Spot the Spy: Exploring Natural Question-Asking in Gaming Environments

Gal Sasson Lazovsky[1] (**galsasson@campus.technion.ac.il**)

Benjamin T. Cook[2] (**benjytcook@gmail.com**)

Vered Pnueli[2] (**vered_pnueli@shenkar.ac.il**)

Yoed N. Kenett[1] (**yoedk@technion.ac.il**)

[1] Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel

[2] Shenkar College of Engineering and Design, Ramat-Gan, Israel

Abstract

Question-asking is a crucial aspect of human interaction. Questions fuel engagement, stimulate thought processes, foster learning, and facilitate information seeking behavior. Yet, scarce empirical research on question-asking, or its relation to related cognitive capacities such as creativity and intelligence, exists. We empirically investigate how people ask questions and the connections between question-asking and creativity through the domain of interactive gaming. To do so, we developed an online game—Spot the Spy—where players are required to find a hidden spy amidst a crowded room, by asking questions that guide them in their investigation. Thus, we dive into the very essence of how creative and strategic thinking collaborate to shape the queries we formulate. We find that players' gameplay correlates with their cognitive abilities, especially with intelligence measures. As such, our game captures insights into the profound ways creative cognition shapes the questions we articulate and navigate within dynamic gaming environments.

Keywords: question-asking; creativity; online game

Introduction

Posing questions is integral to human dialogue, permeating our interactions from early childhood to adulthood (De Simone & Ruggeri, 2021; Ruggeri et al., 2016, 2019). This intrinsic aspect of conversation not only enhances social engagement but also plays a critical role in deepening mutual understanding and likeability (Huang et al., 2017). The pursuit of knowledge and problem-solving endeavors are often propelled by the strategic use of questions (Gottlieb, 2021; Rothe et al., 2018). The selection and timing of questions are important in determining the quality and quantity of information obtained (Nelson, 2005). Yet, much is still unknown about why we ask questions, how questions vary in their quality, or how question-asking relates to various cognitive capacities, such as those facilitating information-seeking behaviors (e.g., creativity, curiosity; Kenett et al., 2023).

From an information-theoretic perspective, asking questions is a mechanism for reducing uncertainty (entropy) and enhancing knowledge acquisition (information gain) (Gottlieb, 2021; Wang & Lake, 2019). Entropy, in this context, refers to the level of unpredictability in a set of observations, which ideally should be minimized (Crupi et al., 2018). Information gain, on the other hand, is the measure of the value added by new data, essentially the reduction in entropy achieved by introducing new information (Coenen et al., 2019). In the context of decision-making, question-asking has been investigated particularly in terms of choosing between alternatives (Rothe et al., 2018). Through a combination of computational studies and behavioral experiments, including game-based methodologies, Rothe et al. (2018) explored how questions facilitate decision-making processes.

Question-asking also requires creativity (Torrance, 1987)—the ability to come up with original and effective ideas. Prior research have contended that the capacity for creativity is intricately linked to one's ability to pose questions, noting a strong correlation between the types of questions asked and individual levels of creativity (Albergaria-Almeida, 2011). Their study reveals that individuals who ask primarily closed questions tend to exhibit lower levels of creativity, while those who pose a variety of questions showcase higher creative aptitude (see also Acar et al., 2023; Raz et al., 2023). Furthermore, the act of problem construction often necessitates queries about the fundamental goals of a problem-solving task, linking this process to question-asking and creative problem-solving as well (Abdulla et al., 2020; Arreola & Reiter-Palmon, 2016; Hu et al., 2010).

Despite the theoretical significance of question asking, little empirical research on question asking exists. This is likely due to methodological challenges in how to empirically collect, and assess, question asking. Various approaches have been proposed to how questions could be categorized and measured. By utilizing Bloom's taxonomy (Bloom et al., 1956), questions can be assessed on a scale from one to six to reflect their cognitive level, from simple factual queries to more complex analytical or evaluative ones (Plack et al., 2007; Raz et al., 2023; Zheng et al., 2008). Another classification for question types was proposed by Mosher and Hornsby (1966), who suggested a few distinct types of questions that are aligned with different searching strategies and thus may reflect how people gather information through question construction.

Investigating question-asking poses a challenge in conventional laboratory experiments, which often fail to capture its spontaneity and real-world context. Exploring the landscape of interactive digital games reveals a rich tapestry of question-asking dynamics, integral to both gameplay and player engagement (Zhangozha, 2020). These interactive platforms mirror complex aspects of the question-asking process and present an opportunity for deeper analysis and examination in more natural settings.

A notable instance is the Akinator, a globally popular online game where a virtual genie guesses a character the player thinks of by asking a series of questions. Research examining the Akinator's methodology in question formulation has analyzed extensive datasets from numerous game sessions (Sasson & Kenett, 2023). This analysis revealed the potential of leveraging online question-based games to gain insights into the patterns and progression of question-asking, laying a foundational understanding for this study. However, this study was limited in scope due to patent protection of the software. Yet, it highlighted that the Akinator's question asking process does not aim to narrow an information space—a popular theory on the aim of question-asking—and that the questions generated by the Akinator can be characterized into focused, yet time-evolving, topics. However, a critical limitation of this study lies in its focus on questions posed by an AI model, not by humans. This distinction is crucial as it overlooks the natural, spontaneous, and often more nuanced way humans formulate and use questions (but see Hwang et al., 2023).

The utilization of online game environments as a research medium offers a unique avenue for collecting extensive, varied data in ecologically valid settings (Jannai et al., 2023; Pedersen et al., 2023; Rafner et al., 2022, 2023; Vickrey et al., 2008). These environments authentically capture diverse player strategies and creative questioning methods, offering a richer dataset than typically available in standard research settings (Testoni et al., 2023). The variety of approaches seen in players within these games provides a wide spectrum of question-asking styles and strategies.

Current Study

Gamification is the process by which game mechanics are added into various processes, programs, and platforms, to create incentive and engaging experiences (Deterding et al., 2011). One genre of gamification that is particularly relevant for this research is coined “Serious Games”. Serious games are designed with the purpose of impacting the players' real-life thoughts and behaviors in order to fulfil a purpose beyond the self-contained aim of the game (Frasca, 2007; Mitgutsch & Alvarado, 2012). Various studies have proved the effectiveness of games in creating engaging systems to attract audiences (Porat et al., 2020), yet empirical evidence about the impact of these games are rather limited (Giessen, 2015; Mitgutsch & Alvarado, 2012).

In this study, we aim to investigate natural processes of human question-asking and explore their connections to creativity and problem-solving. We introduce 'Spot the Spy' (https://spotthespy.itch.io/spot-the-spy-game) – a hidden object game where players engage with the game through the process of question-asking. Hidden object games are a popular genre in which players must locate specific items or characters within a detailed environment. In 'Spot the Spy', players are tasked with identifying a covert spy in a room full of people, employing strategic questioning to narrow down the suspects. This process of inquiry closely resembles the well-known '20-questions' game, a paradigm that has been extensively studied for its insights into question formulation and information gathering strategies (Courage, 1989; Ruggeri et al., 2016, 2019; Sasson & Kenett, 2023; Siegler, 1977; Testoni et al., 2023). In both scenarios, the player's success hinges on their ability to ask effective, narrowing questions that progressively reduce the field of possibilities, thereby demonstrating the intricate relationship between question-asking, creativity, and problem-solving.

Within the context of question-asking strategies, we anticipate that players will adopt a searching strategy that mirrors the Bloom's taxonomy (1956) of question complexity. We hypothesize that players will start with broader and lower-level questions, gradually refining their inquiries as they advance through the game. This evolution from comprehensive queries to highly specific ones is envisaged as an adaptive searching strategy, facilitating players in the efficient elimination of suspects, and ultimately identifying the elusive spy.

Finally, we posit that a player's degree of creative thinking will significantly influence their question-asking behavior. Our secondary hypothesis suggests that individuals with a higher creative aptitude will exhibit a more diverse spectrum of questions, spanning various cognitive levels, compared to those posed by participants with a lower creative inclination.

Methods

Participants

120 individuals recruited through Prolific Academic participated in the study for a compensation of £4. All participants were native English speakers from the United States and the United Kingdom. Participants who did not adhere to the game's instructions were excluded from the analysis. Consequently, the final analyzed sample consisted of 103 participants (50.5% female, 49.5% male; M_{age} = 36.8 years, SD = 12.3 years; mean years of formal education = 15.8 years, SD = 4.5 years). The study received ethical approval from our institution, and all participants provided informed consent prior to their inclusion.

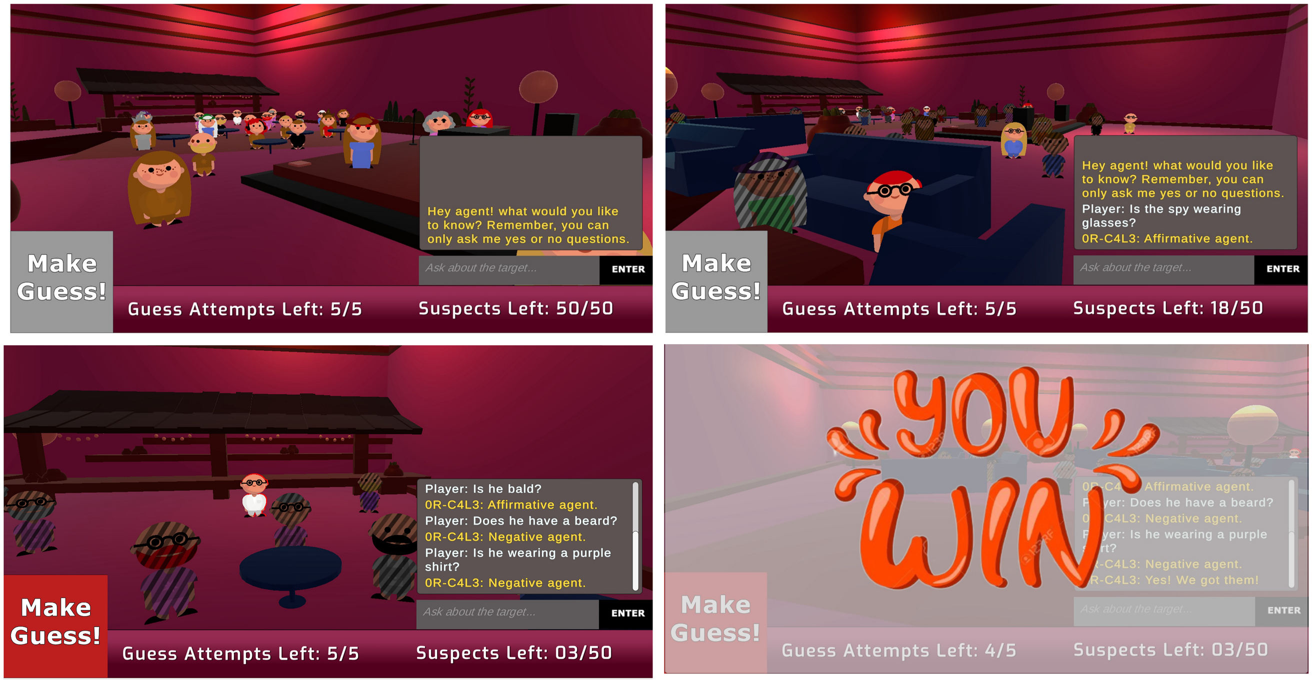

Figure 1: (top left) Opening screen of the game. The player sees a room full of suspects and can wander around the room to explore them. (top right) The player asks questions about the spy to eliminate suspects. Right-clicking on suspects blacks them out to show they are eliminated. In this scene, the player finds out the spy is wearing glasses and thus eliminates all the suspects who are wearing glasses. (bottom left) The player double-clicks a suspect to make a guess. Clicking the "Make Guess!" button reveals if this is the spy. (bottom right) When the player makes a right guess, a winning screen is displayed.

Spot the Spy

Game Design. The game starts with a scene of a room full of characters who differ from each other by a set of visual features such as gender, garments, hairstyles, and accessories. One of the characters is randomly selected at the beginning of the game as the “spy”, and until otherwise known, all characters in the room are considered suspects. The player can explore the different characters while wandering around the room and zooming in on various parts. Then, the player can ask a chatbot agent (see below) yes or no questions about the target (i.e., “Does the spy have gray hair?”), while eliminating suspects based on the answers. When the player arrives at a guess, he can choose a character and click the “Guess” button to get an indication of whether the guess is correct. If the player was right the game ends, otherwise the player can keep asking questions and making guesses, up to 5, until he identifies the spy (**Figure 1**).

An integral component of the game is the interactive chatbot agent, powered by GPT-3.5 Turbo. The chatbot assists players by providing responses to their questions, based on a pre-set list of details about the suspects. The chatbot is prompted to only respond with 'Affirmative agent,' 'Negative agent,' or 'I cannot give the answer to that question' in cases where a yes or no answer isn't applicable.

The game is hosted on the web platform itch.io, ensuring convenient access for players. As they play the game, we collect their questions, answers given to these questions by the chatbot, guesses given by the player and whether they were correct, the time it took the player to ask questions, and the progress of the player after each question (e.g., how many characters he ruled out). Finally, we use a recently developed tool—based on a trained language model—to automatically score the Bloom complexity levels of questions asked during the game (Raz et al., 2024).

Question Effectiveness Measure (QEM). In addition, we introduce a new metric, the Question Effectiveness Measure (QEM), designed to quantitatively evaluate the impact of questions posed by players in narrowing down potential suspects. The QEM considers both the reduction in the pool of suspects following each question and the strategic sequence in which these questions are asked. In developing the QEM, we align with insights from Nelson (2005), who discusses various norms for assessing question usefulness, such as Bayesian diagnosticity and information gain (see also Nelson et al., 2010). Our approach resonates with the concept of expected stepwise information gain, as outlined by Ruggeri et al. (2017), where the informativeness of questions is measured by the reduction of entropy, thereby moving from uncertainty towards certainty. Furthermore, Testoni et al. (2023) emphasize the importance of entropy reduction in effective question-asking, particularly in scenarios where the goal is to efficiently narrow down a set of possibilities, underscoring the relevance of QEM in the context of strategic questioning in gameplay.

Moreover, the impact of object categorization levels on question-asking efficiency provides additional insights into our approach (Mosher & Hornsby, 1966; Ruggeri & Feufel, 2015). This perspective is particularly relevant to our QEM's focus on reducing the number of suspects, similar to decreasing entropy in a set of hypotheses.

The QEM can be formulated as follows:

$$QEM(q_i) = \frac{Number\ of\ Suspects\ Eliminated}{Total\ Number\ of\ Suspects}$$

$$QEM = \sum_{i=1}^{n} \frac{1}{i} \cdot QEM(q_i)$$

where q_i represents the i-th question in the game, with the count of suspects eliminated after each question reflecting the information gained. The Total Number of Suspects includes all suspects, amounting to a total of 50 suspects in this study. The formula for calculating the final QEM score assigns greater significance to the effectiveness of the initial questions in the sequence by applying a weight of $1/i$, where i is the position of the question in the sequence, thereby underscoring the strategic value of the first few inquiries in influencing the overall QEM score.

The QEM aims to quantify the collective efficiency of a series of questions in isolating the target, in this case, identifying the spy. It reflects both the volume of suspects eliminated and the tactical arrangement of the questions. Higher QEM values signify a more impactful sequence of questions, leading to a streamlined and effective identification process.

Alternative Uses Tasks (AUT)

The AUT assesses divergent thinking, a key component of creativity (Acar & Runco, 2019; Guilford, 1967). The AUT requires participants to think of as many creative, unusual, or original uses as possible for a common object within a limited time frame. In this case, participants were asked to think of uses for a broom and a belt. They had two minutes for each object and could provide up to 30 uses for each object. The objective creativity scores of participants' responses were computed automatically using the Maximum Associative Distance (MAD) (Yu et al., 2023), by taking the maximal semantic distance between a response to the cue word—a measure shown to strongly and quantitatively capture subjective ratings of response originality. In addition, we measure AUT fluency as the average number of responses generated by the participant to the object.

Alternative Questions Task (AQT)

Similar to the AUT, in the AQT participants are provided with common objects and are asked to come up with as many original and creative questions about them as they can (Raz et al., 2023). The AQT has been shown to be positively correlated with creative thinking. In this study, participants were asked to think of questions about a pencil and a pillow. They had two minutes for each object and could provide up to 30 uses for each. To score the AQT responses, we compute participants MAD and fluency scores, and a quantitative Bloom score of each response to reflect questions' complexity (Raz et al., 2024).

Intelligence Measures

Participants underwent assessment of fluid intelligence (Gf) and broad retrieval ability (Gr). To measure *Gf*, we employed a series completion task drawn from the Culture Fair Intelligence Test. This task involved identifying the next item in a sequence of three evolving images (small line drawings) over 3 minutes, encompassing 16 items (Cattell & Cattell, 1961). The total number of correct answers forms the *Gf* score. To measure *Gr*, we implemented a category fluency task, where participants listed as many animals as possible in a two minute period, a method previously utilized by Ardila et al. (2006). This approach allowed us to gain a nuanced understanding of the participants' cognitive capabilities in areas crucial for our study.

Forward Flow

We included the Forward Flow task, as conceptualized by Gray et al. (2019). In this task, participants are initially given a cue word and are instructed to spontaneously generate the first word that comes to their mind. Subsequently, they continue this process, producing a series of words where each new word is associated with the preceding one, thus forming a chain of free associations. Computational models are used to quantify participants' breadth of associative search (Beaty et al., 2021). Critically, broader associative ability—as measured via forward flow—has been implicated as critical in creative thinking (Beaty & Kenett, 2023). Here, we used the cue words "bear" and "table", and participants had to complete a sequence of 10 words. This task is particularly valuable in evaluating the fluidity and connectivity of thought, essential components of creative cognition.

Curiosity

Participants completed a 22-item questionnaire designed to measure different facets of curiosity. This questionnaire was compiled from previously established questionnaires that encompass various curiosity dimensions, including social curiosity, workplace or organizational curiosity, and I/D-type curiosity which relates to intellectual cognition, information-seeking behavior, and perceptual curiosity (Collins et al., 2004; Kashdan et al., 2020; Litman, 2008; Litman & Jimerson, 2004; Litman & Spielberger, 2003). Participants were asked to rate the extent to which they agree with curiosity statements in relation to their own experiences and tendencies. To quantify the responses, each answer was scored on a scale from -2 (complete disagreement) to 2 (complete agreement), and these scores were then summed to yield an overall curiosity score for each participant.

Procedure

Participants began the study by providing informed consent, followed by completing a series of tasks on a Qualtrics questionnaire, including the AUT, AQT, fluency task, forward flow task, intelligence assessment, and a curiosity questionnaire. They also provided demographic information. Attention checks were interspersed throughout the questionnaire to ensure engagement and accuracy. After completing these tasks, participants were directed to the 'Spot the Spy' game's website for one gameplay session, until either winning or losing the game.

Results

In assessing the outcomes of the 'Spot the Spy' game, 72% of participants successfully identified the spy, demonstrating a high level of proficiency in strategic thinking and problem-solving within the game's framework. This success rate not only highlights the effective cognitive strategies employed by the majority of players but also suggests that the game's instructions and setup were clear and well-conceived, facilitating player engagement and understanding.

Examining gameplay strategies revealed that, on average, participants posed approximately 8.71 questions (SD = 5.59) throughout their gameplay. This high standard deviation indicates a diversity in the questioning approach, with some participants asking more questions than others. Regarding the number of guesses made, among those who won the game, the average was 1.84 guesses (SD = 1.20). This metric provides insight into the precision of players' final decisions, indicating that most winners were able to accurately identify the spy with minimal guesswork. The relatively low standard deviation points to a consistent pattern of careful and calculated guessing among the successful participants.

A particularly interesting finding emerged when analyzing the number of suspects eliminated by participants. The average number of suspects eliminated was 30.07 suspects (SD = 19.92). However, a deeper look into this data uncovered a distinct pattern based on the game's outcome. Players who eliminated more than 40 suspects consistently won the game, suggesting a robust and comprehensive approach to how narrowing down suspects is linked to higher success rates. Conversely, most participants who lost the game eliminated less than 10 suspects. This disparity suggests that a more thorough and persistent investigative process may be key to succeeding in the game. It is unclear whether players who did not eliminate suspects chose to not use this tool or did not understand how to use it.

To further analyze this pattern, we conducted a logistic regression analysis (**Figure 2**). Given the imbalance in our dataset (72% winning rate), we utilized the Synthetic Minority Over-sampling Technique (SMOTE) to balance the class distribution (Chawla et al., 2002; Fernandez et al., 2018). SMOTE is an effective approach for handling imbalanced datasets by artificially generating synthetic samples for the minority class. In our case, this technique was necessary to mitigate the bias toward the majority class (winners) and ensure a more reliable and generalizable model. After applying SMOTE, the class distribution was balanced with equal counts for both 'win' and 'loss' labels in the training set, which significantly improved the model's predictive ability. The balanced model achieved an accuracy of approximately 95%, demonstrating its effectiveness in predicting game outcomes. In terms of precision, recall, and F1-Score, the model exhibited a high precision (0.80) and perfect recall (1.00) for predicting losses, along with perfect precision (1.00) and a recall of 0.94 for wins. These metrics indicate a balanced performance between precision and recall for both outcomes, with F1-Scores of 0.89 for losses and 0.97 for wins. Overall, the logistic regression analysis provided deeper insights into the patterns of gameplay success, reinforcing our earlier observations about the importance of strategic questioning and problem-solving in 'Spot the Spy' using the ability to visually eliminate suspects.

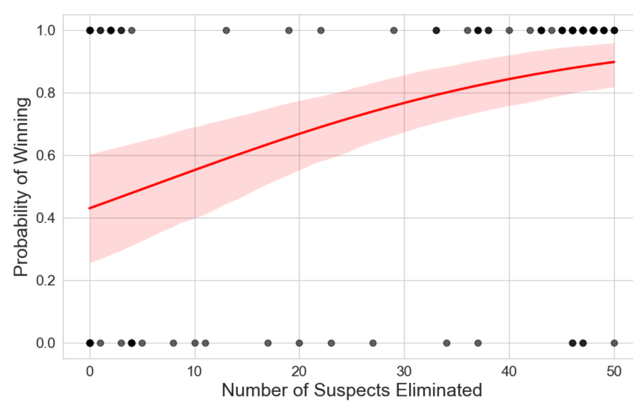

Figure 2: Logistic regression of winning on number of suspects eliminated. The probability of winning is higher when eliminating more suspects.

Correlations of Study Variables

Next, we computed a Pearson's correlation matrix to elucidate the interplay between variables both from the questionnaire and from the game (**Figure 3**). The QEM showed a significant negative correlation with the number of guesses players made in the game, $r = -.43$, $p < .001$, indicating that players who asked more effective questions tended to make fewer guesses to correctly identify the spy. This suggests that effective questioning is a key component in the game's problem-solving process. Furthermore, the QEM showed a significant positive correlation with the intelligence score, $r = .41$, $p < .001$, indicating that players with higher general intelligence as facilitated in the task tended to ask more effective questions. Additionally, there was a notable correlation between the intelligence score and the number of suspects eliminated during the game, $r = .41$, $p < .001$, reinforcing the idea that higher intelligence is associated with more efficient gameplay strategies.

Gr was also significantly correlated with the QEM, $r = .33$, $p < .001$, indicating that *Gr* contributes to the effectiveness of questioning in the game. A negative correlation was observed between *Gr* and number of guesses, $r = -.27$, $p = .006$, suggesting that players with higher *Gr* may proceed with greater caution, requiring fewer guesses to identify the spy. There was a significant correlation between *Gr* and number of responses in both the AUT, $r = .38$, $p < .001$, and the AQT, $r = .40$, $p < .001$. This finding indicates that participants who showed divergent thinking also tended to exhibit a similar proficiency in creating diverse responses in the AUT and AQT, pointing to a broader skill in creative ideation.

Interestingly, the curiosity score revealed a significant negative correlation with the game Bloom's score, $r = -.22$, $p = .023$, hinting that higher curiosity might be associated with different strategic approaches, possibly reflecting more explorative behavior. Additionally, a significant positive correlation was found between the AUT fluency and the curiosity score, $r = .31$, $p = .002$, and between the AQT fluency and the curiosity score, $r = .23$, $p = .023$, indicating that individuals who are more fluent in creative tasks tend to exhibit higher levels of curiosity.

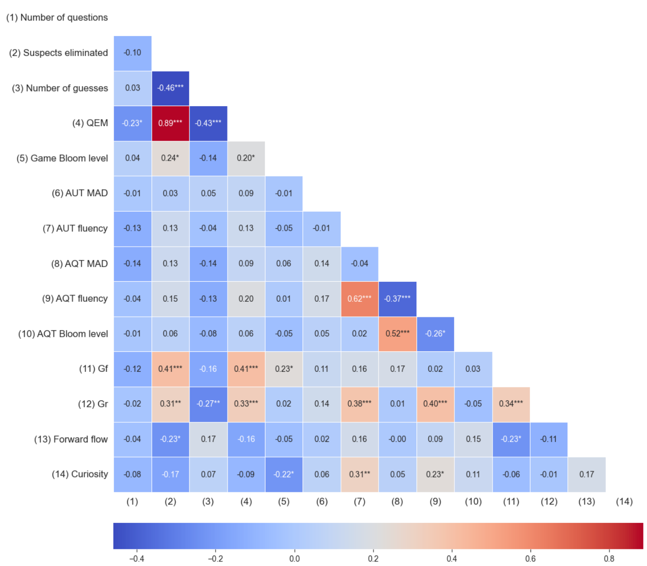

Figure 3: Pearson-correlations matrix of all variables.

Discussion

The practice of question-asking, an essential component of human interactions, plays a pivotal role in fostering engagement, stimulating thought processes, and driving curiosity. Its importance spans over various domains, from educational settings to everyday decision-making, emphasizing its relevance in understanding human cognition and creativity. This study delves deeper into the dynamics of question-asking, particularly in the context of interactive gaming, which mimics real-life scenarios by requiring strategic inquiry and problem-solving.

The primary goal of this study was to examine the intricate relationship between question-asking and cognitive abilities, such as creativity. To achieve this, we designed the interactive game 'Spot the Spy,' where participants engage in a hidden object game necessitating strategic questioning to identify a covert spy among many characters. This exploratory analysis sought to quantify and analyze the effectiveness of the questions posed by players and to correlate this with their cognitive abilities.

Our findings revealed several noteworthy patterns, providing valuable insights into the cognitive processes leading to effective question-asking in problem-solving contexts. A significant majority of participants (72%) successfully identified the spy, demonstrating a high level of proficiency in strategic thinking and questioning. Correlations between various cognitive measures and gameplay strategies were observed, such as the relationship between intelligence scores, the effectiveness of the questions (QEM), and the number of suspects eliminated during the gameplay. Moreover, the number of suspects eliminated predicts with high accuracy the game outcome.

Yet, the study entails several limitations. Being exploratory in nature, the research primarily serves as a starting point for more in-depth investigations. The sample size, while adequate for preliminary analysis, could be expanded in future studies to enhance the generalizability of the findings. Also, the game environment, though designed to mimic real-life problem-solving scenarios, still possesses inherent limitations of artificial settings. Future research could focus on diversifying the contexts in which question-asking is analyzed, including more varied real-world scenarios.

Conclusion

This study represents a significant step forward in elucidating the role of question-asking in cognitive processes, particularly within the context of interactive gaming. By developing and utilizing the innovative 'Spot the Spy' game, we were able to quantitatively analyze the relationship between question-asking strategies and various cognitive abilities in natural context, “outside of the lab”.

Our findings not only underscore the intricate connection between creativity, problem-solving, and question asking, but also serve as a proof-of-concept for the potential of intweractive gaming as a tool for cognitive research. The success of 'Spot the Spy' in eliciting meaningful data on question-asking strategies demonstrates the viability of using game-based environments for investigating complex cognitive abilities. Future research is needed to further unravel the complexities of question-asking. This exploratory analysis paves the way for a deeper understanding of the cognitive underpinnings of one of our most fundamental human abilities – asking questions, opening doors to more extensive and varied investigations in this field.

Acknowledgements

We thank Shahar Hod, Rotem Schwartz, and Elad Kalif for their help in conceptualizing, designing, and developing the game.

References

Abdulla, A. M., Paek, S. H., Cramond, B., & Runco, M. A. (2020). Problem finding and creativity: A meta-analytic review. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(1), 3–14. https://doi.org/10.1037/aca0000194

Acar, S., Berthiaume, K., & Johnson, R. (2023). What kind of questions do creative people ask? *Journal of Creativity*, *33*(3), 100062. https://doi.org/10.1016/j.yjoc.2023.100062

Acar, S., & Runco, M. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 153–158. https://doi.org/10.1037/aca0000231

Albergaria-Almeida, P. (2011). Critical thinking, questioning and creativity as components of intelligence. *Procedia - Social and Behavioral Sciences*, *30*, 357–362. https://doi.org/10.1016/j.sbspro.2011.10.070

Ardila, A., Ostrosky-Solís, F., & Bernal, B. (2006). Cognitive testing toward the future: The example of Semantic Verbal Fluency (ANIMALS). *International Journal of Psychology*, *41*(5), 324–332. https://doi.org/10.1080/00207590500345542

Arreola, N. J., & Reiter-Palmon, R. (2016). The effect of problem construction creativity on solution creativity across multiple everyday problems. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(3), 287–295. https://doi.org/10.1037/a0040389

Beaty, R. E., & Kenett, Y. N. (2023). Associative thinking at the core of creativity. *Trends in Cognitive Sciences*, *27*(7), 671–683. https://doi.org/10.1016/j.tics.2023.04.004

Beaty, R. E., Zeitlen, D. C., Baker, B. S., & Kenett, Y. N. (2021). Forward flow and creative thought: Assessing associative cognition and its role in divergent thinking. *Thinking Skills and Creativity*, *41*(100859). https://doi.org/10.1016/j.tsc.2021.100859

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of educational objectives: The classification of educational goals*. David McKay Company.

Cattell, R. B., & Cattell, A. K. S. (1961). *Measuring intelligence with the culture Fair Tests*. Hogrefe.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, *26*(5), 1548–1587. https://doi.org/10.3758/s13423-018-1470-5

Collins, R. P., Litman, J. A., & Spielberger, C. D. (2004). The measurement of perceptual curiosity. *Personality and Individual Differences*, *36*(5), 1127–1141. https://doi.org/10.1016/S0191-8869(03)00205-8

Courage, M. L. (1989). Children's inquiry strategies in referential communication and in the game of twenty questions. *Child Development*, *60*(4), 877–886. https://doi.org/10.2307/1131029

Crupi, V., Nelson, J. D., Meder, B., Cevolani, G., & Tentori, K. (2018). Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. *Cognitive Science*, *42*(5), 1410–1456. https://doi.org/10.1111/cogs.12613

De Simone, C., & Ruggeri, A. (2021). What is a good question asker better at? From unsystematic generalization to adult-like selectivity across childhood. *Cognitive Development*, *59*, 101082. https://doi.org/10.1016/j.cogdev.2021.101082

Deterding, S., Khaled, R., Nacke, L., & Dixon, D. (2011). Gamification: Toward a definition. *CHI 2011 Gamification Proceedings*, 12, 1-79.

Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. https://doi.org/10.1613/jair.1.11192

Frasca, G. (2007). *Play the message: Play, game and videogame rhetoric* [Unpublished PhD dissertation. IT University of Copenhagen. Denmark].

Giessen, H. W. (2015). Serious games effects: An overview. *Procedia - Social and Behavioral Sciences*, *174*, 2240–2244. https://doi.org/10.1016/j.sbspro.2015.01.881

Gottlieb, J. (2021). The effort of asking good questions. *Nature Human Behaviour*, *5*(7). https://doi.org/10.1038/s41562-021-01132-6

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). "Forward Flow": A New Measure to Quantify Free Thought and Predict Creativity. *American Psychologist*, 74(5), 539-554. hhttps://doi.org/10.1037/amp0000391

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, *1*(1), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Hu, W., Shi, Q. Z., Han, Q., Wang, X., & Adey, P. (2010). Creative scientific problem finding and its developmental trend. *Creativity Research Journal*, *22*(1), 46–52. https://doi.org/10.1080/10400410903579551

Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology*, *113*(3), 430–452. https://doi.org/10.1037/pspi0000097

Hwang, K., Challagundla, S., Alomair, M. M., Chen, L. K., & Choa, F.-S. (2023). Towards AI-assisted multiple choice question generation and quality evaluation at scale: Aligning with Bloom's taxonomy. *NeurIPS'23 Workshop on Generative AI for Education (GAIED)*, 1--8.

Jannai, D., Meron, A., Lenz, B., Levine, Y., & Shoham, Y. (2023). Human or not? A gamified approach to the Turing test. *ArXiv*. http://arxiv.org/abs/2305.20010

Kashdan, T. B., Disabato, D. J., Goodman, F. R., & McKnight, P. E. (2020). The five-dimensional curiosity scale revised (5DCR): Briefer subscales while separating overt and covert social curiosity. *Personality and Individual Differences*, *157*, 109836. https://doi.org/10.1016/j.paid.2020.109836

Kenett, Y. N., Humphries, S., & Chatterjee, A. (2023). A thirst for knowledge: Grounding curiosity, creativity, and aesthetics in memory and reward neural systems. *Creativity Research Journal*, *35*(3), 412–426. https://doi.org/10.1080/10400419.2023.2165748

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, *44*(7), 1585–1595. https://doi.org/10.1016/j.paid.2008.01.014

Litman, J. A., & Jimerson, T. L. (2004). The measurement of curiosity as a feeling of deprivation. *Journal of Personality Assessment*, *82*(2), 147–157. https://doi.org/10.1207/s15327752jpa8202_3

Litman, J. A., & Spielberger, C. D. (2003). Measuring Epistemic Curiosity and its diversive and specific components. *Journal of Personality Assessment*, *80*(1), 75–86. https://doi.org/10.1207/S15327752JPA8001_16

Mitgutsch, K., & Alvarado, N. (2012). Purposeful by design?: A serious game design assessment framework. *Proceedings of the International Conference on the Foundations of Digital Games, 121-128.*. https://doi.org/10.1145/2282338.2282364

Mosher, F. A., & Hornsby, J. R. (1966). *Studies in cognitive growth*. Wiley.

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4), 979–999. https://doi.org/10.1037/0033-295X.112.4.979

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science*, *21*(7), 960–969. https://doi.org/10.1177/0956797610372637

Pedersen, M. K., Díaz, C. M. C., Wang, Q. J., Alba-Marrugo, M. A., Amidi, A., Basaiawmoit, R. V., Bergenholtz, C., Christiansen, M. H., Gajdacz, M., Hertwig, R., Ishkhanyan, B., Klyver, K., Ladegaard, N., Mathiasen, K., Parsons, C., Rafner, J., Villadsen, A. R., Wallentin, M., Zana, B., & Sherson, J. F. (2023). Measuring cognitive abilities in the wild: Validating a population-scale game-based cognitive assessment. *Cognitive Science*, *47*(6). https://doi.org/10.1111/cogs.13308

Plack, M. M., Driscoll, M., Marquez, M., Cuppernull, L., Maring, J., & Greenberg, L. (2007). Assessing reflective writing on a pediatric clerkship by using a modified Bloom's taxonomy. *Ambulatory Pediatrics*, *7*(4), 285–291. https://doi.org/10.1016/j.ambp.2007.04.006

Porat, R., Erel, L., Pnueli, V., & Halperin, E. (2020). Developing ReApp: An emotion regulation mobile intervention for intergroup conflict. *Cognition and Emotion*, *34*(7), 1326–1342. https://doi.org/10.1080/02699931.2020.1747400

Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T., & Sherson, J. (2023). Creativity in the age of generative AI. *Nature Human Behaviour*, *7*(11). https://doi.org/10.1038/s41562-023-01751-1

Rafner, J., Biskjaer, M. M., Zana, B., Langsford, S., Bergenholtz, C., Rahimi, S., Carugati, A., Noy, L., & Sherson, J. (2022). Digital games for creativity assessment: Strengths, Weaknesses and opportunities. *Creativity Research Journal*, *34*(1), 28–54. https://doi.org/10.1080/10400419.2021.1971447

Raz, T., Luchini, S., Beaty, R. E., & Kenett, Y. N. (2024). Automated scoring of question complexity: A large language model approac*h*. *ResearchSquare.* https://doi.org/10.21203/rs.3.rs-3890828/v1

Raz, T., Reiter-Palmon, R., & Kenett, Y. N. (2023). The role of asking more complex questions in creative thinking. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000658

Rothe, A., Lake, B. M., & Gureckis, T. M. (2018). Do people ask good questions? *Computational Brain & Behavior*, *1*(1), 69–89. https://doi.org/10.1007/s42113-018-0005-5

Ruggeri, A., & Feufel, M. (2015). How basic-level objects facilitate question-asking in a categorization task. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00918

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*(12), 2159–2173. https://doi.org/10.1037/dev0000240

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Psychology*, *53*(9), 1620–1632. https://doi.org/10.1037/dev0000340

Ruggeri, A., Xu, F., & Lombrozo, T. (2019). Effects of explanation on children's question asking. *Cognition*, *191*, 103966. https://doi.org/10.1016/j.cognition.2019.05.003

Sasson, G., & Kenett, Y. N. (2023). A mirror to human question asking: Analyzing the Akinator online question game. *Big Data and Cognitive Computing*, *7*(1), Article 1. https://doi.org/10.3390/bdcc7010026

Siegler, R. S. (1977). The twenty questions game as a form of problem solving. *Child Development*, *48*(2), 395–403. https://doi.org/10.2307/1128632

Testoni, A., Bernardi, R., & Ruggeri, A. (2023). The efficiency of question-asking strategies in a real-world visual search task. *Cognitive Science*, *47*(12), e13396. https://doi.org/10.1111/cogs.13396

Torrance, E. P. (1987). Teaching for creativity. *Frontiers of Creativity Research: Beyond the Basics*, *189*, 215.

Vickrey, D., Bronzan, A., Choi, W., Kumar, A., Turner-Maier, J., Wang, A., & Koller, D. (2008). Online word games for semantic data collection. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 533–542. https://aclanthology.org/D08-1056

Wang, Z., & Lake, B. M. (2019). Modeling question asking using neural program generation. *ArXiv*. https://arxiv.org/abs/1907.09899v4

Yu, Y., Beaty, R., Forthmann, B., Beeman, M., Cruz, J. H., & Johnson, D. R. (2023). A MAD method to assess idea novelty: Improving validity of automatic scoring using maximum associative distance (MAD*)*. P*sychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000573

Zhangozha, A. R. (2020). On techniques of expert systems on the example of the Akinator program. *Artificial Intelligence Scientific Journa*l, 25(2), 7-13.

Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy debunks the "MCAT myth." *Science*, *319*(5862), 414–415. https://doi.org/10.1126/science.1147852