Proceedings of the Nineteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2023)

Investigating the Influence of Behaviors and Dialogs on Player Enjoyment in Stealth Games

Wael Al Enezi, Clark Verbrugge

School of Computer Science, McGill University, Montréal, Canada
wael.alenezi@mail.mcgill.ca, clump@cs.mcgill.ca

Abstract

The player's perception of AI behavior significantly influences their overall game experience. This perception is shaped by both interactive encounters and careful observations, particularly in genres like stealth, where gameplay revolves around planning strategies based on AI enemy movement.

This paper aims to derive general insights into the player experience concerning two crucial gameplay elements that impact the perception of NPC intelligence. The first element pertains to the actual behavior of opponent NPCs, while the second focuses on the dialogues employed to highlight NPC decision-making. We conducted a user study to assess whether players can discern between complex and simple NPC behavior during gameplay in a specific scenario of a top-down stealth game prototype. We introduced variations in spoken dialogs to determine their effect on player perception. In the end, our findings revealed that when simple dialogs were used, players derived greater enjoyment from a more complex AI behavior. However, using contextual dialog allowed a simple behavior to match a complex behavior in player enjoyment.

Introduction

It is widely known that creating an efficient or unbeatable opponent behavior in games does not always positively contribute to player enjoyment (Wetzel and Anderson 2017; Livingstone 2006). At the same time, players enjoy playing against NPCs that show intelligent or human-like behavior (Soni and Hingston 2008), and thus some evidence of intelligent decision-making is desirable. Common methods to convey NPC intelligence to players include both creating intricate NPC behaviors with reasonable efficiency that intelligence would be naturally inferred and artificially foregrounding NPC decisions through visual or oral artifacts, such as by having NPC's "bark" dialog that correlates with their internal perceptions or game state (Redding 2009). This paper explores the relationship between these two features on the player's experience and whether one is more essential for improving it. In other words, is it worth investing in the algorithmic effort and extra computation time to create complex NPC behavior, or is it better to cover simple behavior with dialog lines that give the illusion of intelligence?

Our investigation is based on a user study conducted on a stealth game scenario where the player must explore a space while avoiding being captured by opponent NPCs. NPC intelligence is thus required to effectively search the area and best exploit prior observations of a fleeing player to complete the capture. Such a scenario presents an exciting challenge for players to outsmart NPCs, with difficulty and a sense of fairness or realism depending on how well and naturally NPCs search. Hard-coded search paths and "cheating" approaches (allowing NPCs knowledge of the exact player location) are simple strategies to implement but may easily be perceived as unnatural. More complex NPC AI that uses dynamic, local information, however, implies a more significant design and implementation effort, but this may or may not be apparent to the player. Both approaches may be mitigated by using relevant NPC dialog that gives the appearance of intelligence.

The major contribution of this work is a user study to evaluate the effect of NPC behaviors and dialog lines on player perception. The user study was based on playing the game we modified and survey data. We consider three forms of NPC behavior for searching for an intruder and two forms of dialog lines with different levels of contextual information. We divide the users of this study into two groups; each group compares a simple NPC behavior and a heuristic and more effective behavior based on a relatively novel, geometry-aware path prediction (Al Enezi and Verbrugge 2021).

Related Work

Literature showed that players typically find NPCs, perceived to act more humanly, are more fun to play against (Soni and Hingston 2008). Consequently, there are several attempts to evaluate players' perception of AI behaviors, for example, in the form of competitions aimed at testing human-like opponents (Togelius 2014), which can be described as a "Turing test" for game agents (Turing 2009). 2K BotPrize Contest was held to evaluate bots for Unreal Tournament 2004, a First-Person shooter game, human players were asked to judge whether they played against a human or a bot. None of the bots passed the test; however, those who showed a certain level of error and randomness managed to confuse humans in their judgment (Hingston 2009).

Another study approached this problem from a different angle. Bots and human players were asked to predict player positions on a simplified top-down presentation of Counter-strike. Results showed that, by using a particle filter and hidden semi-Markov models, players and bots made more or less the same correct and wrong judgements (Hladky and Bulitko 2008). This, in theory, should pave the way for creating a more human-like behavior. On the other hand, a survey conducted with video game professionals revealed that some expressed apprehension that players may not perceive the increased complexity of NPCs or comprehend their behavior easily. This phenomenon is commonly referred to as "the black hole of AI" (Johansson, Eladhari, and Verhagen 2012). This highlights the importance of evaluating the limits of this phenomenon. A number of studies have investigated different aspects of NPC behavior in modern commercial games that may impact a player's enjoyment.

Lankoski & Björk propose several relevant design patterns for NPC believability, including evidence of intentionality, which they explored through an analysis of a single character from *Oblivion* (Lankoski and Björk 2007). Warpefelt & Strååt consider the issue through an analysis of NPC behavior in a number of RPG and FPS games, observing that immersion is weakened by failing to properly limit NPC knowledge and ensure environmental awareness (Warpefelt and Strååt 2013).

Several studies have also tried to address player enjoyment directly. For example, several works have explored the possibility of refining games by defining mathematical models that determine the progress of a game. The goal is to define the progress model of a game to adjust player enjoyment through the "acceleration" in the game progress model in sport and board games (Sutiono, Purwarianti, and Iida 2014; Sutiono et al. 2015; Iida, Takeshita, and Yoshimura 2003). Player enjoyment was also studied in NPC AI for several game genres, for example, in turn-based strategy games (Wetzel and Anderson 2017), board games (Iida, Takeshita, and Yoshimura 2003), and first-person shooter (Soni and Hingston 2008; Hingston 2009).

Other studies have attempted to use learning NPCs to predict player enjoyment by defining interest value determined by a set of metrics, where player enjoyment is determined by questioning players after playing each NPC behavior variation (Yannakakis and Hallam 2007b,a). Their study focused on a modified version of the same issue within a more limited setting, resembling a Pacman-like game. It aimed to evaluate the accuracy of specific quantitative measures correlated with properties of fun, as part of exploring a quantitative metric for enjoyment. Our work, in contrast, is aimed at investigating whether certain aspects or combinations of AI development and design were recognizable and noticeably preferable to players.

While NPC behavior directly impacts player experience, many other factors can affect that experience. Dialog lines or "barks" are widely used in commercial games to improve the game experience (Redding 2009). These dialog lines are usually a pre-scripted set of lines that are played to players in certain events. While several studies examined the possibility of generating dialog by using natural language generation (Brusk et al. 2007; Kacmarcik 2006), the literature lacks studies that investigate the association between the nature of information conveyed in the dialog and NPC behavior and how these two factors affect the game experience.

Methodology

This section describes our method for assessing the player perception of two main features in stealth scenarios: NPC search behavior and NPC dialog lines. We aim to identify how these two variables impact players' perceptions of NPCs. We investigate the perceived enjoyability and difficulty and report the possible features that affected these aspects based on participants' feedback.

In our study, each participant plays a set of game episodes. In each episode, the player plays against a team of NPCs. Each NPC is provided with a Field of View (FOV) that represents the scope of its visual senses. The goal of NPCs is to maintain the player in their FOV, and the purpose of the player is to increase their score by staying out of the NPCs' FOV. To encourage the player to move and ensure NPCs exhibit a mix of chasing and searching behaviors, the player's score is also improved by collecting coins that spawn in the area. Figure 1 shows a screenshot of what players see while playing the game. In general, the main components of an episode are:

- Map: The map structure must offer some amount of challenge to both players and NPC guards. High connectivity with multiple paths between locations in the map makes it easier for players to evade guards by effectively cycling around obstacles, while too many dead-ends make it difficult for players to escape. We thus chose the "skeld" map from the game Among Us[1]. This map has the advantage of not overly favoring either structure: it has a few major cycles, which gives players multiple paths to reach any place on the map, but is also organized in the form of rooms, most of which are dead-ends, which makes it more challenging for players. We labeled several locations on the map to allow players to associate certain regions with a name to make them more recognizable. In addition, to make the NPC movement observable at all times, we fitted the whole map in one frame, as seen in figure 1.
- Guards: These are NPCs the players play against. The number of guards, the size of their FOVs, and relative speeds also affect the challenge. Based on initial prototyping, we selected a team of 4 guards, each with a 90° FOV with a short, finite range. They all had the same speed. Guards are tasked with chasing the player when the player is in the FOV of at least one of the guards. They all receive commands from a central behavior manager. When the player is out of sight, the guards follow one of the search behaviors defined in this study. In addition, these guards will announce occasional dialog lines that we described. Each team of guards had one search behavior and one type of dialog variation. To differentiate each team, we gave them a color, so players could

[1] https://www.innersloth.com/games/among-us/

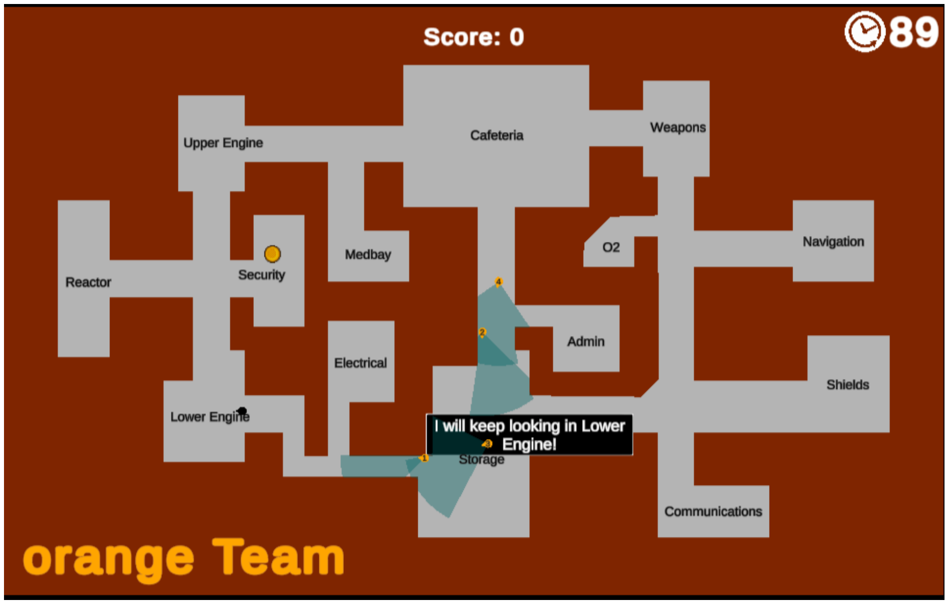


Figure 1: A screenshot of the game the participants played. The player (black disk) is in the "Lower Engine" room, while four guards are in the center area, each with a cone-shaped, translucent FOV. The orange disk in the "Security" room is a coin. In addition, dialog lines are announced visually and verbally to players using the Text-To-Speech function.

easily remember them. To prevent any biases, we randomly assigned the colors and shuffled the order of the teams for each session.

- Intruder: This is the character the player controls, represented as a black circle. To give an advantage to players, we set the player movement speed to be 1.5 times the speed of a guard. To reduce variation, at the start of the episode, the player is always randomly placed in front of one of the guards. The player is tasked with escaping and subsequently avoiding the guards while also collecting coins spawned on the map. The player has a score, which increases when a coin is picked up, and gradually decreases when any of the guards spot them.

- Coins: Once the player is out of the guards' sight, a coin is randomly placed, ensuring that it spawns further than a fixed distance from the current player's position. As the player collects the coin, they increase their score, and another coin is randomly spawned on the map. However, if at any point the player is spotted, any spawned coin will disappear and randomly reappear once the player is out of sight.

- Time: Each episode is set for 99 seconds, a duration sufficient to give players enough time to observe and form ideas of the guard behavior while not exhausting player attention over the multiple game-plays required in our study. Furthermore, to give players a goal, we allowed the game to finish once they achieved a score of 100, which is still challenging to achieve since it meant players had to collect at least five coins to reach that score.

Guard Team's Variables

For each guard team, we defined two main variables that determine their behavior. The first determines how guards move in space, and the second determines the set of predefined "barks".

Guard Search Behavior This parameter determines the guards' behavior when searching for the player. We test three main behaviors:

Heuristic This is a method defined in (Al Enezi and Verbrugge 2021). We chose this method, believing it to be a relatively natural multi-agent search compared to the other methods we considered in this paper. Figure 2 shows a screenshot of how this method works. Once the player is out of sight, the method associates a numerical value to the closest line segment of a skeletal road map to the player's last position. This value corresponds to the player's likelihood of being near the associated line segment.

This method uses the straight skeleton (Giesen et al. 2009) of the map as a road map to propagate the probability of finding a target near a specific segment of the skeleton. Afterward, guards collaborate by choosing a separate spot in the walkable area to investigate. More details can be found in (Al Enezi and Verbrugge 2021).

Cheating In this behavior, all guards simply use the shortest path to the player's current position. This method is considered cheating because knowing the player's location is not justified to players. This behavior is relatively common in commercial video games (Švelch 2020).

Random Here, each guard individually requests a random position in the walkable area and takes the shortest path to-

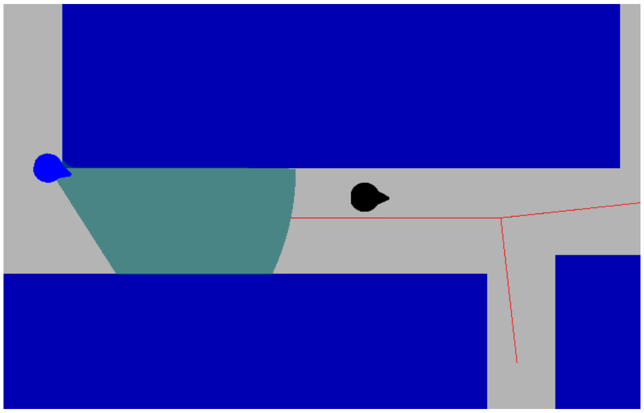


Figure 2: The skeletal road map propagation method. The player character is represented as a black circle. They escaped a guard, represented as a dark blue circle, and their FOV is in a lighter blue partial disk. The red lines represent the roadmap (straight skeleton (Giesen et al. 2009)) with non-zero probability of the player being near them.

ward it using the NavMesh.

Dialog Lines In order to enhance the immersion of players in commercial video games, it is a commonly used technique for NPCs to vocalize dialog lines during various situations. These lines serve multiple purposes, such as capturing the player's attention during important events, making NPCs appear more human-like, and aiding in storytelling. The focus of this paper is specifically on the dialog lines that NPCs exclaim to announce their intentions, thereby revealing the decisions and reasoning behind their actions. In this case, dialog lines prove valuable as they shed light on the decision-making process of guards, which may otherwise remain unclear to players. By employing dialog lines, this aspect can be simplified since players often do not devote sufficient attention to individual guards to discern the reasons behind their choices.

In our game, the guards actively pursue or search for the player and vocally communicate using a selection of appropriate dialog lines. We drew inspiration from well-known games such as "The Last of Us" (Gregory 2014) and "Left4Dead" (Ruskin 2012) to develop this feature. The lines spoken by the guards serve different purposes, such as alerting others when they spot the player, acknowledging when they lose sight of the player, or expressing their intention to search a specific location.

In our research, we established two primary variations that were distinguished by a simple factor: whether they included mentioning room names or not. Varying the level of information conveyed through the dialog can influence the perceived intelligence of NPCs and potentially impact a player's performance and enjoyment. The dialog variations we examined were as follows:

Abstract This collection of dialog lines provides players with no explicit contextual information. In this case, guards announce observations and possible future actions. Each spoken line is accompanied by a specific set of conditions that must be met before it is announced. These conditions may include the current state of the guards, the most recent timestamp of the player's known location, and the planned distance of the speaker's path. The dialog lines can be categorized into the following subcategories:

Preconditions **time_since_player_seen (t)**	**Lines**
$t > 40$ seconds	I finally see them! They are still here! Everyone come here!
$0 \leq t \leq 40$ seconds	Over there! Here they are! Through here! Enemy sighted! Here! On me! I see them

Table 1: Abstract lines a guard can use on spotting a player.

Preconditions **speaker_path_distance (p)** **search_elapsed_time (t)**	**Lines**
$p > 30\%$ **of the map diagonal and** $t < 2.5$ **seconds**	**I'll go from the other end!** **I'll ambush them!** **I'll take the longer path**
$p < 20\%$ **of the map diagonal**	**I will search around!** **They must be nearby!** **I need to check the nearby** **I'll to check this corridor!** **I'll check this hall!**

Table 2: Abstract lines a guard can use to announce their intentions.

Spotting The Player In this situation, when a guard visually detects the player, they select a dialog line corresponding to their search duration. Table 1 illustrates various examples of these lines. To prevent repetitive dialog in the same scenario, we incorporated multiple lines that guards can rotate between. When searching for an appropriate line, we prioritize the list in descending order based on the number of preconditions.

Announcing Intentions Guards can announce their next move; however, for the abstract dialog type, lines are generic and do not refer to specific rooms. Table 2 shows examples of these lines.

Filler Lines As guards search for the player, they announce lines that may add more human-like features. They do not announce clear intentions but observations or opinions that the player might already know. Additionally, guards can reply to each other by replying with a line from a set of lines grouped as a reply for the announced line. Table 3 shows an example of a group of lines and the group of

Preconditions search_elapsed_time (t)	Lines
$t > 40$ **seconds**	**I still can't find them** **They're good!** **I'm tired of searching!** **It's like they vanished!**
Replies	**Yeah! but we have to find them!** **You can say that again!** **Yeah! we need to do better!** **I'm sure they're still here!**

Table 3: An example of two sets of filler abstract lines. The first row is a line initiated by a guard, and the second is a set of replies a random guard can respond to.

Type of lines	Lines
Spotting the player	**They are in {IR}!** **On me in {IR}!** **Through {IR}!** **In {IR}!** **Come to {IR}!** **Alert in {IR}!**
Announcing intentions	**I need to clear out {GR}!** **They might be hiding in {GR}!** **I will keep looking in {GR}!** **I'll keep looking in {GR}!**
Filler lines	**I'm checking {GR}! You?** **{GR}! And you?** **In {GR}... you?**
Filler lines - Response	**Ok! I'm going to check {GR}!** **I'm going {GR}!** **Good! I'm going to {GR}!**

Table 4: Examples of lines for the different subgroups of contextual dialogs. {IR} is the name of the room the player was last seen in. {GR} is the destination room of the speaker guard.

replies.

Contextual The second set of dialogs provides players with additional contextual information, as these lines are linked to specific locations within the game level. Each room in the game level is assigned a name, such as "Cafeteria" or "Storage", for easier identification. All the rooms can be observed in the game screenshot depicted in Figure 1. This dialog group consists of the same subset of lines as the abstract dialog group. Table 4 presents a selection of contextual lines.

Participant Recruitment To recruit players for our study, we created an online portal that hosted a web-based game version. Participants were mainly drawn from our graduate and undergraduate students in different departments responding to a mass email. Player data was anonymously gathered for their game performance and survey answers.

Experiment Setup

The First-person view might reduce the ability of the player to observe guard motion. Thus, we chose a top-down perspective to provide players with full info on guard locations.

In each session, participants started by having the option

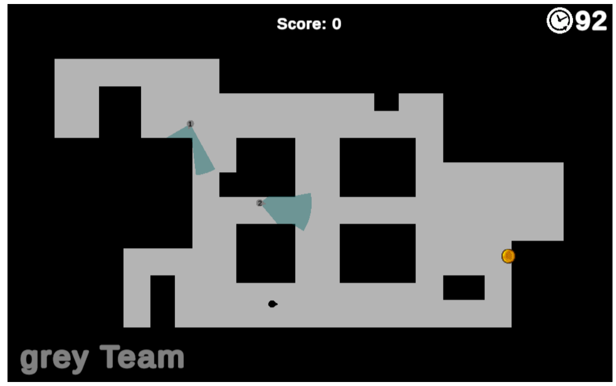


Figure 3: A screenshot of the tutorial map the participants played.

to play a tutorial level of two guards and a simple map taken from the game Metal Gear Solid. The goal of this level is to familiarize players with the game mechanics. In addition, guards had no dialog lines in this level for simplicity. Figure 3 shows a view of the map. Players could replay the tutorial level as often as they wanted before starting the actual study session.

After the players passed the tutorial level, they were randomly assigned to one of two groups in our study. The first aimed to compare guards with random goals and guards with the heuristic method (Al Enezi and Verbrugge 2021). The second group compared the cheating guards to the heuristic method. In addition, both groups compared the dialog variations we defined.

Eventually, each participant played four episodes, a combination of two NPC behaviors and two dialog variations. The order in which these episodes were played is randomized to get more representative results. We identified each team of guards to players by assigning them a unique color. These colors were randomly assigned to the teams to prevent any biases related to colors and behaviors.

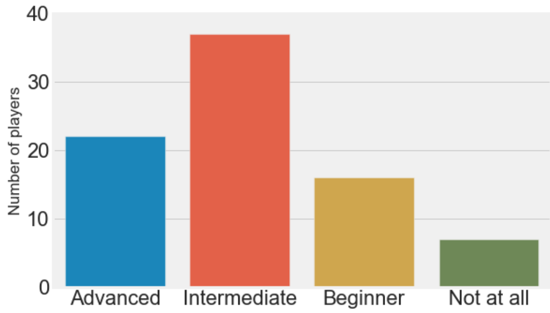
Survey

After the player played all four episodes, we asked them to choose the team they enjoyed playing against the most and the team they found to be the most difficult.

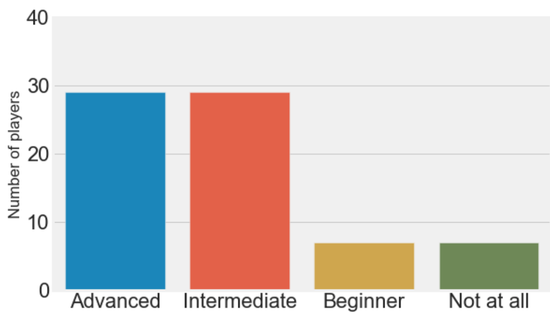
Results

Our study had 154 participants who completed the study. Each user played four games comprising four combinations of two dialog variations and two guard behaviors. One is the heuristic method, and the second is chosen randomly between cheating and random behaviors. This divided the participants into two separate groups. The first group compared the heuristic and cheating guard behaviors, and we collected the gameplay and survey responses of 72 participants. The second group compared heuristic and random guard behavior, and 82 participants were allocated to this group.

Participation Before commencing the study, we requested participants to assess their level of familiarity with

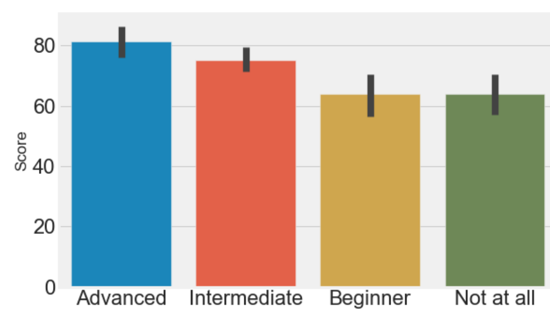


(a) “Random” behavior group

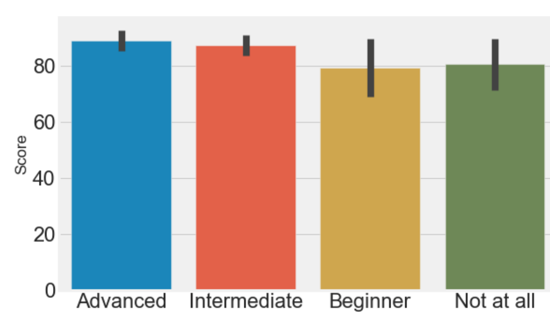


(b) “Cheating” behavior group

Figure 4: The distribution of how players rated their experience in video games for both experiments. The number of participants in the “Random” group, which consisted of the participants comparing the heuristic and random guard behavior, is 82, and the number for the other group is 72.



(a) “Random” behavior group



(b) “Cheating” behavior group

Figure 5: The scores participants achieved in the study grouped by their experience level for the two study groups. The error bars represent 95% confidence intervals.

video games. The distribution of participant ratings as video game players are displayed as a bar chart in Figure 4. Notably, we discovered that approximately 70-80% of the participants considered themselves to possess intermediate to advanced skills in playing video games. Herefore, their outcomes potentially reflect those of the broader video game player population.

Performance Despite the straightforward design of our game, we were curious whether participants’ familiarity with video games would impact their performance, potentially giving advanced players an advantage over other groups. Figure 5 shows a bar chart of players’ scores in both groups. We found no significant differences in scores observed among the intermediate to advanced experience levels. Furthermore, we also found that the scores demonstrated relative consistency as participants progressed through the rounds. Thus, we can conclude that the game possessed low complexity, enabling participants to grasp the game mechanics during the initial rounds. We next explore how the team of guards impacted the players’ scores.

Dialog & Guard Behavior On Player Scores To analyze the impact of guard behavior and dialog type on player performance, we compared scores based on these two variables. Figure 6 shows a bar chart that illustrates the distribution of players’ scores according to dialog type and guard behavior for the group that compared the “Random” and heuristic behaviors. Surprisingly, we observed that players scored higher when playing against the heuristic method. This outcome suggests that despite the randomly moving guards lacking knowledge of the intruder’s whereabouts, their unpredictable nature made them more challenging to anticipate in the game level. On the other hand, heuristic guards exhibited relatively more predictability, resulting in a greater opportunity for players to score against them. Furthermore, participants in the “Random” group appeared to benefit more from contextual dialogs compared to abstract dialogs. Given the unpredictable behavior of the guards in this group, contextual dialogs may have facilitated participants’ understanding of the guards’ plans, thereby enhancing their performance against the guards.

In the case of the “Cheating” group, participants achieved comparable scores when facing both the “Cheating” and heuristic behaviors and dialog types. It is intriguing to note that players scored similarly against the cheating guards, who had constant knowledge of the player’s location, and the heuristic guards, who had no such information. However, upon closer examination, we discovered that participants were able to exploit the cyclical nature of the game level to their advantage. They reported maneuvering around

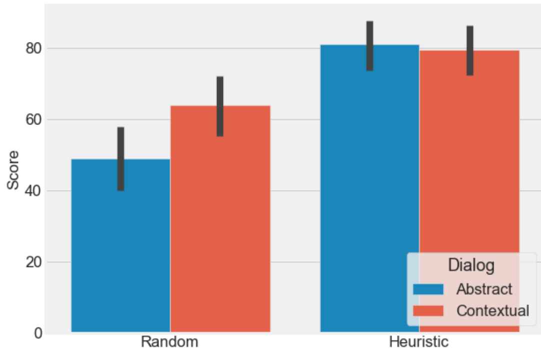


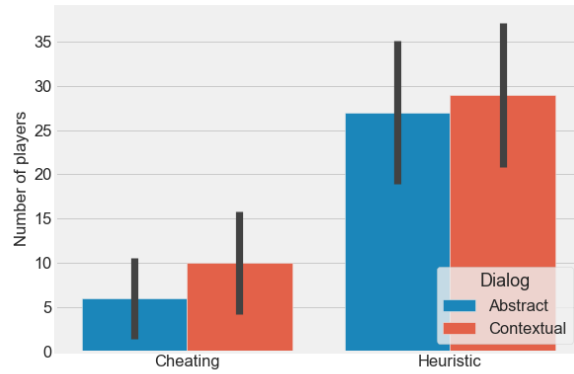
Figure 6: Players in the "Random" behavior group had higher score when they played against the heuristic method regardless of the dialog type the guards had. The error bars represent a 95% confidence interval.

the guards in a circular pattern while collecting the coins. This strategy allowed them to effectively navigate the game without being hindered significantly by the cheating guards' knowledge of their location.

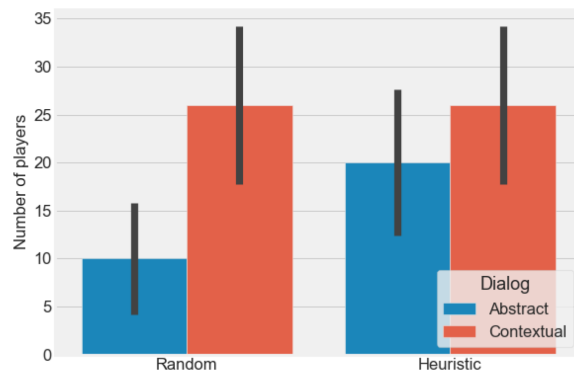
Survey Responses As participants finished the study, we asked them to rate the most enjoyable and difficult team they faced. First, we analyze players' responses regarding their most enjoyable guard behavior and dialog type.

Enjoyment Figure 7 shows the ratings for both participant groups. In the "Cheating" group, most players enjoyed playing against the heuristic method rather than the cheating guards. We suspect this result is due to the manner guards simply converged toward the players, and as a response, players simply used the cycles in the game level to their advantage to avoid the guards. On the other hand, the heuristic guards show more variation in their search, which forces the players to follow different strategies. This is further confirmed by a Chi-Square goodness-of-fit test showing that the guard's behavior had a significant influence on players' enjoyment, $\chi^2(1, 72) = [22.22], p < 0.001$. Table 5 further shows the Chi-Square goodness-of-fit for the guard behaviors and dialogs for players' choices of their most enjoyable teams. Regarding the dialog type impact, there is no significant impact in this case on players' enjoyment when we compare players' ratings of their most enjoyable teams. We suspect this outcome is the result of players easily noticing the difference between the behaviors which impacted their experience the most.

As for the players in the "Random" group, dialog type evidently impacted players' enjoyment, as shown in Table 5. When we compare both guard behaviors for the abstract dialog type, players clearly prefer the heuristic method. There are two possible causes for this observation; one, the heuristic method allowed players to score better, so they enjoyed it more. Two, the heuristic method showed a more exciting behavior than the random. However, changing the dialog type to contextual greatly improved players' enjoyment to match



(a) "Cheating" behavior group



(b) "Random" behavior group

Figure 7: Players in the "Cheating" group voted for the heuristic search behavior as most enjoyable. However, dialog type had a more significant impact on "Random" group choices. Contextual dialogs drastically improved players' enjoyment of the random search behavior. The error bars represent 95% confidence interval.

that of the heuristic rating. This outcome can have several interpretations, two of which are: one, when random guard behavior is paired with the contextual dialog, players scored more and thus enjoyed that team more. The second possible interpretation is that having the guards announce contextual dialogs helped in making them seem more attractive to players.

Difficulty Figure 8 shows a bar chart of players' votes for the most challenging team they faced in the study. In the "Cheating" group, although players' scores had no significant difference between the different teams, more players considered cheating guards with contextual dialogs to be the most difficult compared to abstract dialog. This outcome could be explained when guards explicitly announced the name of the room of the player's current location, and it overtly showed that guards knew the players' location and confirmed a sense of unfair advantage the guards had, so

	vs Random		vs Cheating	
	χ^2	p	χ^2	p
Behavior	1.21	0.26	22.22	<0.001
Dialog	1.21	0.01	0.5	0.47

Table 5: The Chi-square goodness-of-fit test results of the players' most enjoyable behaviors and dialogs. The sample size for comparing the cheating vs. heuristic method is 72, and for the Random vs. heuristic is 82 participants.

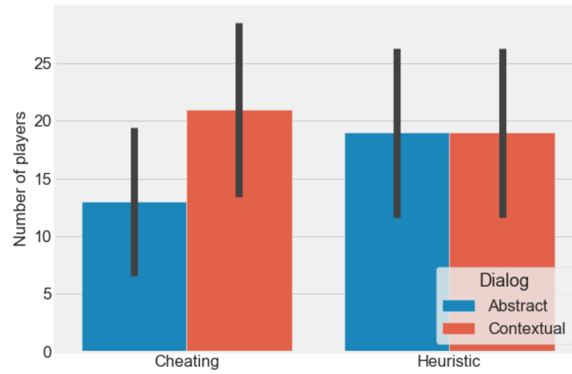
	vs Random		vs Cheating	
	χ^2	p	χ^2	p
Behavior	2.39	0.12	0.22	0.63
Dialog	1.21	0.2	0.88	0.34

Table 6: The Chi-square goodness-of-fit test results of the players' most challenging behaviors and dialogs. The sample size for comparing the cheating vs. heuristic method is 72, and for the Random vs. heuristic is 82 participants.

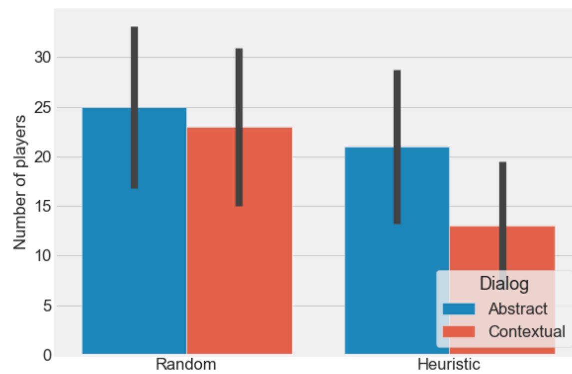
players considered it more difficult. Furthermore, players were equally distributed in choosing the most challenging team in the heuristic method, so dialog types did not impact the perceived difficulty.

When we compare the latter observation with how players voted for the most challenging team in the "Random" group, we found that the dialog type impacted the perceived difficulty of the heuristic team. We believe that this difference in players' opinions in the two groups is caused by the fact that it was easier for players to differentiate cheating and heuristic behaviors than random and heuristic behaviors. So in the "Random" group, the abstract dialog made the heuristic method appear more challenging to players. This could be explained by the fact that guards were more transparent in their decisions when using contextual dialogs.

To confirm the significance of this result, Table 6 shows the Chi-Square goodness-of-fit for the guard behaviors and dialogs for players' choices of their most enjoyable teams. The results show no statistical significance; however, we believe that further testing with a larger sample size may give more insight.



(a) "Cheating" behavior group



(b) "Random" behavior group

Figure 8: Players did not have any search behavior or dialogs that significantly affected the difficulty of the team. However, a slight trend was observed where contextual dialogs made cheating guards appear more difficult for a higher number of players while it made heuristic behavior to appear easier. The error bars in the data indicate a 95% confidence interval.

Conclusion & Future Work

Developing complex game AI can be difficult and represents an ongoing research challenge. Intelligent decision-making, however, is more of a mechanism in games than a goal and can be replaced by more straightforward, cheaper approaches when the difference is not evident to players. Our stealth game investigation and user study showed that players are able to notice the difference in NPC behaviors when the behaviors are sufficiently distinctive, at least when NPC behavior is clearly unjustified by the local knowledge an NPC should logically possess. This remains true even if the NPCs give the illusion of intelligence through appropriate and contextual dialog. On the other hand, other simple approaches, such as randomization, can combine effectively with detailed contextual dialogs to give an overall positive experience equivalent to (or even better than) one given by using more intelligent opponents.

In future work, we would like to further explore techniques to effectively but invisibly minimize intelligence requirements. Modeling player attention in relation to the relative impact of NPC behavior may further reveal opportunities for reduced development effort while maintaining an enjoyable and apparently realistic impression of NPC actions. Many techniques are employed in practical games, and continued academic examination and analysis would help validate and solidify when and how such approaches are most effective.

Acknowledgments

This work supported by the COHESA project, through NSERC Strategic Networks grant NETGP485577-15.

References

Al Enezi, W.; and Verbrugge, C. 2021. Skeleton-based multi-agent opponent search. In *2021 IEEE Conference on Games (CoG)*, 1–8. IEEE.

Brusk, J.; Lager, T.; Hjalmarsson, A.; and Wik, P. 2007. DEAL: Dialogue management in SCXML for believable game characters. In *Proceedings of the 2007 conference on Future Play*, 137–144.

Giesen, J.; Miklos, B.; Pauly, M.; and Wormser, C. 2009. The scale axis transform. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, 106–115.

Gregory, J. 2014. A Context-Aware Character Dialog System - Game Developer's Conference. https://www.gdcvault.com/play/1020386/A-Context-Aware-Character-Dialog. Accessed: 2023-08-14.

Hingston, P. 2009. A Turing test for computer game bots. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(3): 169–186.

Hladky, S.; and Bulitko, V. 2008. An evaluation of models for predicting opponent positions in first-person shooter video games. In *2008 IEEE Symposium On Computational Intelligence and Games*, 39–46. IEEE.

Iida, H.; Takeshita, N.; and Yoshimura, J. 2003. A metric for entertainment of boardgames: its implication for evolution of chess variants. In *Entertainment Computing*, 65–72. Springer.

Johansson, M.; Eladhari, M. P.; and Verhagen, H. 2012. Complexity at the cost of control in game design. In *Proceedings of the 5th Annual International Conference on Computer Games and Allied Technology (CGAT 2012). Global Science & Technology Forum*, 22–29. Citeseer.

Kacmarcik, G. 2006. Using natural language to manage NPC dialog. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 2, 115–117.

Lankoski, P.; and Björk, S. 2007. Gameplay Design Patterns for Believable Non-Player Characters. In *DiGRA Conference*, 416–423.

Livingstone, D. 2006. Turing's test and believable AI in games. *Computers in Entertainment (CIE)*, 4(1): 6–es.

Redding, P. 2009. Aarf! Arf Arf Arf: Talking to the Player with Barks. https://www.gdcvault.com/play/1308/Aarf-Arf-Arf-Arf-Talking. Accessed: 2023-08-14.

Ruskin, E. 2012. AI-driven Dynamic Dialog through Fuzzy Pattern Matching. Empower Your Writers! - Game Developer's Conference. https://www.gdcvault.com/play/1015317/AI-driven-Dynamic-Dialog-through. Accessed: 2023-08-14.

Soni, B.; and Hingston, P. 2008. Bots trained to play like a human are more fun. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 363–369.

Sutiono, A. P.; Purwarianti, A.; and Iida, H. 2014. A mathematical model of game refinement. In *International Conference on Intelligent Technologies for Interactive Entertainment*, 148–151. Springer.

Sutiono, A. P.; Ramadan, R.; Jarukasetporn, P.; Takeuchi, J.; Purwarianti, A.; and Iida, H. 2015. A Mathematical Model of Game Refinement and Its Applications to Sports Games. *EAI Endorsed Transactions on Creative Technologies*, 2(5).

Švelch, J. 2020. Should the monster play fair?: Reception of artificial intelligence in alien: isolation. *Game Studies*, 20(2): 243–260.

Togelius, J. 2014. How to run a successful game-based AI competition. *IEEE Transactions on Computational Intelligence and AI in Games*, 8(1): 95–100.

Turing, A. M. 2009. Computing machinery and intelligence. In *Parsing the turing test*, 23–65. Springer.

Warpefelt, H.; and Strååt, B. 2013. Breaking immersion by creating social unbelievabilty. In *Proceedings of AISB 2013 Convention. Social Coordination: Principles, Artefacts and Theories (SOCIAL. PATH)*, 92–100.

Wetzel, B.; and Anderson, K. 2017. What You See Is Not What You Get. In *Game AI Pro 3: Collected Wisdom of Game AI Professionals*, 31–47. CRC Press.

Yannakakis, G. N.; and Hallam, J. 2007a. Capturing player enjoyment in computer games. In *Advanced Intelligent Paradigms in Computer Games*, 175–201. Springer.

Yannakakis, G. N.; and Hallam, J. 2007b. Towards optimizing entertainment in computer games. *Applied Artificial Intelligence*, 21(10): 933–971.