Info Vis

Article

Information Visualization
2023, Vol. 22(4) 307–322

DOI: 10.1177/14738716231180892
journals.sagepub.com/home/ivi
Sage

Impact of incidental visualizations on primary tasks

João Moreira[1], Daniel Mendes[2] and Daniel Gonçalves[1]

Abstract
Incidental visualizations are meant to be seen at-a-glance, on-the-go, and during short exposure times. They will always appear side-by-side with an ongoing primary task while providing ancillary information relevant to those tasks. They differ from glanceable visualizations because looking at them is never their major focus, and they differ from ambient visualizations because they are not embedded in the environment, but appear when needed. However, unlike glanceable and ambient visualizations that have been studied in the past, incidental visualizations have yet to be explored in-depth. In particular, it is still not clear what is their impact on the users' performance of primary tasks. Therefore, we conducted an empirical online between-subjects user study where participants had to play a maze game as their primary task. Their goal was to complete several mazes as quickly as possible to maximize their score. This game was chosen to be a cognitively demanding task, bound to be significantly affected if incidental visualizations have a meaningful impact. At the same time, they had to answer a question that appeared while playing, regarding the path followed so far. Then, for half the participants, an incidental visualization was shown for a short period while playing, containing information useful for answering the question. We analyzed various metrics to understand how the maze performance was impacted by the incidental visualization. Additionally, we aimed to understand if working memory would influence how the maze was played and how visualizations were perceived. We concluded that incidental visualizations of the type used in this study do not disrupt people while they played the maze as their primary task. Furthermore, our results strongly suggested that the information conveyed by the visualization improved their performance in answering the question. Finally, working memory had no impact on the participants' results.

Keywords
Incidental visualizations, glanceable visualization, user studies, information visualization

Introduction

In 1996, Weiser and Brown[1] defined Calm Technology as any device that interacts with people via auditory or visual channels, while allowing information to be conveyed from the periphery to the center of human attention and back. For information visualization, this brought new opportunities regarding how information was received. New ways to convey data came to life such as glanceable,[2] ambient,[3] and incidental visualizations,[4] all inheriting from Peripheral Displays, no longer requiring users to stand still in front of a monitor.

Glanceable visualizations focus on conveying information at-a-glance so that people do not need to lose too much time understanding data. For example, a graphic shown in a smartwatch[5] that allows people to see data without needing to look at a visualization for too much time, that is accessible at any moment.

[1]Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[2]Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

Corresponding author:
João Moreira, Instituto Superior Técnico, Universidade de Lisboa. Rua Alves Redol, 9, Room 532, Lisboa 1000-029, Portugal.
Email: joaopedro.moreira@inesc-id.pt

Ambient visualizations focus on having information embedded in the environment, always available for people to see. For example, for real-time residential energy feedback.[6] Finally, the main purpose of incidental visualizations is to provide relevant information in pertinent moments without disrupting an ongoing task.

Let us imagine a company where each person operates a machine that produces t-shirts, requiring operators to be constantly focused on their tasks. During each shift, besides the usual work, operators need to keep track of three things: when a stock of raw material goes below a certain threshold; when one t-shirt printer reaches a certain temperature; and when the quantity of each t-shirt produced is unbalanced. The company determined that it needed a way to easily inform operators of all this information. For the first, they decided to use a line chart depicting how raw materials are being spent for the past few days. For the second, they decided to use a bar chart depicting the temperature of each printer. Finally, for the third, they decided to use a donut chart depicting the proportion of each t-shirt from all produced. The next step was to decide where to put this information. Initially, they thought of putting three ambient visualizations next to the machine. However, the company would rather not have the information always available because the information is not always needed, and having three visualizations continuously being displayed would just create unnecessary visual pollution. Another solution could be having several glanceable visualizations in a wearable, allowing each operator to see the information at-a-glance every time they needed it. Nevertheless, this would still require them to mechanically look at the information. Since each operator is always busy managing each machine, this solution would probably disturb their ongoing tasks. The company concluded that it needed incidental visualizations. On one hand, information would be displayed to the operators only when it was needed. On the other, they would appear to them at a glance in their visual field, avoiding disturbing their current task.

Although there has been recent work covering both glanceable[7–9] and ambient[10] visualizations, there is still unexplored theory regarding incidental visualizations. In particular, until now, it has only been shown that it is possible to perceive information displayed for very short periods at specific moments.[4] However, regarding the ongoing tasks that are being conducted by people, it is not known yet how much seeing an incidental visualization disrupts their performance.

In this paper, we present the results of an empirical online between-subjects user study with 80 participants to study the extent to which an incidental visualization disrupts the participants in performing an ongoing primary task. Forty of them perceived an incidental visualization while performing a primary task, and the other 40 participants performed the primary task without seeing the visualization. The primary task was playing a browser-based maze game, where the goal was to move a character across the maze as quickly as possible using the keyboard. At the same time, participants had to keep track of their movements during gameplay. By analyzing the data, we show that the performance of the primary task was mostly not affected due to incidental visualizations, and we show that having information conveyed this way allowed for additional insights regarding the primary task. Our major contributions with this paper are: user study to evaluate incidental visualizations during primary tasks; how people and their ongoing primary tasks are influenced by incidental visualizations; key insights for future research with incidental visualizations.

Related work

In this section, we will present three main topics: peripheral displays, Visual Memory, and incidental visualizations. Regarding the former, we will explain their origins, goals, strengths, and how Information Visualization has been used. In the latter, we will present two topics, Operation Span Tasks and Subitizing. The first because we wanted to consider Working Memory in our study,[11] and the second because we used subitizing theory for several design decisions in our user study. Finally, regarding incidental visualizations, it has been explored recently how people perform on pre-attentive graphical perception tasks.[4] In particular, the authors wanted to validate if certain known results like the ranking proposed by Cleveland and McGill[12] would be the same for incidental visualizations.

Peripheral displays

One well-known definition of peripheral displays was given by Matthews et al., "A display is peripheral when a particular person uses it in an operational or automatic way."[13] It is a display that people can perceive by moving their sight to the periphery and back again with little to no disrupting effects. This inherits from the definition of Calm Computing,[1] where Weiser and Brown first introduced it as technology that exists as part of its environment.

In their work in 2005,[13] Matthews made an in-depth analysis of peripheral displays, exploring in which contexts they were used, how they could be evaluated,[14] and which guidelines should be followed to create them.[2,14] Peripheral displays always have a scope of use, a set of supported activities, and a criticality level of use. The scope of use defines its utility for the activities it will support. Then, criticality manages the importance of the information being sent.

Glanceable Displays: Eventually, peripheral displays branched into a new topic called glanceable displays. Glanceability "refers to how quickly and easily the visual design conveys information after the user is paying attention to the display."[2] As with its parent field, these displays are designed to allow information to be perceived as quickly as possible. However, it differs by the amount and type of information conveyed. A peripheral display assumes it is available at the periphery, but it is not specified how much time people can spend looking at it. Regarding a glanceable display, information must be simple enough that allows people to understand it without taking too much time.[5,7,15,16] Furthermore, these devices do not require people's attention as notification systems do.[17]

One particular branch of research that emerged from glanceable displays is glanceable AR, where the focus is to convey glanceable information using augmented reality head-worn devices.[9,18–20] Feiyu Lu et al., for example, has been studying how people could interact with glanceable widgets[20] in AR, and how could that information be part of people's daily routine.[9] Their results show promise, as people felt relatively positive about glanceable AR in general. However, there is still a big issue with this approach, which is occlusion and focus disruption, challenges that may not just exist in AR. Understanding the best place to put a glanceable display is an ongoing challenge,[21] since achieving calm technology assumes that people look at technology as part of the environment.[1]

When Information Visualization is used in glanceable displays, designers create glanceable visualizations, where now the information is conveyed via graphics. In the last few years, we have seen instantiations of this concept mainly on wearable devices such as the smartwatches,[5,7,15,16] mostly because smartwatches are usually very accessible and easy to look at regarding the effort that people need to put into it. Even so, smartphones are also a possibility.[22] Gouveia et al. showed that glanceable feedback has a positive effect.[15] Neshati et al. tested how could a Line Chart be displayed in a smartwatch, where the challenge was having such a small display to use,[5] and ended up defining several design guidelines for this graphic in particular. Similarly, Blascheck et al. were attempting to find effective visual idioms to display in a smartwatch.[16] They tested the Bar Chart, Donut Chart, and Radial Bar Chart, and found out that Radial Bar Char had the worst results. They later replicated their study to understand if changing the size of the visualizations would make a significant difference,[8] and they concluded that it would not.

Ambient Displays: Similar to glanceable displays, we can have Ambient displays,[3] also inheriting from peripheral displays. Although they may also be glanceable, their focus is to present information while being embedded in the environment. As a result, esthetics steps in as one of the key features in designing Ambient displays.[23] Then, new opportunities for displaying information originate. Since people can look at ambient displays for as long as they want, these displays can focus on giving real-time awareness of underlying data from a specific context of use.[24–26]

However, once again, as with the challenges of glanceable AR, cognitive load is one major issue.[10] As pointed out by Shelton et al., being frequently layered with digital information received from multiple devices that require our attention may place stress on cognitive load. They eventually proved that indeed there is a significant increase in cognitive load, so they argue that it is important to validate how much having an ambient display disrupts people. Finally, as with glanceable visualizations, there are also Ambient visualizations, which inherit from Ambient displays, but now information is specifically conveyed via graphics.[23,27] Furthermore, Ambient visualizations may be designed as a physical visualization,[28] and since they are embedded, sometimes data may be related to people's daily routine.[29]

Visual memory

In cognitive psychology, many of the most used tools for measuring Working Memory are Working Memory Span Tasks[30] such as the counting,[31] operation,[32] and reading[33] tasks. However, using these tools takes time, for both running participants and scoring their data. In particular, for Operation Span Tasks, Unsworth et al. developed a simplified version to be easily applied in any study.[34] In an operation span task, participants are required to solve a series of math operations. At the same time, they need to remember a set of unrelated words.[32]

In the Unsworth et al. version, participants have control using a mouse, and they are allowed to complete the tasks without the instructions of the experimenter. First, a math operation is presented to the participants. After solving the operation, each participant chooses to move to the next phase, where a value will be shown. Then, the participant must decide if the value presented is the correct result of the previous operation. After choosing, a letter appears for 800 ms, which participants must attempt to memorize. Depending on the set size of the experiment, participants may repeat these steps several times. In the end, the more steps it has, the more letters must be memorized. After solving all operations, participants must select from a set of letters, the ones that appeared. The final score was calculated in different ways: OSPAN score, total number correct, math errors, speed errors, and accuracy errors. In particular, the total number of correct answers shows if participants were able to recall in which order each letter appeared.

Subitizing: Subitizing is a well-known phenomenon in cognitive psychology that relates to people's working memory. In 1871, Jevons[35] experimentally addressed a philological question posed by Hamilton in 1859: "How many objects can the mind embrace at once?." Subitizing has been addressed since that time,[36,37] and, even now, there is still debate about how this phenomenon occurs.[38] In particular, visual subitizing, when compared with the auditory and tactile modalities, is by far the most studied. In summary, in the visual modality, the subitizing range is defined as four,[38] but presenting familiar patterns can increase the range. Therefore, accuracy to memorize visual items drops significantly from more than four items, except if patterns can see detected, like for example the six sides of one normal dice. However, Katzin et al. argue that subitizing is, in fact, an effective enumeration mechanism, and they propose a new definition of subitizing: "a subprocess of counting that takes place when facilitating factors are present, and yields accelerated and more accurate enumeration."

Incidental visualizations

Since it is a relatively new topic, there is little related work regarding incidental visualizations apart from the Graphical Perception study we mentioned.[4] We know they inherit from peripheral displays because the information is supposed to be available at the periphery. Their scope of use is always side-by-side with an ongoing primary task, which is any task that does not involve looking at the incidental visualization. Then, by default, it supports any activity, since there is no explicit context of use. Finally, regarding criticality, due to the information being shown for short exposure times, perceiving it cannot be crucial for the ongoing primary task.

Then, they inherit from glanceable displays because of their context of use. They convey glanceable information while people are performing primary tasks. However, there is one specific trait that can occur in glanceable Visualization that cannot occur in incidental visualizations by definition, which is interaction. For example, the glanceable AR approach we have seen has glanceable information being shown, but it also supports interaction with a person. In incidental visualizations, were it to be any interaction, perceiving it would no longer be a secondary task. Furthermore, issues like occlusion should never occur because that would compromise the ongoing primary task.

With Information Visualization, they differ from glanceable visualizations because the act of receiving information is not supposed to be the primary goal. Instead, performing the ongoing primary task is. Information is received as a secondary task that is performed simultaneously. Besides, incidental visualizations are not limited to wearable or portable devices, as in most seen cases. For example, we do not limit its design to small screens. Finally, when comparing Ambient visualizations, the major difference is in the information availability. While Ambient information is supposed to be always available in the environment, incidental information appears only during another task, and it is only available at specified moments. Therefore, incidental visualizations end up targeting situations where people cannot afford to stop doing their current primary task, cannot look away at a visualization for too long, and cannot have additional situated displays in the environment.

Discussion

In summary, we concluded that incidental visualizations have yet much to be explored. In particular, we cannot assume that the results obtained until now regarding peripheral, glanceable, and ambient displays can automatically be used. The context of use, supported tasks, and criticality of incidental visualizations are different from current research approaches, and the incidental scenario has not been approached yet. Peripheral displays may allow long explore times, glanceable displays are required on-demand, and ambient displays try to be embedded in the environment to avoid inducing more cognitive load. However, incidental visualizations appear for short exposure times, always during an ongoing primary task, and only at specific moments without people requiring them.

Regarding Visual Memory, cognitive psychology may play an important role. In particular, the Operation Span Task appeared to be a valuable tool to apply in our study, since an incidental Visualization is presented for short exposure times. It would be important to understand if somehow this individual characteristic affects how the information is retained while a person is performing a primary task. Furthermore, subitizing is also a phenomenon that should be taken into consideration, not only for incidental visualizations but also in any scenario where visualizations are presented such as glanceable and ambient visualizations. However, we found it odd that subitizing was

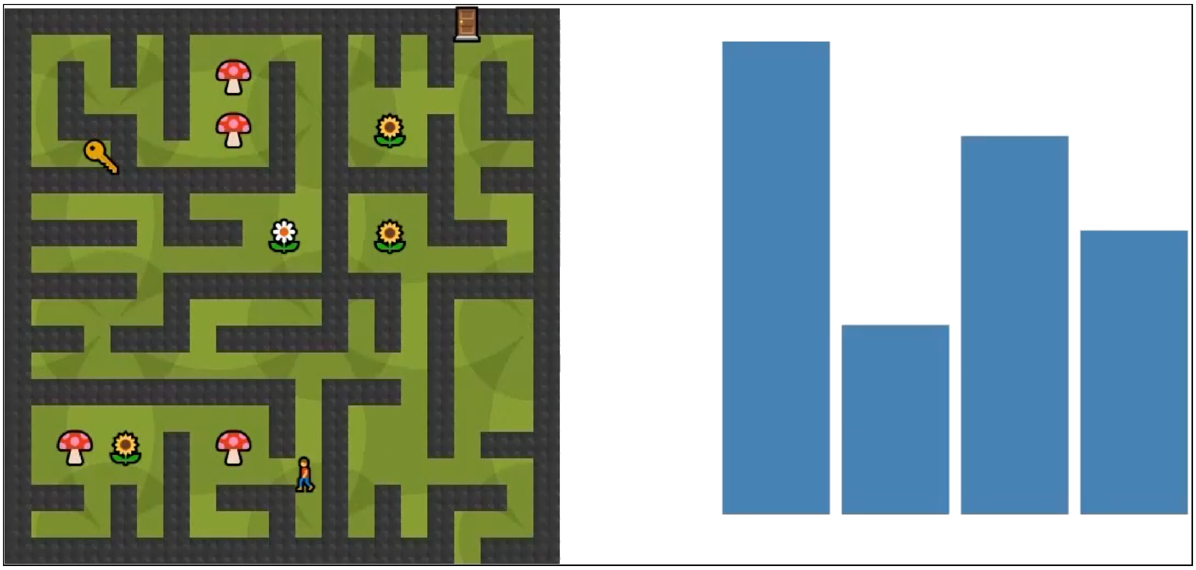

Figure 1. A user study was conducted to evaluate how disruptive are Incidental Visualizations on people's ongoing primary tasks. This image shows one state of the main phase of the user study, where the participant had to glance at the Incidental Visualization as it appeared. As in the image, the bar chart appeared next to the maze. The first bar encoded the number of times the "up" direction was pressed since the maze started, the second bar encodes the "down" direction, the third the "left" direction, and forth the "right." Participants then had to rank the directions pressed from more pressed to less pressed, using both their memory and information perceived in the bar chart.

never mentioned in the studies we reviewed, probably because having just four marks is not enough to display useful information for most cases. Still, we argue that this cognitive phenomenon is crucial to ensure accurate quick perceptions of incidental visualizations where exposure times are brief. Furthermore, the subitizing range has been explored in graphical perception tasks, and accuracy is overall high.[4] In the next section, we will explain how our user study was designed, and how we reached our goals.

User study

Our goal was to study the extent to which an incidental visualization disrupts the participants in performing an ongoing primary task. We aimed to answer these four research questions:

- RQ1: Does perceiving an incidental visualization disrupt the primary task performance?
- RQ2: Does perceiving an incidental visualization increase the amount of stress felt by people?
- RQ3: Are incidental visualizations able to provide actionable information while people are performing a primary task?
- RQ4: Does working memory influence how well people perform at the primary task and how well they perceive the incidental visualization?

To reach our goal, we conducted an online between-subjects design user study on the scientific crowdsourcing platform Prolific (https://www.prolific.co/) with 80 participants. Crowdsourcing platforms have been widely used for the past years[39,40] since it helps with one of the most challenging situations in research: large samples. Each participant was asked to play a maze game (primary task), designed to be cognitively demanding. Its goal was to finish different mazes with the highest score possible. At some point during the maze, the game asked a question that required participants to be aware of the game from the beginning (Figure 1). However, before the question was asked, for 40 participants, an incidental visualization, depicting information needed to answer it, was presented while the user was solving the maze (independent variable). This way both the primary task and the visualization's data were related. The other 40 participants were not presented with this additional information. Therefore, by looking at both groups, we were able to understand how exactly incidental visualizations impacted people while playing, and whether they could still effectively convey relevant information to the users. To ensure the visualization shown was incidental and not ambient, it only appeared for short exposure times, instead of being constantly displayed to each participant. Then, to ensure it was not a glanceable visualization, it appeared without the participants having to search for it.

In this section, we will explain how the task was implemented, how we managed mischievous or careless behavior from the participants (since it was an online study), and how the game was played. Then, we present the incidental visualizations we designed and the information they conveyed. Afterward, we briefly explain the interface participants had to deal with during the study. Finally, we present the several steps participants had to go through during the study from start to finish.

Tasks

The primary task of our study was to play several instances of a maze game in the web browser, based on the one created by Chirp Internet (https://www.the-art-of-web.com/javascript/maze-game/). People have to move a character through the maze until they reach a door. However, they can only finish the maze if they have previously picked up a key. The performance of the maze run is given by a score that can increase if the player catches treasures, which are represented by several emojis (presents, flowers, diamonds, and mushrooms), and when the key is picked up. Then, the score can decrease in two situations. First, for every step taken by the character, the score is decreased by one point. Then, if the character bumps into a monster (crocodiles or snakes), the score decreases by 10 points. If the score gets to zero, the players lose the game. The initial score was defined by adding two distances: from the exit to the key, and from the entrance to the key. This gives the minimum number of steps needed to finish the maze.

The authors generate each maze from scratch using the Recursive Division algorithm, which allows for each maze to be entirely different regarding walls, treasures, keys, and monster placement. Therefore, the players get a unique experience in every maze run. We removed the text that contained the instructions to finish the maze and left only the text containing the score. This way, we avoided having more than the essential information available.

Imposing Focus: In the scenario that we described previously, incidental visualizations are to be seen at a glance, and they only appear to people for short exposure times. At the same time, each participant was supposed to maintain constant focus on the primary task. To increase participants' pressure, if the player took more than one second to move another step, the player would lose the game. This way, we guaranteed that the primary task was performed with a high degree of attention. Looking away could be enough to make the player lose focus. This pressure timer was visually shown by interpolating the character's opacity. At first, the character is opaque, but after one second the character becomes transparent (Figure 2). This event is reset every time the player moves the character.

Figure 2. If the player stops moving the character for a few milliseconds, it starts to become invisible, until eventually the game is lost.

Figure 3. Example of a corridor that would lead to making repetitive movements, thus making it possible for them to look away while playing.

To identify the pressure timer used, we conducted a preliminary study where we asked 14 participants to play the game where the pressure timer was 1 s. Therefore, if a player took more than one second to move, the player would lose the game. Our goal was to check people's completion percentage for this value in particular. Each participant played six different mazes, all generated in a 19×19 grid, and the order in which they played each maze was generated via a Latin Square distribution. In summary, in two of the mazes, 10 participants finished the maze. In another two, 12 participants were also able to finish. At the other, 13, and in the last one, 14. Thus, for each maze, at least 70% of the participants were able to finish without succumbing to the pressure timer. Therefore, we concluded that one second on the pressure timer was a value that allows playing the game without making it too easy. Furthermore, we kept the 19×19 grid as the official size for the mazes in our study, since these results were obtained for this maze size.

Avoiding Repetitive Movements: To avoid repetitive movements that would allow the participant to repeat them while looking away from the maze, each keypress calculated the next cell that allowed more than the movement in the opposite direction. Therefore, if the character entered a corridor (in Figure 3, from the left), it would be automatically skipped. As a result, this made each keypress an important decision for the player because there was always more than one possible direction to choose.

Furthermore, we kept track of movements with a buffer of previous steps. If a pattern was detected at least two times in a row, we would give players a score penalty (explained later). Regarding wall collisions, they were easier to detect because it was a collision with an object in the maze.

Score: To calculate the score, instead of it changing only due to the players' actions, we implemented the game as a time-attack. As a result, the score represented how many seconds players had to finish the maze. Additionally, in the main phase, we presented the high score for the current maze each participant played.

Penalties also needed some reconsideration. First, we added penalties for movement patterns detected. In case players repeated a movement more than two consecutive times, they would receive a 2-s penalty. However, for each consecutive repetitive movement, the penalty increased exponentially, thus forcing players to stop. For example, the penalty starts at two, then four, then 16, and then 256 (an instant lost game at any point). Bumping into a wall for the first and second time yields a 2-s penalty, but this penalty grows exponentially after that. Hitting monsters results in a 10-point penalty.

We made all the treasures mandatory, and nothing could increase the score at any point. Therefore, the only thing that maximized the player's score was to finish the game as quickly as possible, including picking all the treasures (in any order), the key, and then going to the door. Meanwhile, the player had to avoid movement patterns, avoid going into walls or monsters, and avoid standing still for more than one second. All these changes were to avoid any unwanted unpredictable effects in our statistical analyses. If we were to consider treasure bonuses, we would need to consider each possible combination of treasure-picking orders. As a result, our analyses would become too complex and would be deviated from our goals. Our version makes the primary task more simple to understand, and more straightforward since players now do not need to calculate which treasures are worth picking up.

Incidental visualization

Our incidental visualization appeared for just one second, next to the maze on the right side. It conveyed information regarding the number of turns each player took in each direction.

Contextualized Information: Our incidental visualization conveyed information about the primary task since we wanted to understand how that would affect participants when asked about how they played the game. Therefore, we presented to the player information generated during the corresponding maze run.

Subitizing research has proven that, up to four items, people can count how many items there are.[38] Furthermore, it has been proven that pre-attentive graphical perception for incidental perception is effective in the subitizing range.[4] Therefore, we used the number of times each direction was chosen because there are only four possible directions for the character to move, which fits the subitizing range. We kept track of the keys pressed to generate the necessary data.

Visual Idiom: Each direction represents a category with a quantitative value (the number of times each was chosen). In Information Visualization, the comparison of items is one of the most well-known tasks.[41] Therefore, participants needed to compare the number of times each key was pressed. For that reason, we chose the Bar Chart as our visual idiom. Since the famous Cleveland and McGill graphical perception study,[12] it is known that length comparison on a common axis is a more precise task. Furthermore, Bar charts can be used for comparing categories that have quantitative values. With four possible directions, we encoded information with 4 bars. To make the visualization incidental, we made sure the bar chart appeared next to the maze, as seen in Figure 1, and to make it be shown for only one second. We chose this time interval because it has been concluded that one second was enough to allow people to perform graphical perception during short exposure times.[4] Finally, we designed the visualization to only present the four bars to maximize the ink-to-data ratio, and we normalized the bars so that the maximum value always had the highest height possible.

Since each player plays each maze differently, it would be impossible to choose a specific moment to present the incidental visualization. The combinations of keys pressed during each moment in the maze most likely differ between players. Therefore, we presented the visual idiom only when each direction key was pressed a specific number of times. As soon as the amounts could be ranked, with no direction key having equal values (e.g. up 1, down 2, left 3, and right 4), the visualization would appear.

Perception Accuracy: To access the perception accuracy of data shown in the incidental visualization, players had to answer a question 1 s after the visualization was shown. We asked players to rank each direction from most pressed to less pressed, via a pop-up window that paused the ongoing maze and score timer. They could answer by choosing one of each button, each representing a direction (Figure 4). First the most pressed, then the second most pressed, and finally the third most pressed. The fourth was assumed chosen after the third. Therefore, accuracy could either be

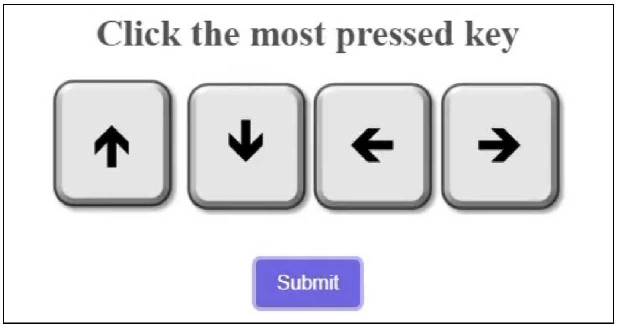

Figure 4. Players had to rank the directions from most to less pressed, by submitting each button, one at a time.

zero, one, two, or three. After all three buttons were submitted, the maze would proceed normally.

There were two bias effects that we attempted to prevent when creating our Bar Chart. First, the visualization height was always adjusted to the highest bar value. This way, we made sure that visualizations triggered at different moments by different players did not change significantly in shape. For example, the set of values up 1, down 2, left 3, and right 4, and the set of values up 8, down 9, left 10, and right 11, using our approach, have the same shape. Second, the order in which the bars are placed in the Bar Chart corresponds to the buttons shown to the players. This way, we facilitated the mapping between the data they saw and the button they needed to press (Figure 5).

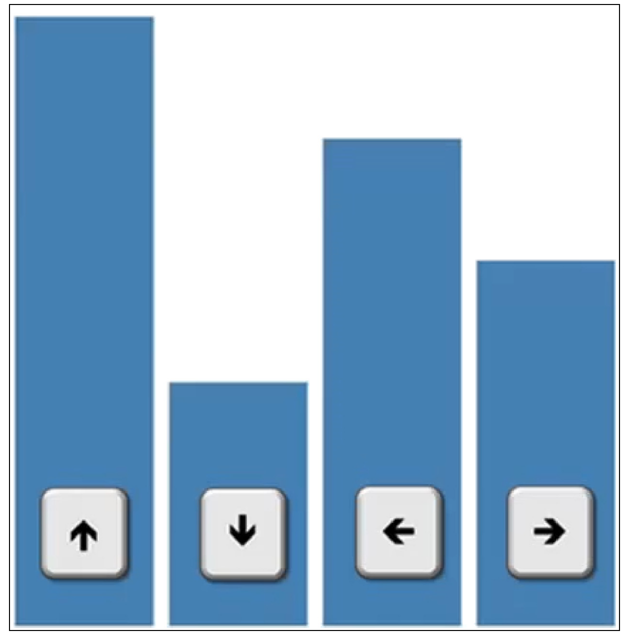

Figure 5. Illustration to explain how the order of the buttons corresponded to the order of the bars.

Study workflow

We divided our study into four phases: OSPAN Task,[34] instructions, warm-up, and main. In the first phase, each participant had to go through a simplified version of the operation span task, which allowed us to later correlate Working Memory with our results. In the second, participants watched a video that served as a global tutorial for the entire study (https://figshare.com/s/4172cf9797f9202b4afb), and how they should behave during the study to ensure that trustworthy data was collected. The third phase allowed participants to play three unique mazes as many times as they wanted, thus allowing them to get familiarized with the webpage interface and the game that was explained in the video. Furthermore, the players got to experience the real trials even though data was not collected. Only after they felt confident were they advised to move to the next part. In the main phase, people played six unique mazes (Figure 6), all different from the ones played during warm-up (Figure 7), where each maze corresponded to a different trial. Everything else was the same, except they could not replay any of the mazes. Each participant played the same mazes, but the order in which they played each one was generated using a Latin square distribution to avoid any bias.

Data: To reach our goal, we monitored specific metrics for the entire study. To answer RQ1, we needed to save each player's score for each of the six mazes, to then look at both conditions (with and without an incidental visualization) and compare them. The score was saved as a quantitative value. We also tracked how many times each participant, for each maze, bumped into walls and monsters, how many repetitive patterns were made, and the average time each key was pressed, all quantitative values. This last metric was calculated by tracking how fast the character was being moved. Next, to answer RQ2, we needed to monitor which mazes was the player unable to complete by taking more than one second to move. If the player was unable to complete a particular maze, the maze would be flagged a loss due to pressure (0 or 1, which is a dichotomous variable). It is worth mentioning that loss due to pressure is not the same as losing because of idle behavior. Therefore, the idle cases were discarded for this analysis. By comparing the percentage of people that did not complete each maze, in both groups, it would be possible to see if there is a significant impact induced by the visualization. Then, to answer RQ3, we look at how perfect the raking of each participant was (Figure 4), a value that was either zero, one, two, or three (quantitative value). Then, again, we compared both groups.

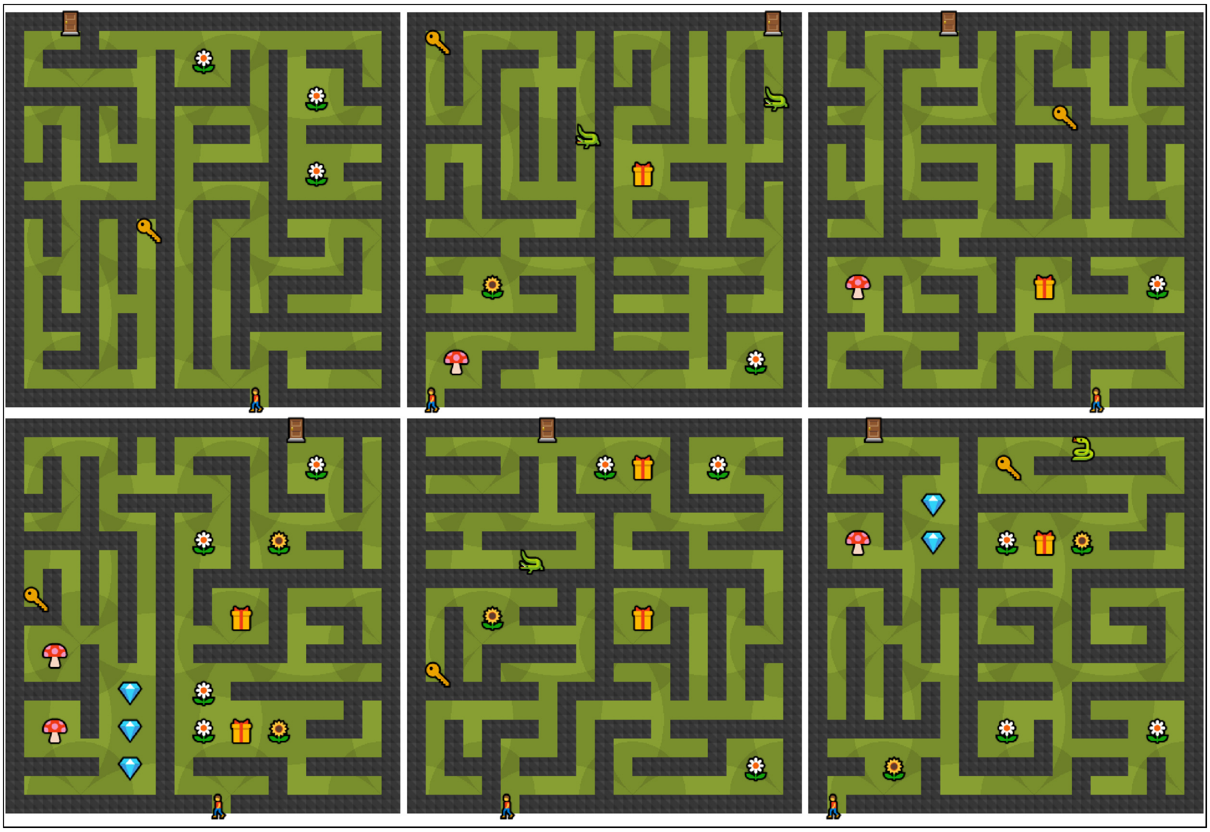

Figure 6. All the used mazes during the study. They were all randomly generated using the recursive division algorithm.

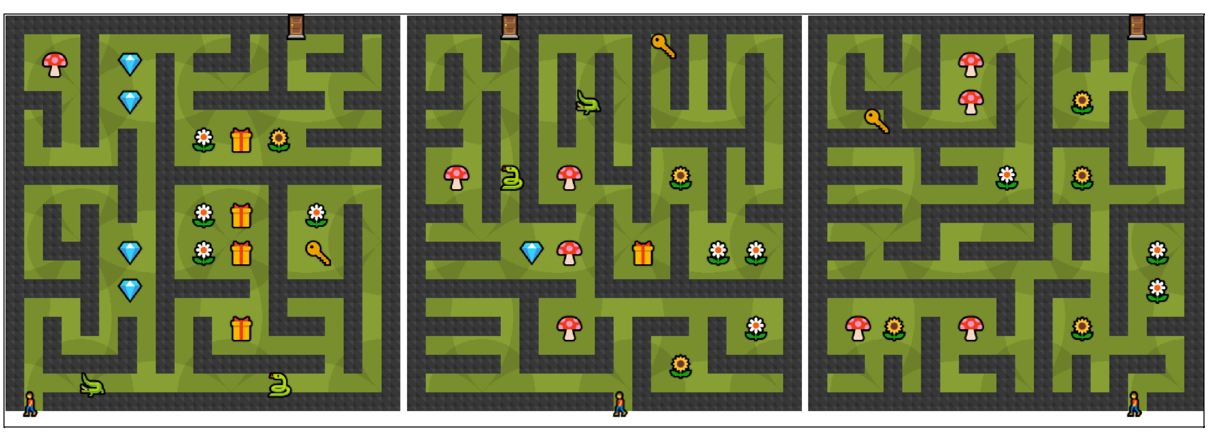

Figure 7. All the used warm-up mazes during the study. They were all randomly generated using the recursive division algorithm.

To answer RQ4, we analyzed three metrics. First, we used the OSPAN task, inspired by the one used by Castro et al.[42] Again, using as the baseline the subitizing range, our OSPAN task was implemented with a set of four images, and it went as follows. Each participant saw an image fading in and then fading out after 1 s. Afterward, a math problem needed to be solved, and again, it faded in and out the same way. Then, a value would appear, and participants had to decide if the value was either right or wrong. This workflow was repeated four times, and, at last, the four images appeared. Participants were then tasked to choose the images in the order of appearance. The score was calculated by checking how many images were correctly chosen, and weighted by how many correct answers were given. For example, if the order was correct

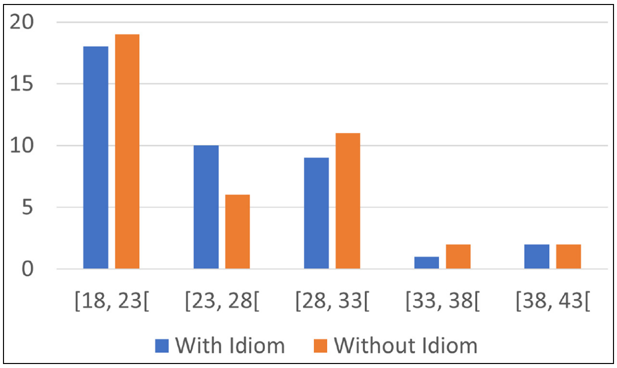

Figure 8. Age distribution of the final 80 participants, 40 of which played the maze with the visualization, and 40 without.

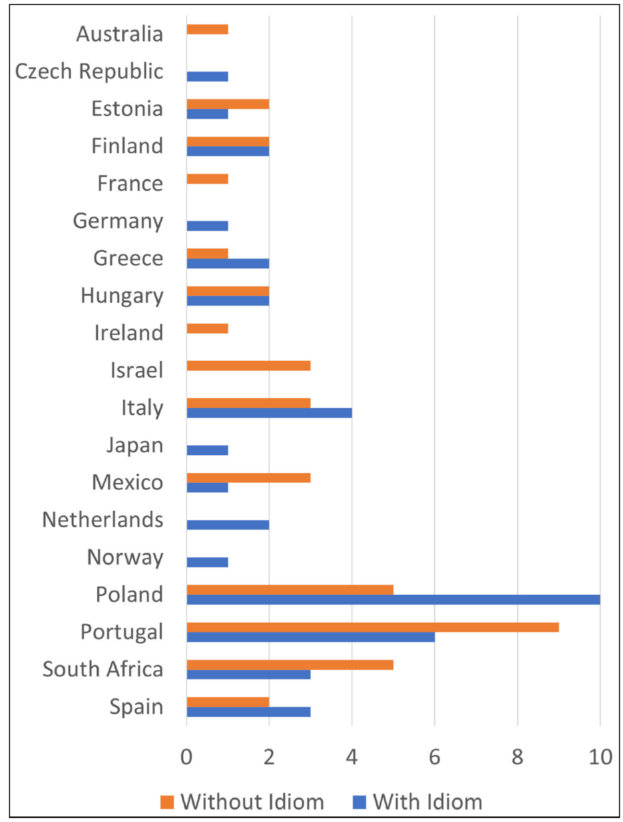

Figure 9. Country distribution of the final 80 participants, 40 of which played the maze with the visualization, and 40 without.

(100%) and if just one question was wrong (75%), the final result would be 75% (0.75). These values would later be used to see if visual memory was influencing how people performed at the questions asked when an incidental visualization was shown. The second additional metric was gathered at the end of each maze run during the main phase. In the cases where participants finished a maze, before starting the next one, they were asked to rate their performance using a Likert scale with values ranging from one (Very Poor) to five (very Good). This way, we could later check if the participants' self-perception of how they performed differed with and without an incidental visualization. Finally, at the end of the study (after all six mazes) we asked participants to rate how much were they attempting to beat each maze's high score using again a Likert scale with values ranging from one (Nothing) to five (A lot). Again, we would like to know if it somehow would significantly affect how they played each maze.

Participants: We used Prolific filtering to get people who fulfilled some minimum requirements to ensure we had no Ineligible Workers.[43] First, we did not impose restrictions on their current country of residence. Then, we required all participants to use a desktop computer, and we asked for a balanced sample (50% male and 50% female). Next, each participant had to be fluent in English since the study was conducted in this language. Regarding medical conditions, we asked for people that had normal or corrected to normal vision. Finally, we asked for people that had some experience playing video games, since the primary task was to play a game. In the end, our participants ranged between 18 and 43 years old, and we got participants from 19 different countries. In total, we gathered 90 participants. However, due to data problems that we will explain just next, 10 were rejected. Of the remaining 80, mostly between 18 and 33 years old (Figure 8), 40 played the game with an incidental visualization and the other 40 without. Furthermore, we got participants from across 19 countries (Figure 9). On average, each participant took 12 min to complete the study, all paid £7.50 per hour.

Interface

To prevent as much as we could any mischievous or accidental data-compromising effect, we looked at the work by Gadiraju et al.[43] There are Ineligible Workers (not filling prerequisites), Fast Deceivers (participate too fast), Rule Breakers (do not follow the rules), Smart Deceivers (exploit weaknesses), and Gold Standard Preys (not paying attention).

Since there is little to no control over participants in an online study, we implemented some validations to avoid any issues with the data collected. First, we monitored when people were idle. When using platforms like Prolific, participants get paid by the hour. Therefore, we had to track any Smart Deceiver. Second, we checked whether people lost focus from the browser tab where the study was being conducted. We did not want participants to be distracted after they started the study, which would make them the Gold Standard Preys.

For idle behavior, we implemented three strategies. First, we added timers to all pop-up boxes. Depending on the type of box it was, the study was either canceled or a default action was taken. Then, in some pop-ups like the initial ID submission, if participants took more than 15 s, they would be removed from the study and would be blocked from starting again. Finally, we implemented an idle behavior check at the beginning of each maze run that worked similarly to the pressure timer. However, this one took longer to trigger, and only happened before the first step in each maze. If a participant took too much time starting the maze, the study would consider it a loss due to idle behavior.

Regarding focus, we monitored when participants meddled with the browser tab where the study was being conducted. If they changed the zoom, left, or minimized the tab, at any moment during the study (except during the maze runs), participants had a limited time to return to the study tab again. Otherwise, the study would be canceled, and, again, the participant would be unable to restart it. If this happened during any maze run, the maze would be considered lost due to validation issues. Again, this was not considered the same as a pressure timer lost maze.

Data analysis

We did two separate analyses for our study. First, we wanted to find out if there would be significant differences for each of our variables, between our two groups. The one with participants that saw the incidental visualization, and the one without the visualization. Then, we wanted to explore possible associations or correlations between variables. From now on, we will refer to our metrics as score (maze performance), wall (number of times the player bumped into a wall), monster (number of times the player bumped into a monster), finish (if the player finished the maze), speed (time between each key press, in milliseconds), patterns (number of times a pattern was made), rating (performance rating for each maze), stepsScore (score for the question asked), and highScore (rating for the high score question). Each trial was analyzed independently. The datasets are available at this link (https://figshare.com/s/295a599145dcf5da7667).

Regarding group differences, our study had a between-subjects design and one independent variable with two groups. Two of our dependent variables were ordinal, the "rating" and "highScore." Therefore, for these two, we used the Mann-Whitney *U* test which is a rank-based non-parametric test that can be used to determine if there are differences between two groups on a continuous or ordinal dependent variable. Then, for the "finish" variable, which is dichotomous, we used the chi-square test of homogeneity, which requires a minimum sample size to provide a valid result. When this requirement failed, we used Fisher's exact test. Finally, for the remaining six variables, that were continuous, we used the independent-samples *t*-test that is used to determine if a difference exists between the means of two independent groups on a continuous dependent variable. More specifically, it would let us determine whether the difference between these two groups was statistically significant. In the cases where the variables were not normally distributed, we decided to use the Mann-Whitney *U* test.

Regarding correlations, we wanted to find correlations in three sets of variables. The first one was between the "OSPAN" and "stepsScore" variables. We wanted to see if working memory influenced how participants performed at the question for each trial. The second set was between the "rating" variable and all other continuous variables, and the third was between the "highScore" variable and all other continuous variables. In these two last sets, our idea was to understand how people perceive their performance and how much they were aiming for the highscore correlated with the results of the other metrics. For the first set of variables, we used the Mantel-Haenszel chi-square test since we treat both ordinal variables as intervally scaled. Then, for the second and third sets, we used Spearman's correlation, a non-parametric measure of the strength and direction of association between two variables measured on at least an ordinal scale.

Results

Although we want to avoid comparing between mazes, from now on, to simplify, we will explain the results per maze (as depicted in Figure 6). Furthermore, we decided to keep data outliers since we know these values were not measurement errors. For all mazes, there were no dependent variables with normal distributions on both groups simultaneously, as assessed by Shapiro-Wilk's test ($p < 0.05$), except the "speed" variable on mazes 2, 4, and 6, and the "score" variable on maze 5. For mazes 1, 3, 4, 5, and 6, we did not have enough data to analyze the "monsters" variable. For maze 3, we did not have enough data to analyze the "patterns" variable.

For all mazes, a Mann-Whitney *U* test was run on each variable to determine if there were differences between groups. However, for the "speed" variable on mazes 2, 4, and 6, and the "score" variable on maze 5, we used the Independent-samples *t*-test. The "wall" variable presented statistically significant differences on mazes 1, 2, 4, and 5. The "speed" variable presented statistically significant differences in maze 1. The “stepsScore” variable presented statistically significant differences on mazes 1, 2, 3, and 4. The “rating” variable presented statistically significant differences in maze 4. Finally, there were no statistically significant differences in each variable on maze 6.

The median value for “wall” was statistically significantly higher with a visualization for mazes 1 (from 2 to 1, $U = 263.500$, $z = -2.511$, $p = 0.012$), 2 (from 2 to 1, $U = 306$, $z = -2.770$, $p = 0.006$), 4 (from 4 to 2, $U = 300.500$, $z = -3.844$, $p < 0.001$), and 5 (from 4 to 2, $U = 280.500$, $z = -2.959$, $p = 0.003$). Therefore, when participants were presented with an incidental visualization, in four mazes, they bumped more into walls.

The median value for speed was statistically significantly lower with a visualization for maze 1 (from 507 to 546 $U = 578.500$, $z = 2.458$, $p = 0.014$). Therefore, when participants were presented with an incidental visualization, in one maze, they played slower.

The median value for “stepsScore” was statistically significantly higher with a visualization for mazes 1 (from mean rank = 35.43 to mean rank = 23.57, $U = 248.500$, $z = -2.799$, $p = 0.005$), 2 (from 2 to 1, $U = 367$, $z = -1.986$, $p = 0.047$), 3 (from 2 to 1, $U = 365.500$, $z = -3.048$, $p = 0.002$), and 4 (from 3 to 2, $U = 235.500$, $z = -4.858$, $p < 0.001$). Therefore, when participants were presented with an incidental visualization, in four mazes, they performed better at the question asked.

The median value for “score” was statistically significantly lower with a visualization for mazes 4 (from 61.50 to 67, $U = 868$, $z = 2.742$, $p = 0.006$) and 5 (from 50.32 ± 9.713 to 55.28 ± 7.968, $t(61) = -2.188$, $p = 0.033$). Therefore, when participants were presented with an incidental visualization, in two mazes, they performed worse at the maze.

Finally, the median value for “rating” was statistically significantly lower with a visualization for maze 4 (from 3 to 4, $U = 818$, $z = 2.289$, $p = 0.022$). Therefore, when participants were presented with an incidental visualization, in one maze, they were less self-confident.

Regarding the number of participants that finished each maze, the results of the chi-square test of homogeneity showed us that when comparing the groups with and without the incidental visualization, there was a non-statistically significant difference. Therefore, in all six mazes, introducing an incidental visualization did not make people lose more times due to the pressure timer.

Next, in all mazes, the Mantel-Haenszel test of trend showed a statistically non-significant linear association between the “stepsScore” and “OSPAN” variables. Therefore, in all six mazes, memory capacity did not influence the participant’s accuracy at answering the question.

Finally, in all mazes except the 6 one, we found some statistical correlations between our dependent variables. However, in all cases, the correlations turned out to be weak.

Table 1. Brief summary of the results, considering the cases when an incidental visualization was presented.

	1	2	3	4	5	6
score						
wall						
monster						
finish						
speed						
patterns						
rating						
stepsScore						
highScore						

In purple cells are the values that increased significantly, and in brown cells the values that decreased significantly. For example, we can see that for four mazes, the score at the questions asked (stepsScore) was significantly higher.

Analysis

In Table 1, you can see filled in brown the variables whose values significantly decreased due to having an incidental visualization. In purple, we can see the variables whose values significantly increased due to having an incidental visualization. Then, regarding correlations, in no maze did the OSPAN score correlated with the questions’ performance, and every time significant correlations using the “rating” and “highScore” variables existed, they were weak. In Figure 10 we can see the results for the dependent variables where existed significant differences, and here is the summary of the statistically significant results, where each trial corresponded to a maze played:

- Participants’ performance in the game significantly decreased only in two trials when using an incidental visualization;
- Participants significantly hit walls more times in four trials when using incidental Visualizations;
- Participants played significantly faster in one trial when using incidental Visualizations;
- Participants were significantly less confident in one trial when using incidental Visualizations;
- Participants performed significantly better at the question asked in four trials when using incidental Visualizations.

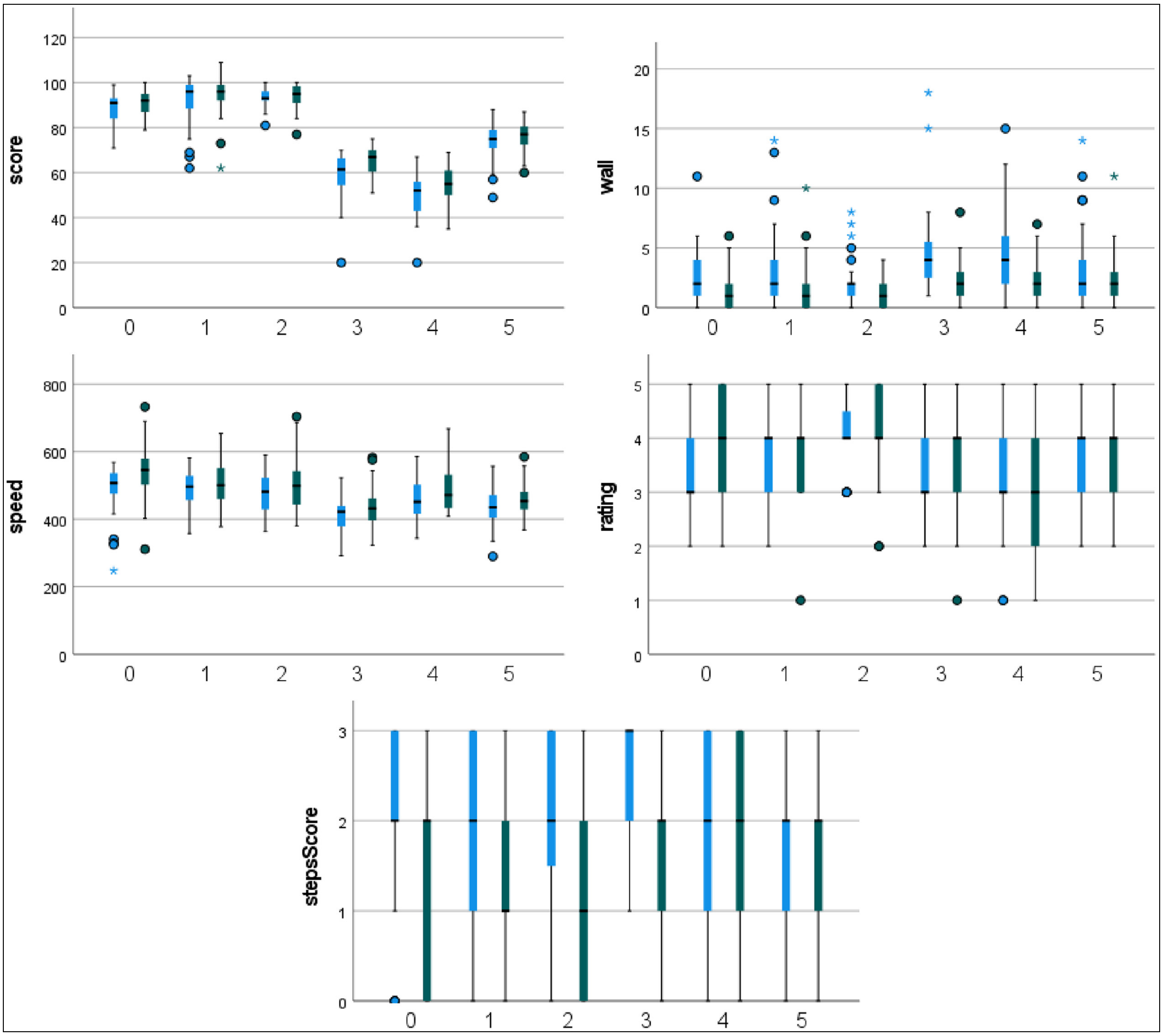

Figure 10. Clustered boxplots for the variables with significant differences. The x-axis corresponds to each maze's index (0 is the first maze, and 5 is the sixth). For each maze, the left boxplot corresponds to the results with the visualization and the right without. We decided to keep data outliers, represented with circles and stars since we know these values were not measurement errors.

Discussion

In this section, we will discuss how our results answer each one of the questions that we presented at the beginning of this document. We also present in Table 2 a brief summary of our results regarding the influence between the tasks performed, incidental visualizations, working memory, and the people's maze runs.

Task performance

Our RQ1 was: Does perceiving an incidental visualization disrupt the primary task performance? By looking at the times players bumped into a wall while playing, we can conclude that the incidental visualizations had a disruptive effect. However, that was not enough to make players always get the corresponding penalty in performance. Only in two of those trials did the performance significantly get lower. Furthermore, only in one of those four disrupted trials did the participant's confidence get lower. Therefore, we argue that the answer to our RQ1 is no. Perceiving an incidental visualization does not disrupt the primary task performance.

Pressure

Our RQ2 was: Does perceiving an incidental visualization increase the amount of stress felt by people? In all

Table 2. Brief summary of influences between several factors. People's performance at the tasks did not get disrupted by the presence of an incidental visualization.

Task		People	Visualization
People	Performance was not disrupted		
Visualization	Accurately conveyed information	Did not influence losses	
Memory	Did not influence performance	Did not disrupt confidence	Did not influence performance

An incidental visualization allows people to receive information during the tasks, and did not make people lose more. Finally, memory did not influence performance at the task and on perceiving the visualization, and it did not influence people's confidence at the tasks.

trials, the completion rate never changed significantly, indicating that introducing an incidental visualization did not change the stress players were feeling. Additionally, this also indicates that participants' focus on their primary task did not differ significantly, even though in some cases performance decreased. Therefore, although they did not play all trials perfectly, if one maze was lost due to the pressure timer, it was not because of the incidental visualization. Therefore, we argue that the answer to our RQ2 is no. Perceiving an incidental visualization does not increase the amount of stress felt by people.

Augmented perception

Our RQ3 was: Are incidental visualizations able to provide actionable information while people are performing a primary task? Our results show that in four trials, having the incidental visualization significantly increased people's awareness of the underlying data of the primary task (directions taken). In the other two, it did not differ significantly. Therefore, in the worst-case scenario, the participant's perception remains unchanged. In only one trial, the awareness of the underlying data may have compromised the maze's score. However, since this happened in one of the six trials, we believe it may have happened because of some specific characteristic of the maze generated. We argue that the answer to RQ3 is yes. Incidental visualizations can provide actionable information while people are performing a primary task.

Working memory

Our RQ4 was: Does working memory influence how well people perform at the primary task and how well they perceive the incidental visualization? First, our results regarding the OSPAN showed that visual memory does not influence how people performed at the question asked at each trial. Therefore, designers may not need to worry about people's memory if they want the information to be perceived incidentally and processed a few moments after. Then, we checked if introducing an incidental visualization would make people pursue more the high score or if it made them more or less confident about their performance. In one trial, players felt their performance was worse when an incidental visualization was presented, which it was. However, the same results do not hold for another trial, where participants' confidence did not significantly change, even though the score did. Finally, in no trial did participants feel the urge to significantly pursue the high score just because the visualization was presented. Therefore, we argue that the answer to RQ4 is no. Working memory does not influence how well people perform at the primary task and how well they perceive the incidental visualization.

Limitations and design guidelines

The biggest limitation in our work is the scenario. All these results most likely only apply to the primary task we tested, and we only used the bar chart as the visual idiom conveying the information. Furthermore, although the maze game turned out to be efficient for designing a user study with an incidental visualization, it may not help to understand people in the real world, and it would be interesting to test different real-world tasks outside a computer. Finally, we only tested incidental visualization using a bar chart, but there may be more alternatives worth studying that encode values with different marks and channels that still allow accurate perception and low destructiveness. These are the major implications of our work:

- Incidental Visualizations can be perceived during an ongoing primary task, and its performance will not usually be influenced by their presence.
- People's focus on an ongoing primary task is usually not disrupted due to incidental Visualizations. People did not lose more because of them, and usually played at the same speed.
- An incidental Visualization can effectively convey information when perceived at-a-glance for short exposure times. Participants' performance at the questions asked usually increased while they were present.

Conclusion and future work

Incidental Visualizations are a specific type of Peripheral Displays that share characteristics with Glanceable Displays, but not with Ambient Displays. Furthermore, they are unique in their context of use because they are supposed to be used side by side during a primary task for short exposure times. We conducted an empiric user study as a starting point to evaluate how Incidental Visualizations impact people while they perform primary tasks. We concluded that overall, they do not decrease the primary task performance, which gives future researchers and designers new windows of opportunities to explore this topic even further.

Acknowledgement

I would also like to thank my colleague Tomás Alves for his help in discussing the work, and with some minor revisions.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by FCT, through grants PTDC/CCI-CIF/28939/2017, UIDB/50021/2020, and SFRH/BD/143496/2019.

ORCID iD

João Moreira https://orcid.org/0000-0002-5461-5217

References

1. Weiser M and Brown JS. Designing calm technology. *PowerGrid J* 1996; 1: 75–85.
2. Matthews T, Forlizzi J and Rohrbach S. Designing glanceable peripheral displays. *UC Berkeley*, September 2006; 23.
3. Tomitsch M, Lehner A and Grechenig T. Towards a taxonomy for ambient information systems. In: *Proceedings of the workshop for ambient information systems, colocated at Pervasive 2007*, Toronto, Canada, May 13, 2007, p.4247.
4. Moreira J, Mendes D and Gonçalves D. Incidental visualizations: Pre-attentive primitive visual tasks. In: *Proceedings of the international conference on advanced visual interfaces*, 2020, pp. 1–9.
5. Neshati A, Sakamoto Y, Leboe-McGowan L, et al. G-sparks: Glanceable sparklines on smartwatches. In: *Proceedings of graphics interface 2019*, Kingston, Ontario, 28–31 May 2019. DOI: 10.20380/GI2019.23.
6. Rodgers J and Bartram L. Exploring ambient and artistic visualization for residential energy use feedback. *IEEE Transactions Vis Comput Graph* 2011; 17(12): 2489–2497.
7. Islam A, Aravind R, Blascheck T, et al. Preferences and effectiveness of sleep data visualizations for smartwatches and fitness bands. In: *CHI conference on human factors in computing systems*, 2022, pp.1–17.
8. Blascheck T and Isenberg P. A replication study on glanceable visualizations: Comparing different stimulus sizes on a laptop computer. In: *IVAPP 2021-12th International conference on computer vision, imaging and computer graphics theory and applications*, pp.133–143. SCITEPRESS-Science and Technology Publications.
9. Lu F and Bowman DA. Evaluating the potential of glanceable ar interfaces for authentic everyday uses. In: *2021 IEEE virtual reality and 3D user interfaces (VR)*, 2021, pp.768–777. IEEE.
10. Shelton B, Nesbitt K, Thorpe A, et al. Assessing the cognitive load associated with ambient displays. *Pers Ubiquitous Comput* 2022; 26(1): 185–204.
11. Liu Z, Crouser RJ and Ottley A. Survey on individual differences in visualization. *Comput Graph Forum* 2020; 39(3): 693–712.
12. Cleveland WS and McGill R. Graphical perception: theory, experimentation, and application to the development of graphical methods. *J Am Stat Assoc* 1984; 79(387): 531–554.
13. Matthews T, Rattenbury T and Carter S. Defining, designing, and evaluating peripheral displays: an analysis using activity theory. *Human–Computer Interaction* 2007; 22(1–2): 221–261.
14. Matthews T. Designing and evaluating glanceable peripheral displays. In: *Proceedings of the 6th conference on Designing Interactive systems*, 2006, pp.343–345. ACM.
15. Gouveia R, Pereira F, Karapanos E, et al. Exploring the design space of glanceable feedback for physical activity trackers. In: *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, 2016, pp.144–155. ACM.
16. Blascheck T, Besancon L, Bezerianos A, et al. Glanceable visualization: Studies of data comparison performance on smartwatches. *IEEE Transactions Vis Comput Graph* 2019; 25(1): 630–640.
17. Tarantino L, Angelucci D, Bonomo A, et al. Design and applications of GLANCE: GLanceable alarm notification for a user centered experience. *Appl Sci* 2021; 11(2): 669.
18. Daskalogrigorakis G, McNamara A and Mania K. Holo-box: Level-of-detail glanceable interfaces for augmented reality. In: *ACM SIGGRAPH 2021 Posters*, 2021, pp.1–2. Association for Computing Machinery.
19. Lu F. [dc] glanceable ar: Towards an always-on augmented reality future. In: *2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, 2021, pp.717–718. New York: IEEE.
20. Lu F, Davari S and Bowman D. Exploration of techniques for rapid activation of glanceable information in head-worn augmented reality. In: *Symposium on spatial user interaction, SUI '21 Association for Computing Machinery*, Virtual Event, USA, 2021, pp.1–11. ACM.
21. Imamov S, Monzel D and Lages WS. Where to display? How interface position affects comfort and task switching

time on glanceable interfaces. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp.851–858. New York: IEEE.
22. Wilkinson D, Bahirat P, Namara M, et al. Privacy at a glance: the user-centric design of glanceable data exposure visualizations. *Proc Priv Enh Technol* 2020; 2020(2): 416–435.
23. Skog T, Ljungblad S and Holmquist L. Between aesthetics and utility: designing ambient information visualizations. In: *IEEE Symposium on information visualization 2003 (IEEE Cat. No.03TH8714)*, pp. 233–240. DOI: 10.1109/INFVIS.2003.1249031.
24. Morrison-Smith S, Chilton LB and Ruiz J. AmbiTeam. *XRDS* 2021; 28(2): 60–65.
25. Bressa N, Korsgaard H, Tabard A, et al. What's the situation with situated visualization? A survey and perspectives on situatedness. *IEEE Transactions Vis Comput Graph* 2022; 28(1): 107–117.
26. Bentley F, Tullio J, Metcalf C, et al. A time to glance: Studying the use of mobile ambient information. In: *Workshop at Pervasive 2007 designing and evaluating ambient information systems*, 2007, p.1.
27. Skog T. Activity wallpaper: ambient visualization of activity information. In: *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, 2004, pp.325–328.
28. Moere AV. Beyond the tyranny of the pixel: Exploring the physicality of information visualization. In: *2008 12th International conference information visualisation*, 2008, pp.469–474. New York: IEEE.
29. Pousman Z, Stasko J and Mateas M. Casual information visualization: depictions of data in everyday life. *IEEE Transactions Vis Comput Graph* 2007; 13(6): 1145–1152.
30. Conway ARA, Kane MJ, Bunting MF, et al. Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev* 2005; 12(5): 769–786.
31. Case R, Kurland DM and Goldberg J. Operational efficiency and the growth of short-term memory span. *J Exp Child Psychol* 1982; 33(3): 386–404.
32. Turner ML and Engle RW. Is working memory capacity task dependent? *J Mem Lang* 1989; 28(2): 127–154.
33. Daneman M and Carpenter PA. Individual differences in working memory and reading. *Journal Verbal Learn Verbal Behav* 1980; 19(4): 450–466.
34. Unsworth N, Heitz RP, Schrock JC, et al. An automated version of the operation span task. *Behav Res Methods* 2005; 37(3): 498–505.
35. Jevons WS. The power of numerical discrimination. *Nature* 1871; 3(67): 281–282.
36. Mandler G and Shebo BJ. Subitizing: an analysis of its component processes. *J Exp Psychol Gen* 1982; 111(1): 1–22.
37. Haroz S and Whitney D. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions Vis Comput Graph* 2012; 18(12): 2402–2410.
38. Katzin N, Cohen ZZ and Henik A. If it looks, sounds, or feels like subitizing, is it subitizing? A modulated definition of subitizing. *Psychon Bull Rev* 2019; 26(3): 790–797.
39. Borgo R, Micallef L, Bach B, et al. Information visualization evaluation using crowdsourcing. *Comput Graph Forum* 2018; 37(3): 573–595.
40. Borgo R, Lee B, Bach B, et al. Crowdsourcing for information visualization: promises and pitfalls. In: Archambault D, Purchase H and Hoßfeld T (eds) *Evaluation in the crowd. Crowdsourcing and human-centered experiments*. Cham: Springer International Publishing, 2017, pp.96–138.
41. Quadri GJ and Rosen P. A survey of perception-based visualization studies by task. *IEEE Trans Vis Comput Graph* 2022; 28: 5026–5048.
42. Castro SC, Quinan PS, Hosseinpour H, et al. Examining effort in 1d uncertainty communication using individual differences in working memory and NASA-TLX. *IEEE Transactions Vis Comput Graph* 2022; 28(1): 411–421.
43. Gadiraju U, Kawase R, Dietze S, et al. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp.1631–1640.