elucidate human question-asking. Given the increasing popularity of online language models for dialogue, such as chatGPT (<https://openai.com/blog/chatgpt/>, accessed on 1 December 2022), our work highlights fruitful new and exciting avenues of research based on such question-asking games.

Author Contributions: Conceptualization, G.S. and Y.N.K.; methodology, G.S.; formal analysis, G.S.; investigation, G.S.; writing—original draft preparation, G.S.; writing—review and editing, G.S. and Y.N.K.; supervision, Y.N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the US-Israel Binational Science Foundation (grant number 2021040).

Institutional Review Board Statement: Ethical review and approval were waived for this study since no human data was collected as part of this study.

Informed Consent Statement: Participant consent was waived since no human data was collected as part of this study.

Data Availability Statement: Data collected in this study is available upon request from the corresponding author.

Acknowledgments: We thank Talia Wise for her help in editing this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ruggeri, A.; Xu, F.; Lombrozo, T. Effects of explanation on children's question asking. *Cognition* **2019**, *191*, 103966. [[CrossRef](#)]
2. De Simone, C.; Ruggeri, A. What is a good question asker better at? From unsystematic generalization to adult-like selectivity across childhood. *Cogn. Dev.* **2021**, *59*, 101082. [[CrossRef](#)]
3. Ruggeri, A.; Lombrozo, T.; Griffiths, T.L.; Xu, F. Sources of developmental change in the efficiency of information search. *Dev. Psychol.* **2016**, *52*, 2159–2173. [[CrossRef](#)] [[PubMed](#)]
4. Rothe, A.; Lake, B.M.; Gureckis, T.M. Do people ask good questions? *Comput. Brain Behav.* **2018**, *1*, 69–89. [[CrossRef](#)]
5. Getzels, J.W. The problem of the problem. *New Dir. Methodol. Soc. Behav. Sci. Quest. Fram. Response Consistency* **1982**, *11*, 37–49.
6. Gottlieb, J. The effort of asking good questions. *Nat. Hum. Behav.* **2021**, *5*, 823–824. [[CrossRef](#)]
7. Nelson, J.D. Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychol. Rev.* **2005**, *112*, 979–999. [[CrossRef](#)]
8. Wang, Z.; Lake, B.M. Modeling question asking using neural program generation. *arXiv* **2019**, arXiv:1907.09899.
9. Crupi, V.; Nelson, J.D.; Meder, B.; Cevolani, G.; Tentori, K. Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. *Cogn. Sci.* **2018**, *42*, 1410–1456. [[CrossRef](#)]
10. Coenen, A.; Nelson, J.D.; Gureckis, T.M. Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychon. Bull. Rev.* **2019**, *26*, 1548–1587. [[CrossRef](#)]
11. Hawkins, R.; Goodman, N. Why do you ask? The informational dynamics of questions and answers. *PsyArXiv* **2017**. [[CrossRef](#)]

12. Boyce-Jacino, C.; DeDeo, S. Cooperation, interaction, search: Computational approaches to the psychology of asking and answering questions. *PsyArXiv* **2021**. [[CrossRef](#)]
13. Myung, J.I.; Pitt, M.A. Optimal experimental design for model discrimination. *Psychol. Rev.* **2009**, *116*, 499–518. [[CrossRef](#)] [[PubMed](#)]
14. Gureckis, T.M.; Markant, D.B. Self-directed learning: A cognitive and computational perspective. *Perspect. Psychol. Sci.* **2012**, *7*, 464–481. [[CrossRef](#)] [[PubMed](#)]
15. Damassino, N. The questioning Turing test. *Minds Mach.* **2020**, *30*, 563–587. [[CrossRef](#)]
16. Bloom, B.S.; Engelhart, M.D.; Furst, E.J.; Hill, W.H.; Krathwohl, D.R. *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook 1: Cognitive Domain*; David McKay Company, Inc.: New York, NY, USA, 1956.
17. Hu, W.; Shi, Q.Z.; Han, Q.; Wang, X.; Adey, P. Creative scientific problem finding and its developmental trend. *Creat. Res. J.* **2010**, *22*, 46–52. [[CrossRef](#)]
18. Zhangozha, A.R. On techniques of expert systems on the example of the Akinator program. *Artif. Intell. Sci. J.* **2020**, *25*, 7–13.
19. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
20. Hass, R. Modeling Topics in the Alternative Uses Task. *PsyArXiv* **2018**. [[CrossRef](#)]
21. Vayansky, I.; Kumar, S.A.P. A review of topic modeling methods. *Inf. Syst.* **2020**, *94*, 101582. [[CrossRef](#)]
22. Rehurek, R.; Sojka, P. *Gensim—Python Framework for Vector Space Modelling*; NLP Centre, Faculty of Informatics, Masaryk University: Brno, Czech Republic, 2011; Volume 3.
23. Bianchi, F.; Terragni, S.; Hovy, D. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv* **2020**, arXiv:2004.03974.
24. Angelov, D. Top2Vec: Distributed representations of topics. *arXiv* **2020**, arXiv:2008.09470.
25. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
26. Zope, B.; Mishra, S.; Shaw, K.; Vora, D.R.; Kotecha, K.; Bidwe, R.V. Question answer system: A state-of-art representation of quantitative and qualitative analysis. *Big Data Cogn. Comput.* **2022**, *6*, 109. [[CrossRef](#)]
27. Daoud, M. Topical and non-topical approaches to measure similarity between Arabic questions. *Big Data Cogn. Comput.* **2022**, *6*, 87. [[CrossRef](#)]
28. Simanjuntak, L.F.; Mahendra, R.; Yulianti, E. We know you are living in bali: Location prediction of Twitter users using BERT language model. *Big Data Cogn. Comput.* **2022**, *6*, 77. [[CrossRef](#)]
29. Rafner, J.; Biskjær, M.M.; Zana, B.; Langsfjord, S.; Bergenholtz, C.; Rahimi, S.; Carugati, A.; Noy, L.; Sherson, J. Digital games for creativity assessment: Strengths, weaknesses and opportunities. *Creat. Res. J.* **2022**, *34*, 28–54. [[CrossRef](#)]
30. Kumar, A.A. Semantic memory: A review of methods, models, and current challenges. *Psychon. Bull. Rev.* **2021**, *28*, 40–80. [[CrossRef](#)] [[PubMed](#)]
31. Kumar, A.A.; Steyvers, M.; Balota, D.A. A critical review of network-based and distributional approaches to semantic memory structure and processes. *Top. Cogn. Sci.* **2022**, *14*, 54–77. [[CrossRef](#)]
32. Abraham, A.; Bubic, A. Semantic memory as the root of imagination. *Front. Psychol.* **2015**, *6*, 325. [[CrossRef](#)] [[PubMed](#)]
33. Beaty, R.E.; Kenett, Y.N.; Hass, R.W.; Schacter, D.L. Semantic memory and creativity: The costs and benefits of semantic memory structure in generating original ideas. *Think. Reason.* **2022**, 1–35. [[CrossRef](#)]
34. Kenett, Y.N.; Faust, M. A semantic network cartography of the creative mind. *Trends Cogn. Sci.* **2019**, *23*, 271–274. [[CrossRef](#)]
35. Kenett, Y.N.; Humphries, S.; Chatterjee, A. A thirst for knowledge: Grounding creativity, curiosity, and aesthetic experience in memory and reward neural systems. *Creat. Res. J.* **2023**, 1–15. [[CrossRef](#)]
36. Arreola, N.J.; Reiter-Palmon, R. The effect of problem construction creativity on solution creativity across multiple everyday problems. *Psychol. Aesthet. Creat. Arts* **2016**, *10*, 287–295. [[CrossRef](#)]
37. Paravizo, E.; Crilly, N. Computer games for design creativity research: Opportunities and challenges. In Proceedings of the International Conference on Design Computing and Cognition, Glasgow, UK, 4–6 July 2022; pp. 379–396.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.