Article in Press

IJIMAI journal

Modulating the Gameplay Challenge Through Simple Visual Computing Elements: A Cube Puzzle Case Study

Jose Ribelles, Angeles Lopez, V. Javier Traver*

Institute of New Imaging Technologies, Universitat Jaume I, Castellón (Spain)

Received 24 April 2021 | Accepted 3 March 2022 | Early Access 2 May 2022

LA UNIVERSIDAD EN INTERNET

Abstract

Positive player's experiences greatly rely on a balanced gameplay where the game difficulty is related to player's skill. Towards this goal, the gameplay can be modulated to make it easier or harder. In this work, a modulating mechanism based on visual computing is explored. The main hypothesis is that simple visual modifications of some elements in the game can have a significant impact on the game experience. This concept, which is essentially unexplored in the literature, has been experimentally tested with a web-based cube puzzle game where participants played either the original game or the visually modified game. The analysis is based on players' behavior, performance, and replies to a questionnaire upon game completion. The results provide evidence on the effectiveness of visual computing on gameplay modulation. We believe the findings are relevant to game researchers and developers because they highlight how a core gameplay can be easily modified with relatively simple ingredients, at least for some game genres. Interestingly, the insights gained from this study also open the door to automate the game adaptation based on observed player's interaction.

Keywords

Computer Game, Gameplay Modulation, Player Experience, Visual Computing.

DOI: 10.9781/ijimai.2022.05.001

I. Introduction

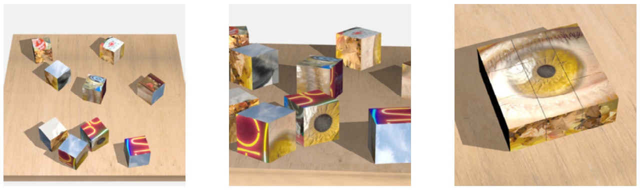
A good understanding of the factors affecting the player experience can be very important for producing better games. Accordingly, much research effort has been devoted to both gaining theoretical insights on, and quantifying, this experience. Game heuristic and guidelines can be proposed to elicit desirable human (player) emotions and responses. Engagement is one of the concepts that has been studied, with the Game Engagement Questionnaire (GEQ) [1] being proposed to measure it. Validation studies [2] of game experience scales have been performed [3]. For game challenge, a recent scale has been developed and validated [4], [5]. Flow [6] is also widely studied in the context of video games, for better understanding it [7], [8], measuring it [9]–[13], relating it experimentally to learning and other conditions [14]–[17], or even providing some design intuitions or guidelines to produce it [8], [18].

Many of the aspects of the player experience are interrelated, so that flow, engagement [1], enjoyment [19]–[22], and others such as immersion [23], may overlap and share common attributes. For instance, flow is one of the factors within the GEQ [1]. Interestingly, most conceptualizations of the player experience share the view that the *optimal complexity level* is a key ingredient of an enjoyable game, and it is particularly essential for flow. This optimal experience relies on a balance between the challenges of the game and the player's skill; in essence, if the game is too simple or too challenging for the player, it will lead to either boring or frustrating experiences, respectively [24].

* Corresponding author.

E-mail address: vtraver@uji.es

Our work has to do mainly with the game challenge; in particular, we delve into this issue: whether the player experience can be modulated with relatively simple manipulations of images present in the game. By *image manipulation* we mean modifying some other aspects of the image [25] (e.g. color, edges, texture) or distorting it somehow (e.g. blurring, geometric deformations, frequency filtering), but respecting some aspects of its contents, so that it can be recognized and potentially distinguished from others, yet with different cognitive abilities or effort with respect to the original image. This possibility can be partially grounded on the information quality, which is understood as one of the mechanisms underlying flow [8]. We generally refer to these alterations as *visual computing* [26]. Arguably, visual computing is particularly suitable for some game types or genres such as puzzle or card games, but it could also be applied to other games where the modification of visuals of some of its elements, may have an impact on the gameplay. To address this issue, we focus on a controlled case study, a 2D video game version of the Cube blocks puzzle (Fig. 1), and explore whether, and how, changes to the images used for the sides of the cubes affect the actual gameplay and the player perception of their experience.



(a) (b) (c)

Fig. 1. A rendering of Cube blocks, a classic puzzle game: (a,b) two views of the unsolved puzzle and (c) once the puzzle is solved for a target image of a human eye. The prototype game used in this paper consists of solving a series of these puzzles, in a 2D (top-view) interface.

Please cite this article in press as:
J. Ribelles, A. Lopez, V. J. Traver. Modulating the Gameplay Challenge Through Simple Visual Computing Elements: A Cube Puzzle Case Study, International Journal of Interactive Multimedia and Artificial Intelligence, (2022), http://dx.doi.org/10.9781/ijimai.2022.05.001

The contributions of this work are as follows. The possibility of modulating the gameplay with visual computing elements is hypothesized. A web-based cube blocks puzzle has been developed as a proof-of-concept prototype. A user study has been conducted to find out possible behavioral, performance and opinion differences of players in one control and one experimental group. A detailed analysis of the results has been performed. Results provide interesting findings that can inform both game level designs and, eventually, automatic and dynamic game difficulty adjustment.

In the following, section II reviews related work on image manipulation and gameplay modification of some sort. Section III introduces the core game and its version with visual computing. Section IV details the design and implementation of both the experiment and the interface. Results are carefully analyzed and discussed in section V. Some discussions on the potential and limitations of the work are provided in section VI, and concise conclusions are finally presented in section VII.

II. Previous Work

In general terms, it is well accepted the role that visual contents in all kinds of media play on our emotions, and that they can even affect our beliefs. This fact has ramifications on how deliberate modifications of these contents may end up affecting our attitudes and behaviors [27]. The players' visual attention patterns can inform the level design for game improvement [28]. By masking some kinds of visual information affecting human prior knowledge, the human performance in a platform game has been found to decay notably with respect to the original unmodified game [29]. The relation between playing some games and the perceptual and cognitive abilities that are developed [30]–[34] may eventually bring insights into game design by reversing the problem, namely, how can we affect performance by modulating the gameplay. The interesting relationship between perception, computer graphics and video games are certainly not new, but with a large margin for further exploration [35].

Next, we briefly review ideas previously considered for visual or challenge modifications, which for convenience are also summarized in Table I. As a first simple choice, players may select game levels when available, although this choice not only relies on the players' ability to do the selection adequately, but also on a proper difficulty design by game developers, an area where more research and designer-assistance tools are required [36]. Similarly, modding can be seen as a form of game modification, but it is motivated more by the need of self-expression of gamers [37], than aimed at a carefully planned gameplay modification. Schell [38] mentions a few examples of subtle visual cues that can indirectly guide the player's actions with the goal of providing the player with a sense of perceived freedom, without actually enjoying full freedom. Similarly, in Mirror's Edge [39], some game elements are highlighted in red as explicit cues to navigation.

In games, the term *juiciness* refers to an abundance of audiovisual effects [40], as a form of additional feedback (more than strictly required from a usability point of view), often including second-order motion, which seeks to provide the player with plenty of power and rewards [38]. Since the term juiciness is somehow vague, a recent survey tried to elicit developers' understanding, in an attempt to provide a useful framework for the game design and research communities [41]. Regarding the positive or negative effects of "juicy" games, findings on juicy-vs-dry game versions are somehow disparate and thus results remain essentially inconclusive [42]–[47]. For instance, it has been observed that while the perceived competence is positively affected by juiciness, the actual performance is not really changed [47]. However, a recent large-scale empirical study with four levels of juiciness in a role-playing game [48] reveals that the degree of juiciness has an impact on the valence of the effects on performance, experience and motivation, with moderate amounts of juiciness found to be optimal. Although interesting and akin to our work, this form of game modification has more to do with user interaction and includes more effects (motion, audio) than the ones considered in our work.

TABLE I. Summary of (Visual) Game Modifications and their Main Purpose

Approach	Main purpose
Level selection	Match game challenge to player's skill
Modding	Allow gamers' self-expression
Subtle visual cues	Guide players' actions
Juiciness	Endow the player with rewards and sense of power
Flipped levels	Expand the game variety
Embellishments	Increase the engagement
Non-photorealistic rendering	Provide artistic styles and moods
Color adaptations	Augment accessibility
Visual computing (proposed here)	Support gameplay (challenge) modulation

In some games, previously played levels are flipped horizontally and, optionally, other components such as the art style also changed, which is a simple means of having more levels to play; apparently these alterations can make the game harder to play [49], [50]. Embellishments in game skins are found to increase engagement but decrease performance [51]. Color in some game elements may also have some effects. Performance, immersion and flow have been found to be inferior with red avatars than with blue ones [52]. Some games include color-blindness adaptation [53], [54] some of which may result in (unintended) easier playing for people with normal vision. Non-photorealistic techniques were proposed as a means of providing the players with different styles and moods [55], but their impact in terms of gaming experience was not studied. Some past video games included effects such as toon-rendering, comic-like appearances, sumi-e painting or pointillist style [56]. Finally, procedural content generation techniques [57] can be rather sophisticated, a current research hot topic and orthogonal to our study.

Although all these are means to modify some visual aspect of games, the mechanisms and purposes are not always aimed at modulating the gameplay, as summarized in Table I. Therefore, as far as we know, using visual computing to modulate the challenge in the gameplay has scarcely been considered in the literature, and detailed studies in the sense intended in this work are limited.

To summarize, this work addresses the use of visual modification for modulating the difficulty or challenge of the gameplay by means of a case study. Notice that it is not aimed at making the players achieve the state of flow, which could be seen as an ultimate desirable goal, but it is out of the scope of this present work. Since the flow depends both on the game and on each particular player, the flow goal partially relates to the interesting concept of dynamic difficulty adjustment [36], [58]–[60], and the findings of our study can be expected to complement and inform subsequent research in this area. If we were able to characterize a player in terms of their skills on the one hand, and categorize different image modifications (including no modification at all) in terms of the degree of difficulty they induce on the other hand, then a properly balanced skill-difficulty match could be chosen for an improved game experience (Fig. 2).

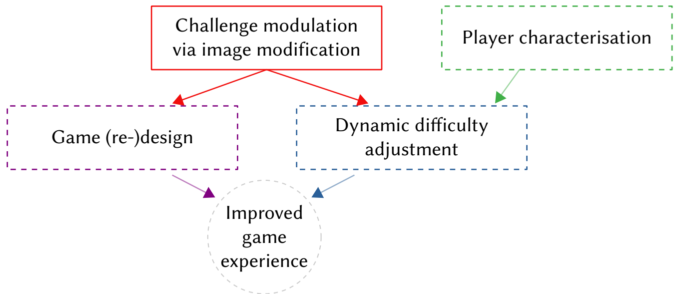


Fig. 2. This work focuses on exploring game challenge modulation through image modifications. Eventually, this might help (re-)design games for an improved game experience. Such an improvement might also be obtained together with player characterization through dynamic difficulty adjustment. Dashed lines are used to highlight that these modules are not part of this work.

III. Methodology

To evaluate the possibility of modulating the gameplay with visual computing elements, we developed a web-based cube puzzle game as a proof-of-concept prototype and conducted a user study. The basic gameplay is first introduced in section A, the selection of the visual concepts for this case study are motivated, and their implementation within the game is discussed in section B. Other visual concepts are reviewed, although intended for future exploration, in Section C. Finally, the actual images used in both versions of the game as well as their grouping to design the puzzles are detailed in Section D.

A. Baseline Gameplay

The Cube Puzzle game consists of a series of 6 puzzles, each with 9 pieces arranged in a 2D 3 3 layout (Fig. 3). Each piece represents a cube, and each of its six sides has a subimage of one of the six possible full images that can be formed with the 9 cubes. Only one of the six faces of each cube is displayed at a given time, and this face can be changed by rotating the cube in any of four directions (by clicking on the triangular marks) to reveal the corresponding neighbor face. In addition, the displayed face can be repeatedly rotated 90˝ clockwise by clicking on the corresponding inner arrows. Each puzzle is solved when all the cubes display the corresponding part of a target image and at the appropriate orientation, i.e. when the target image is formed. Although each cube has six faces, only one target image per puzzle has to be formed. Different puzzles have different images on the sides of the cubes. The images used for the six puzzles in this baseline version of the game are given in the first six rows in Table II. The rest of the information in this table is described where relevant in the following sections. A game is successfully complete when all the 6 puzzles have been solved. Note that this number (6) of puzzles has to do with the number of visual concepts we decided to work with, not with the number of sides of a cube.

B. Visual Concepts Considered

The baseline gameplay was enriched with visual computing (VC) elements (or visual concepts, for short) so that we can test our main hypothesis that this kind of visual concepts can be useful to modulate the gameplay. Much is known on the impact on human visual perception of computer-generated graphics [26], even though a lot of practical insight has yet to be gained by the application of this body of knowledge to games, and to human-computer interfaces at large. For this first study, we decided to focus on the following three visual concepts, which are highly relevant to, and cover, three different areas of human visual perception: edges (*e*), color (*c*), and dynamics (*d*).

Edges can roughly represent a given object, but they can also be insufficient for object recognition [61]. Therefore, edges may challenge some visual-perception-based task.

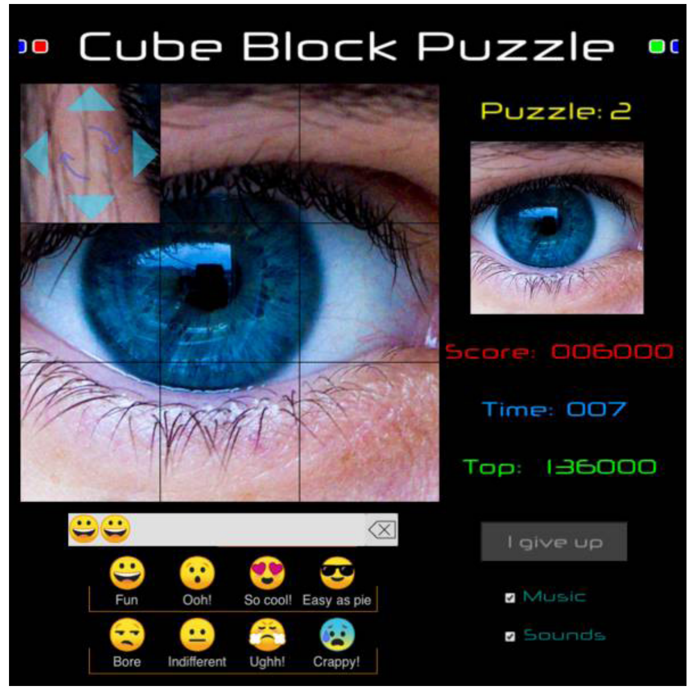


Fig. 3. Game interface with one of the puzzles. In this case, just a single cube (the one in the upper-left corner) remains to be aligned to complete the puzzle. The target image is a human eye in this case, which is displayed as a thumbnail for reference while playing.

Color is another visual attribute whose perception has also potential implications for games. For instance, color perception is known to differ across the visual field of view [62], and it is influenced by 3D shape perception [63]. Interestingly, brain activation from actual observations of color are found to be similar to *implied* observations, i.e. not actually observed, but known from prior knowledge of objects [64].

Motion is a powerful visual cue that can enhance the perception, particularly when other visual conditions are poor [65]. For our purposes, motion and image dynamics in general provide a wide range of concepts where time-varying contents can be used to modulate the discriminability of images. For instance, visual acuity of static stimuli is found to be superior than that of dynamic stimuli in the fovea, but not in the periphery [66], and perception is influenced by temporal frequency and occlusion [67].

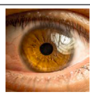
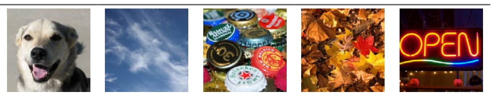
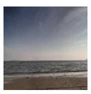
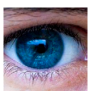
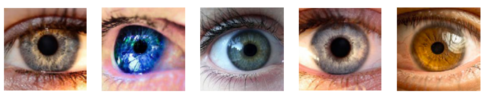
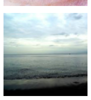
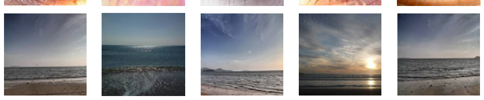
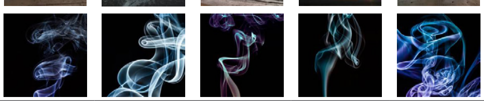
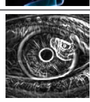
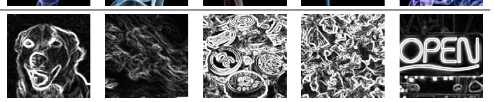
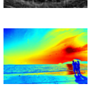
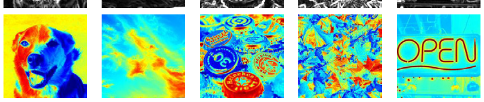
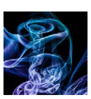
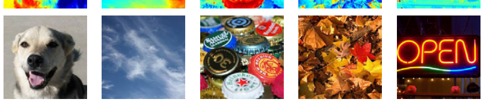
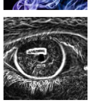
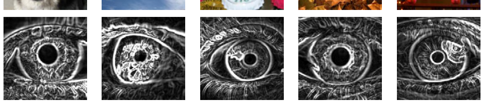
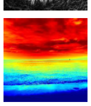
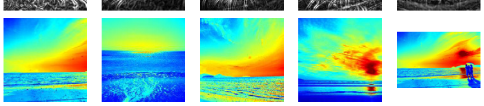
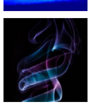
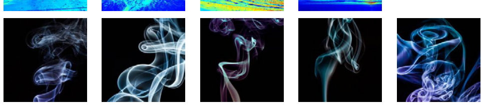
Implementation details. Regarding the implementation of these concepts, for edges (puzzles $\mathsf{A}_{vc:e}$ and $\mathsf{A}_{\widetilde{vc}:e}$ in Table II), spatial image gradients were computed on the corresponding gray-level image, and then scaled for visualization purposes. Algorithm 1 shows the steps used in this work to convert a color image to an edge image. First, the image is converted to gray-scale by using the NTSC standard where each color channel is weighted differently ($w_r = 0.30$, $w_g = 0.59$, $w_b = 0.11$) to account for human color perception, which is more sensitive to green [25]. In order to reduce the noise effect, the gray-scale image is smoothed by average filtering [25], that is, a linear convolution, $I_g * A$, of the gray-scale image I_g with the filter mask A of size $n \times n$, with $A_{ij} = 1/n^2$. Spatial gradients [25] are computed from the smoothed image as $[G_x, G_y] = \nabla I_s = [\partial I_s / \partial x, \partial I_s / \partial y]$. The gradient magnitude is approximated by the sum of the absolute values of the gradient in both directions. We finally apply a scale factor s to the gradient magnitude for visualization purposes, and clamp the result to the valid gray-level range [0,255]. In this work, we used $n = 11$ for smoothing; spatial gradients were approximated with simple neighbor differences,

$$G_x(x,y) = \frac{\partial I_s}{\partial x}(x,y) \approx \frac{I_s(x+1,y) - I_s(x-1,y)}{2} \quad (1)$$

$$G_y(x,y) = \frac{\partial I_s}{\partial y}(x,y) \approx \frac{I_s(x,y+1) - I_s(x,y-1)}{2} \quad (2)$$

and $s = 30$ for scaling to visually emphasize the edges.

TABLE II. Images Used in the Different Puzzles for Each of the 6 Faces of Each Cube, I_i, $i \in \{0; ...; 5\}$. Each Image Is Divided Into 3 × 3 Images in a Regular Grid. These Subimages Are the Faces of Each Cube in the Puzzle (As Shown in Fig. 1 and Fig. 3). The Goal Is to Rotate Individually the Cubes So That the Visible Face of the Cube Correspond to the Respective Subimage in I0, Thus Forming the Target Image

Puzzle	Target image (I_0)	Rest of images (I_1, I_2, I_3, I_4, I_5)
A_{st}		
B_{st}		Same as A_{st}
C_{st}		Same as A_{st}
$A_{\widetilde{st}}$		
$B_{\widetilde{st}}$		
$C_{\widetilde{st}}$		
$A_{vc:e}$		
$B_{vc:c}$		
$C_{vc:d}$		
$A_{\widetilde{vc:e}}$		
$B_{\widetilde{vc:c}}$		
$C_{\widetilde{vc:d}}$		

Algorithm 1. Edge computation
Input: Color image with color channels R, G, B
Output: Edge map as image

$I_g \leftarrow w_r \cdot R + w_g \cdot G + w_b \cdot B$ ⊳ convert to gray-scale image
$I_s \leftarrow$ `averageFilter` (I_g, n) ⊳ smooth image, with filter size n
$G_x, G_y \leftarrow$ `spatialGradients` (I_s) ⊳ vertical & horiz. image gradients
$M \leftarrow |G_x| + |G_y|$ ⊳ approximation to gradient magnitude
return min $(s \cdot M, 255)$ ⊳ scale and clamps to range [0, 255]

For color, a color map ("false color") was applied (puzzles $B_{vc:c}$ and $B_{\widetilde{vc:c}}$ in Table II). In this work, we use the known JET color map that we apply after converting the original image to gray-level scale (Algorithm 2). This color map and pseudo-color processing in general, are common aspects in image processing [25]. Color maps are functions (maps) from a scalar value (an index) in a range (typically gray levels in the range [0,255]) to tuples defining colors (e.g. RGB values). The JET color map (Fig. 4) produces a smooth transition from cold to warm colors as the index varies from 0 (the darkest) to 255 (the lightest).

Algorithm 2. Color modification
Input: Color image with color channels R, G, B
Output: Image in JET color map

$I_g \leftarrow w_r \cdot R + w_g \cdot G + w_b \cdot B$ ⊳ convert to gray-scale image
return `applyColorMap(`I_g`, 'JET')` ⊳ apply JET color map

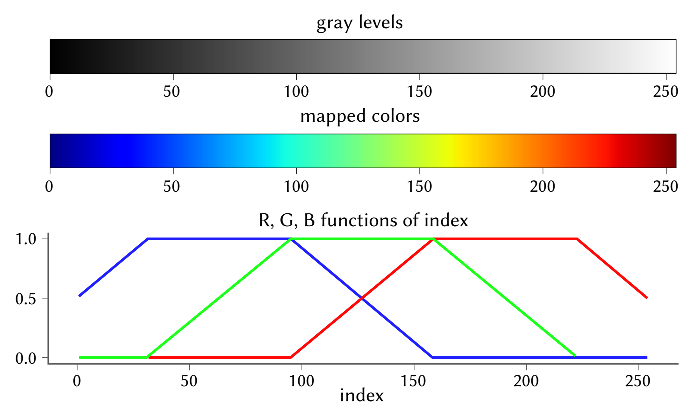


Fig. 4. JET color map. When applied to gray values (top) the corresponding colors (middle) are defined by specific functions (bottom).

Finally, the dynamics concept was implemented by clockwise rotating independently each of the six images around their center at a constant angular speed of about $\omega = 20°/s$. Therefore, when one subimage in one of the cube's sides is visible, the corresponding rotating motion is displayed, rather than a static image, as illustrated in Fig. 5. Such effect is applied to all images in puzzles $C_{vc:d}$ and $C_{\widetilde{vc:d}}$ in Table 2. Note this transformation (Algorithm 3) is computed in real-time on a graphic processing unit (GPU), by means of the fragment shader [68]. To this end, images are provided as textures and faces are provided with 2D texture coordinates. So, for each frame, texture coordinates are transformed by the 2D rotating transformation matrix (3):

$$R(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \quad (3)$$

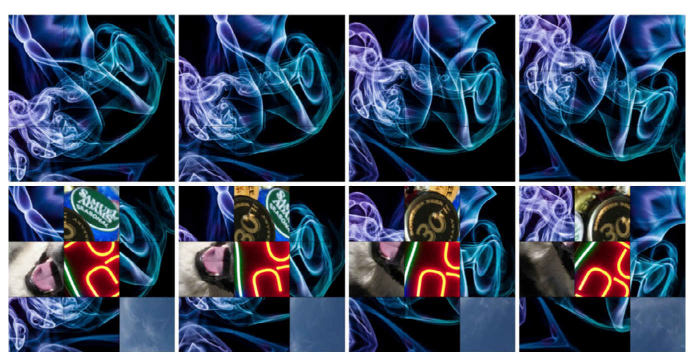


Fig. 5. Illustration of the dynamic effect. All images are rotating, including the target image (above), but only the parts corresponding to the disclosed cube side are visible at a given moment (below). This animation corresponds to 4 seconds, one frame per second, and the target image is visible in five out of the nine cubes.

Algorithm 3. Updating rotation effect, at each frame

Input: angular speed (ω, degrees/second), and elapsed time (t, seconds)

$\varphi \leftarrow$ `rotationAngle` (ω, t) ▷ update rotation angle

$R \leftarrow$ `rotationMatrix` (φ) ▷ create 2D clockwise rotation matrix

Send R to fragment shader ▷ for rotation of all textures coordinates

For the experiment, all the color images were of 512 × 512 pixels, and the visual modifications corresponding to edges and color change were precomputed and used in the video game afterwards, while the dynamic effect was performed in real-time.

C. Other Visual Concepts

Besides these three concepts, we comment on a few other interesting possibilities that might be explored in the future. Due to its fundamental importance in our daily lives, human faces have been extensively studied from both computer and human perception points of view [69]–[71], and might be a good concept to include and study in the context of digital games; for instance, face-based environments may increase users' engagement [72].

Pixelation is a simple but powerful mechanism to modify images so that their perception can be facilitated or hindered [73], and it lends itself to be included in video games for a variety of purposes. However, pixelation in successful games has only or mostly been used for its nostalgic or art style values [74], [75].

Visual illusions are an important area of vision research since they provide great insights into the functioning, abilities and limitations of our visual system [76]. Interestingly, it was recently found that machines cannot (yet) understand nor create such illusions [77]. Consequently, illusions might arguably be another ingredient in video games, a topic worth exploring in the future.

D. Target Images

Regarding the target images, three generic visual entities are considered: human (facial) attributes, natural landscapes, and abstract shapes. The particular images chosen for representing these entities were eyes, beaches, and smokes, respectively (column I_0 in Table II). One main reason for these images is that each of them lend themselves, respectively, to each of the visual concepts chosen. Thus, edges were applied to eye images, color modification to the beach scenes, and dynamics to the smoke images.

Additionally, for each puzzle, two degrees of likeness (`distinct` and `alike`) are considered depending on which images are used for each of the remaining five faces (columns I_1–I_5 in Table II) of the cubes other than the face for the target image. For the `distinct` version, images very different from the target ones are used, whereas the `alike` version uses distracting images that look similar to the target image. For instance, for the eyes case, other human eyes with similar appearance are used as well.

Summarizing, there are two versions of the game: ST (after "standard") as the original one, without visual computing; and VC (after "visual computing"), which uses the three visual concepts (edge, color and dynamics), one per target image. For each version of the game (ST and VC), two factors are considered: the three target images (eyes, beaches and smoke), and the type of distracting images (distinct and alike). In a way, using distracting images might be seen as an additional VC to consider, since it may affect the gameplay as well. Thus, this hypothesis will also be tested experimentally.

Table III collects all this information for convenience and quick reference. We will refer to each of the three puzzle pairs by different target images (A, B, C). The alike versions are denoted as $A_{\tilde{i}}$, $B_{\tilde{i}}$, and $C_{\tilde{i}}$ (for the corresponding $i \in$ {st, vc}). This notation will allow us to compactly and easily refer to the different puzzles, in particular when comparing their results. As an example, the puzzle shown in Fig. 3 corresponds to $A_{\widetilde{st}}$, and that shown in Fig. 5 corresponds to $C_{vc:d}$ as images are rotating. All images used in all puzzles are given in Table II. These are all static images except for puzzles $C_{vc:d}$ and $C_{\widetilde{vc:d}}$, which move as indicated.

TABLE III. The Conditions and Notation to Refer to the 12 Different Puzzles Used, 6 Puzzles Per Game Condition. TABLE 2 Provides Useful Complementary Information to Understand Each Possible Puzzle

Puzzle	Target image	Other 5 images	Likeness	Visual concept	Condition ST	Condition VC
A	eye	not eyes	distinct	edges (e)	A_{st}	$A_{vc:e}$
		other eyes	alike (~)		$A_{\widetilde{st}}$	$A_{\widetilde{vc:e}}$
B	beach	not beaches	distinct	color (c)	B_{st}	$B_{vc:c}$
		other beaches	alike (~)		$B_{\widetilde{st}}$	$B_{\widetilde{vc:c}}$
C	smoke	not smokes	distinct	dynamics (d)	C_{st}	$C_{vc:d}$
		other smokes	alike (~)		$C_{\widetilde{st}}$	$C_{\widetilde{vc:d}}$

IV. Design and Implementation

We review the rationale behind the design of the experiment in section A, the user questionnaire after game completion in section B, and the interface in section C. Details of the user study and of the implementation are given in section D and section E, respectively.

A. Experiment Design

Two user groups are considered, the control group (ST) and the experimental group (VC). A between-subjects protocol was preferred over within-subjects one, since less time and effort is required from each participant, and the subjects do not need to compare directly both versions of the game. Although the between-subjects protocol requires more subjects, this was not found to be an issue, because we planned to do an online experiment and expected to recruit participants with relative ease. Interaction logs were saved per session in order to collect quantitative data in terms of time to completion, number of puzzles successfully solved, number of clicks required, etc. Additionally, subjective data was captured via a final opinion questionnaire (Sect. B) to understand how easy, entertaining, or enjoyable the game was perceived by players. Qualitative feedback from users data was also gathered during and at the end of each puzzle (Sect. C). Table IV summarizes the main descriptions of the experimental design.

TABLE IV. Summary of Main Descriptions of the Experimental Design

Study type	Online game user study for about 1,5 months. Each participant played 6 puzzles for up to about 15 minutes (Sect. IV-D).
Experimental design	Between-subjects (Sect. IV-A): • Control group (ST): original (standard) images. • Experimental group (VC): images with visual modifications.
Assignment to experimental condition	Uniformly at random (either ST or VC)
Statistical hypothesis testing	Mann–Whitney U test (Fig. 8), Kruskal-Wallis H-test (Fig. 8), χ^2 test (Sect. 5-C-1), Z score test (Tables XII and XIII)
Exploratory analysis	Confidence intervals (Fig. 6) and effect sizes, Cohen's d (Fig. 7)
Intervention	Visual modifications and experimental conditions (Table II and Table III)

Online participants were assigned randomly either to the ST or the VC conditions, but they were not given any information about the game conditions so that no expectation could bias their judgment and play, a danger that has been identified [60]. All users are requested to complete 6 puzzles, but they can cancel a particular puzzle if they prefer so (Sect. C). To avoid a presentation-order effect, the three pairs of puzzles were randomly presented, but within each pair, the `distinct` puzzle was always presented before the `alike` one. For instance, one possible presentation order in the control group might be B_{st}, $B_{\widetilde{st}}$, C_{st}, $C_{\widetilde{st}}$, A_{st}, $A_{\widetilde{st}}$, and a possible presentation order in the experimental group might be $A_{vc:e}$, $A_{\widetilde{vc:e}}$, $B_{vc:c}$, $B_{\widetilde{vc:c}}$, $C_{vc:d}$, $C_{\widetilde{vc:d}}$. Thus, a total of $3! = 6$ different presentation orders are possible.

To avoid an effect due to the type or contents of the images, the same images were given to all participants, except obviously for the corresponding VC modifications in the control group. Since it is important to find out the completion times, the game is timed and the player is asked to proceed as quickly as possible. To prevent interruptions from being included in the total elapsed time, the player is offered to take breaks after each puzzle, if desired, and this time is not considered. The elapsed time is displayed while forming a puzzle (Fig. 3) to provide the user with feedback and awareness of the time.

B. Opinion Questionnaire

The contents of the questionnaire (Table V) were designed to learn about how much the player enjoyed the game as a whole (Questions 2 and 4), and their subjective perception of elapsed time (Question 3) and performance (Question 1). We were also interested in finding out whether they enjoyed or hated some particular puzzle (Questions 13–14). To better understand in terms of what they liked or disliked the game, their overall experience was assessed in four dimensions (Questions 5–8). A drawback of the between-subject approach is that, since each participant only plays one version of the game, they cannot be asked which one they prefer. To cope with this, users were given a number of widely known reference games to compare with (Questions 9–12), so that the two versions of the game could eventually be comparable indirectly via these reference games.

No.	Question
1	In general, I found easy to complete the game
2	I found the game entertaining
3	I think I was quick in solving the puzzles
4	Would you play this game again?
5-8	I found the overall experience to be...
5	... entertaining | boring
6	... simple | complex
7	... surprising | dull
8	... exciting | frustrating
9-12	I liked this game (significantly more | more | similarly to | less | significantly less) than
9	... Mahjong
10	... Solitaire
11	... Classical puzzle
12	... Sliding puzzle
13	Which puzzle did you like the most?
14	Which puzzle did you like the least?

The parts of the questionnaire and the possible answers in each question were as follows:

- Questions 1–4 could be answered with a 5-level Likert scale (from "strongly disagree" to "strongly agree"), plus a "No answer" choice.
- Questions 5–8 inquired about particular dimensions of the overall experience, and the possible answers were as given (Table V), plus a "No answer" choice.
- Questions 9–12 seeks to find out how much the player likes this particular game in comparison to the reference games given. A representative image of each reference game is given to help the player recognize the game.
- Questions 13–14 are about which puzzle they liked the most or the least, and they are offered to choose among the six puzzles. To facilitate their recognition, both the target and the remaining images are presented.

Notice that Questions 1 and 3 relate to the *perceived* effort, and the main motivation for including them is to relate them with the measured, actual effort in terms of times and number of clicks required to complete the puzzles. Our intuition was that a lack of correlation between the perceived and the actual efforts might provide some clue on whether the subjects enjoyed the game. For instance, if they underestimate their effort, this might mean they have been losing track of time, which is a sign of engagement [1], without explicitly asking the participant on that.

C. Interface Design

Some interface design details that are particularly important for this study are now briefly discussed. Before starting with the actual puzzles, the player is offered to play a first puzzle with neutral images so as to make sure they understand how to play. We will be referring to this initial "learning" puzzle as `L`.

It turns out that it can be very frustrating for a player not to be able to complete one puzzle that can be found particularly challenging, since this would prevent them to continue with the remaining puzzles. This is addressed with a "Give up" button (Fig. 3). However, to help preventing a player from giving up too easily, this button only becomes active after two minutes from the starting time of each puzzle. As a form of feedback, time is displayed as a countdown in seconds, starting with 120. When it gets to zero, the game is still playable and the player can give up.

The scoring captures the performance in terms of completing each puzzle and parts of it, and the time elapsed (Table VI). The first test puzzle is not included in the scoring. This score was included to provide the player with a sense of an actual game, rather than as a reliable measure of performance to be analyzed. When the sides for all the 3 cubes in a row or in a column are correct, the player is provided with feedback, which is also a form of visual reward and encouragement to keep playing.

TABLE VI. Scoring Scheme, Points are in Thousands

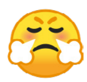
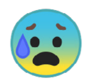
Situation	Points
Completing one row or column	1
Completing one row and column at the same time	4
Puzzle solved	10
For each remaining second once puzzle completed	2

Although the user is asked to fill in a final questionnaire, this only provides us with a form of *overall* opinion, but it was interesting to get more detailed feedback from the user for each of the puzzles, so that the player could express their emotional feelings at a time closer to the moment they experience them. To that end, an Instant Emotional Feedback (IEF) component was designed in the form of emojis (bottom part in Fig. 3). This IEF allows the user to choose, at any time during a puzzle or right before moving to the next puzzle, one or more among 8 possible emojis to express a subset of the emotions most closely resembling their mood.[1]

[1] Notice that this form of IEF might be related to the different approaches of getting feedback by monitoring player's physiological signals [78], [79], which are more costly and obtrusive, but have the potential to predict the players' affective state in real time.

An individual emoji can be selected more than once as a form of indication of the strength of the corresponding emotion. Although these feelings can be roughly categorized into positive and negative emotions (Table VII), we used this just as a pragmatic approximation for usability, since it offers a convenient grouping criterion. For instance, a surprise is not necessarily a positive emotion (it could be a surprise for bad); and the expressions for difficulty can convey a negative valence if a puzzle is felt as *too much* difficult, as this would depart from the ideal flow concept. So, the interpretation of the use of some of these emojis should be taken as a mere approximation of the true feelings.

TABLE VII. Emotions and Emojis for the IEF (Instant Emotional Feedback). Since the Emoji Images Can be Ambiguous in the Associated Feeling, a Short Caption was Added Below

Valence	Underlying emotion, emoji used to convey the emotion, and informal caption			
positive	Entertaining	Surprising	Exciting	Easy
				
	Fun	Ooh!	So cool!	Easy as pie
negative	Boring	Dull	Frustrating	Difficult
				
	Bore	Indifferent	Ughh!	Crappy!

D. Pilot and Final User Study

After testing the game ourselves, four people of different profiles (two of them females; one senior lecturer, one teenager high-school student, and two young students, one undergraduate and one graduate), were asked to play the game to identify potential functionality and usability issues. Each game condition (ST and VC) were assigned to two of them. No problem was identified, and all of them could complete the game. Interestingly, two of the pilot users reported to find the initial puzzle critically important to understand that only one image needed to be formed, and which one it was. The feedback after a row or column is complete was reported to be either useful (to make sure whether a puzzle is fully solved) or satisfying.

For the final study, calls for participation were submitted to mailing lists in our university, both to staff (lecturers and administration) and students, both from different disciplines (humanities, law and economics, health, and sciences & engineering), so different ages and backgrounds were covered. Users were informed on the scientific purpose of the study and that the privacy of the collected data was guaranteed since it would only be treated anonymously.

For the statistical significance analyses, the following tests were used: the non-parametric Mann–Whitney U test (M-W for short), Kruskal-Wallis H-test, χ^2 test, and Z score test, depending on the nature of the data and the comparison purpose. Even though we report p-values, we would like to emphasize the danger of dichotomous thinking associating statistical significance to conclusive evidence (and non-significance to no evidence) [80], that statistical testing may promote. Related to this, and also because our work can be seen as a mixture of confirmatory and exploratory research [81], we complement the significance testing with an estimation-based approach based on confidence intervals and effect sizes.

E. Implementation Details

All parts of the game, including the questionnaire, were implemented using HTML5 and JavaScript, and WebGL for the graphics. Since users did not register for playing and no personal information from them was collected, the logged actions, the results of the different puzzles, and the responses to the questionnaire were associated with randomly generated names. The game condition (ST or VC) was selected uniformly at random, so that approximately half of the participants were assigned to each group.

The game was (and is) made available at https://bit.ly/3dxLFXZ. The link was provided as part of the message for the participation, for convenient quick and ready access. This is the version used during the study, where the game condition is chosen randomly. For the reader convenience, VC and ST versions can be accessed at https://bit.ly/3lRLV8z and https://bit.ly/ 3IyaNfo, respectively.

V. Results and Analysis

Up to 271 users started the game, and 55% of them (148) completed it. According to the random assignment to the control (ST) and experimental (VC) groups, about half of the participants played under each condition (Table VIII). Interestingly, a higher percentage of ST participants compared to VC participants completed the game, which can be seen as a first sign that the VC version of the game might be more challenging, since it may have caused some users to quit at some point before finishing.

TABLE VIII. Number of Participants who Completed the Game

Condition	Started	Completed (%)
ST	126	77 (61.1)
VC	145	71 (49.0)
Total	271	148 (54.6)

We now discuss the players' performance in section A. Then, the instant emotional feedback and the answers to the questionnaire are analyzed in sections B and C, respectively. All these analyses focus on the data from the users who completed the game, to avoid the bias that considering the data from all participants may introduce.

A. Behavioral Results

In order to study how challenging the different puzzles and game conditions were, two play performance metrics are mainly considered: the time taken to complete the game, and the number of clicks (i.e. cube rotations) the users made. Higher values for these metrics can readily be associated to an overall greater difficulty. We analyze these metrics globally in section 1 and per-puzzle in section 2.

B. Global Analysis

On average, VC users took 50% longer and required about 25% more clicks than the ST users for game completion (Table IX). These differences are statistically significant (U = 763.0 for times and U = 1292.5 for clicks, p < 0.0001 both cases), and provide strong evidence on the higher challenge that visual computing introduces. This first finding leads to two relevant questions: which particular visual concepts are offering higher challenge, and how participants subjectively perceive the game. These issues are analyzed in section 2 and section C, respectively.

TABLE IX. Performance Metrics of Game Completion: Mean (Standard Deviation)

Metric	All users	ST	VC
Time (Seconds)	594 (189)	486 (115)	710 (183)
# clicks	429 (106)	384 (68)	479 (117)

Although somehow more anecdotally, 19 users used the "give up" button for particular puzzles. Most of of these users (14) were playing the VC version, which brings further evidence on previous observations. Curiously enough, these 19 users took about 20% longer to complete the game than the remaining 129 users. Individually, any of those players who gave up some puzzle, used this option twice at most. These observations suggest that the give-up option was used sparingly, and only when players sensibly judged they were taking too long to complete some particular puzzle. Most surrenders correspond to puzzle $C_{\widetilde{vc:d}}$ followed by $C_{vc:d}$, a fact that can be further inspected in light of other results discussed in subsequent sections.

C. Per-Puzzle Analysis

The above overall results can be broken down by the different kind of puzzles. The confidence intervals [82], [83] for time and clicks (Fig. 6), computed with bootstrapping (100 repetitions, percentile method [84]), provide both a good first idea of which puzzles require more effort to complete, and qualitative support to the quantitative analysis described subsequently.

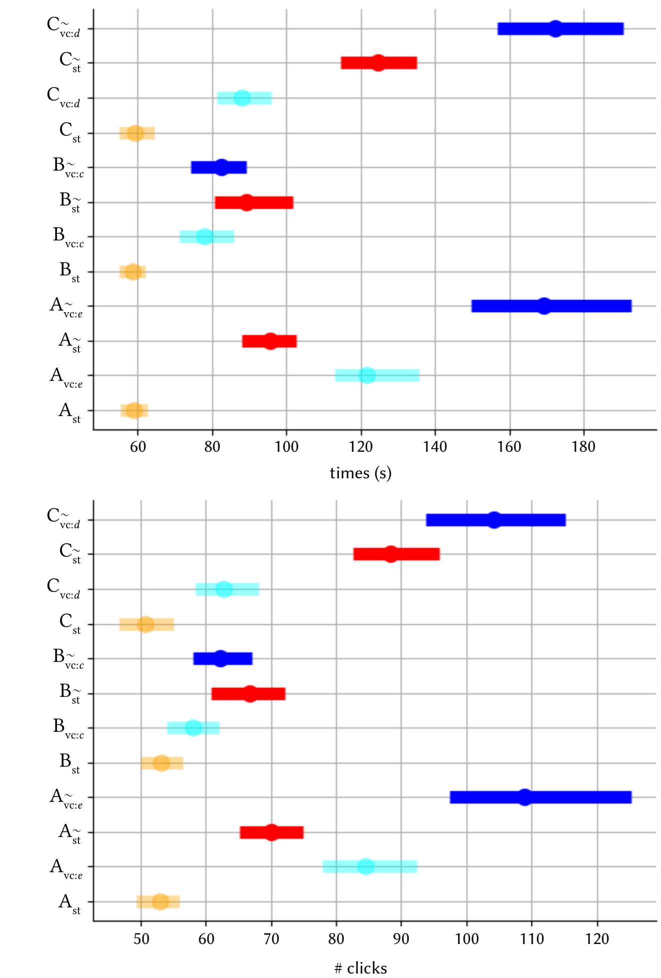


Fig. 6. Times (top) and clicks (bottom) per puzzle, with 95% confidence intervals.

Three comparisons are considered: ST versus VC, `distinct` versus `alike` puzzles, and among puzzle types (namely, with different target images for the same game condition and likeness). These analyses are performed in two complementary manners: effect sizes and hypotheses testing, as reasoned above (Sect. D).

For the effect sizes, Cohen's d [83] was used (Fig. 7(a–d)) where the confidence intervals for the mean were computed with bootstrapping as well. For convenience, the intervals were color-coded according to the known effect size classes (small, medium and large) [83] using the mean as a reference, as follows: large if $|d| \geq 0.8$, medium if $0.5 \leq d < 0.8$, small if $0.2 \leq |d| < 0.5$, and none if $|d| < 0.2$. For the statistical tests, all the comparisons are given graphically (Fig. 8a–c) for higher clarity.

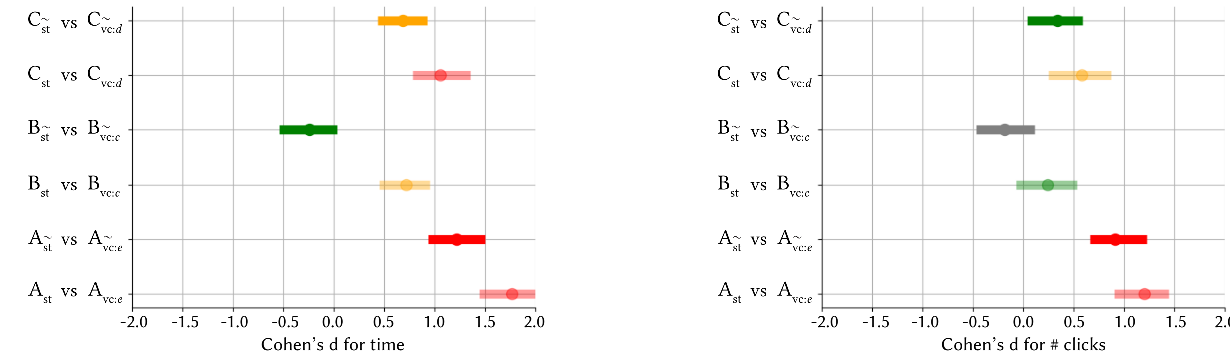
ST versus VC. Players' performance for each of the 6 puzzles plus the initial puzzle (`L`) were pair-wise compared under the ST and VC conditions (Fig. 7a, Fig. 8a). Significant differences were found (Fig. 8a) for both the required completion time and number of clicks, for all puzzles except for `L` and `B~`. These results are consistent with those obtained with Kolmogorov-Smirnov test, with the only difference being in $C_{\tilde{st}}$ vs $C_{\widetilde{vc:d}}$ for times (not for clicks); thus both tests agree in 11 out of 12 comparisons. The statistical differences are also essentially captured with the effect sizes (Fig. 7a). The lack of differences in `L` ($U = 2617.5$, $p = 0.33$ for time, and $U = 2700.0$, $p = 0.45$ for the number of clicks) makes sense because this is just a test puzzle not including any form of visual computing. Thus, this result serves as a verification that no undesirable bias due to differences in subject distribution exists. As for the lack of differences between $B_{\tilde{st}}$ and $B_{\widetilde{vc:c}}$ ($U = 2408.5$, $p = 0.11$ for time, and $U = 2454.0$, $p = 0.14$ for the number of clicks), it might relate to the fact that the alike version in this puzzle (i.e. with distracting images similar to the target image) had a higher impact on the difficulty of $B_{\widetilde{vc:c}}$ than the fact of including the VC. This suggests that the distracting images can be regarded as a form of visual computing itself, as initially hypothesized.

Distinct versus alike versions. Performances were also compared between the distinct and alike versions of the puzzles for both the VC and ST conditions, separately, i.e. A_{st} vs $A_{\tilde{st}}$, $A_{vc:e}$ vs $A_{\widetilde{vc:e}}$ and so on (Fig. 7b, Fig. 8b). With a confidence level $\alpha = 0.05$, significant differences are found (Fig. 8b) in both ST and VC, for the three types of puzzles, except for $B_{vc:c}$ vs $B_{\widetilde{vc:c}}$, a result which agrees with the per-puzzle analysis described above. It can be observed that the medium-big effect sizes (Fig. 7b) correspond to the statistical differences. Notice that the differences found in the ST condition further support the usage of distracting images as a factor for game-difficulty modulation.

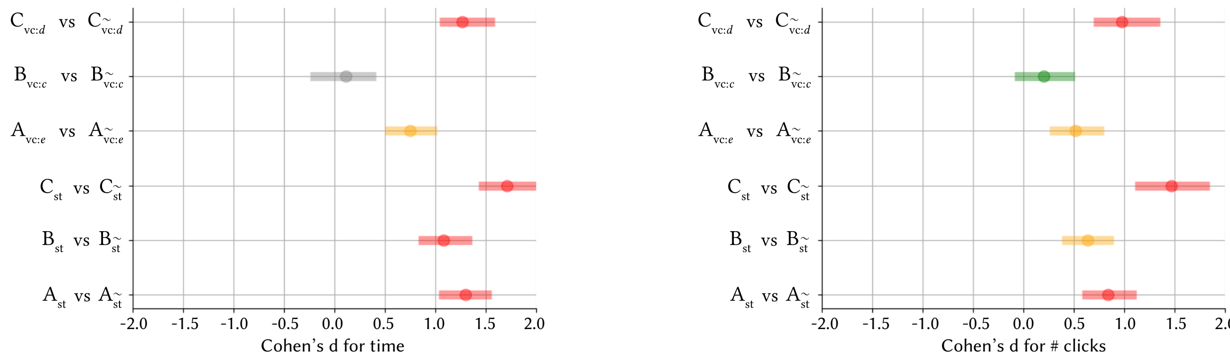
Among puzzle types. Finally, to gain insight on whether some puzzles may be harder than others, performances were compared group-wise for two 3-puzzle groups, namely, the three `distinct` puzzles, and the three `alike` ones, again for VC and ST separately (Fig. 7c–d, Fig. 8c). The differences among the puzzles of the distinct group for the ST condition (A_{st}, B_{st}, C_{st}) seem statistically insignificant, which agrees with Cohen's d close to 0, which means that none of the different images being used (eyes, beaches and smoke) bring any particular challenge with respect to the others. However, when either VC concepts ($A_{vc:e}$, $B_{vc:c}$, $C_{vc:d}$) or alike versions ($A_{\tilde{st}}$, $B_{\tilde{st}}$, $C_{\tilde{st}}$), or both ($A_{\widetilde{vc:e}}$, $B_{\widetilde{vc:c}}$, $C_{\widetilde{vc:d}}$) are introduced, the remarkable differences in both metrics (time and the number of clicks) suggest that the visual computing elements induce differences that affect both the `distinct` and `alike` versions of the puzzles.

Now, to understand which puzzles are harder than others, the three possible pair-wise post-hoc tests are performed for the cases where group differences have been found. From the results (Fig. 8d) the most remarkable outcome is that $C_{\tilde{st}}$ seems the most difficult one among the alike versions under ST. For VC, $A_{vc:e}$ seems the most challenging one among the distinct versions, and $B_{\widetilde{vc:c}}$ the easiest one among the alike versions. Interestingly, both the times and the number of clicks agree in the puzzle ranking. This suggests that both metrics characterize similarly the control and experimental groups. Regarding the effect sizes, the medium to big sizes found in both ST and VC (Fig. 7c-d) are in agreement with the statistical results. Effect sizes also reveal a close symmetry between times and clicks.

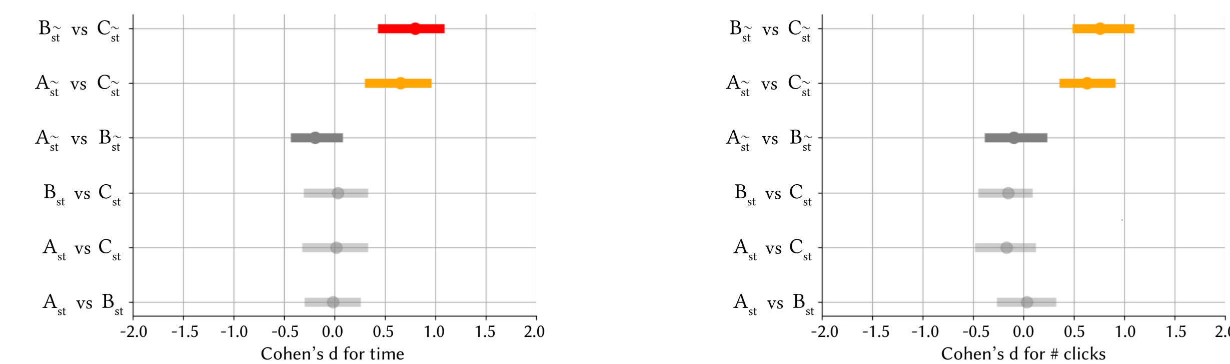
Taken together, these observations can be summarized as follows. First, the game difficulty can be modulated not only by visual computing but also by a proper choice of the target and distracting images. Second, interestingly, the VC elements bring an additional difficulty beyond image contents *even* with the `distinct` versions of the puzzles, which provides an evidence of its intended effectiveness. Finally, remarkable differences exist among the three types of VC elements introduced, which promises to offer flexibility when choosing the desired challenge level for a target game and user profile.



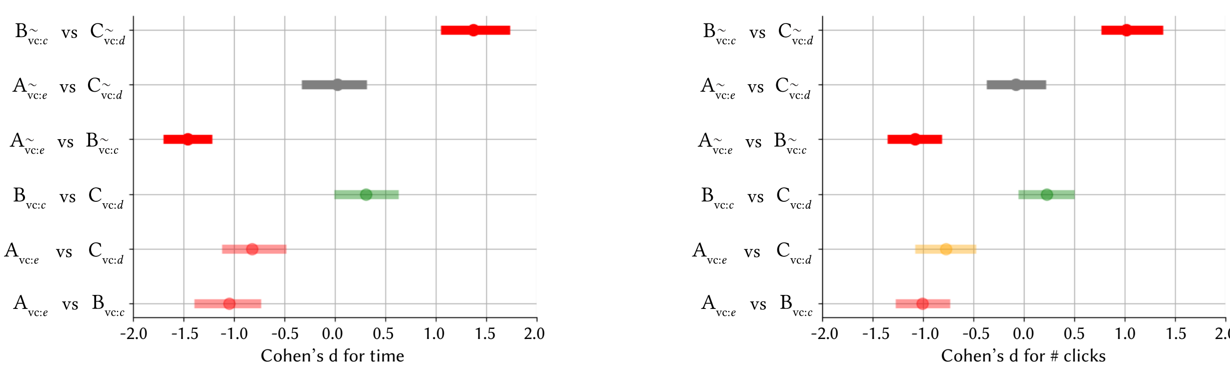
(a) ST vs VC



(b) distinct vs alike



(c) among puzzle types in ST

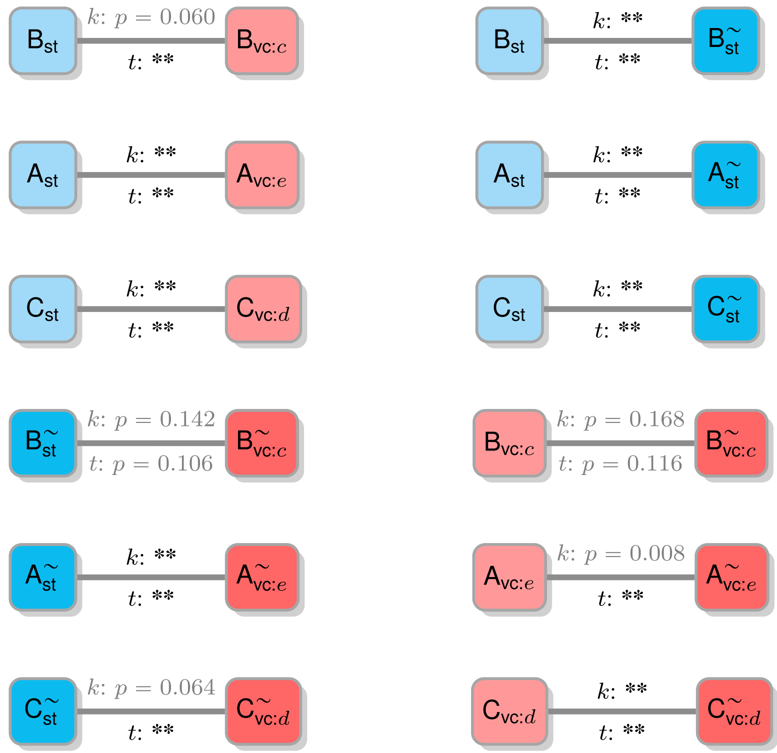


(d) among puzzle types in VC

Legend for interval's color:

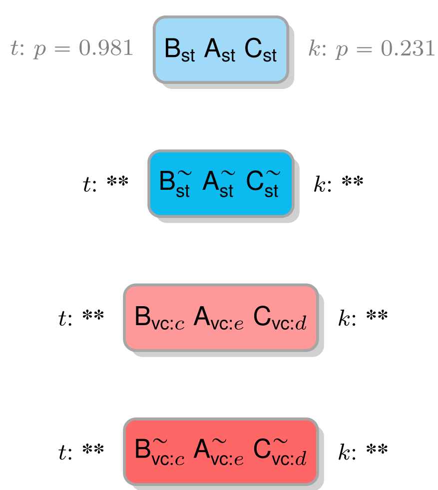
Effect size		distinct vs alike (in a, c, d)
red	big	light distinct
yellow	medium	**dark** alike
green	small	
gray	none	

Fig. 7. Confidence intervals of effect sizes (cohen's d) for times (left) and clicks (right) when comparing different puzzles.



(a) Pairwise ST vs VC

(b) Pairwise distinct vs alike



(c) Group-wise within distinct and alike

(i) ST

Group	Rank	
	Time	# clicks
distinct	—	—
alike	$A^{\sim}_{st} < C^{\sim}_{st}$ $B^{\sim}_{st} < C^{\sim}_{st}$	$A^{\sim}_{st} < C^{\sim}_{st}$ $B^{\sim}_{st} < C^{\sim}_{st}$

(ii) VC

Group	Rank	
	Time	# clicks
distinct	$B_{vc:c} < A_{vc:e}$ $C_{vc:d} < A_{vc:e}$	$B_{vc:c} < A_{vc:e}$ $C_{vc:d} < A_{vc:e}$
alike	$B^{\sim}_{vc:c} < A^{\sim}_{vc:e}$ $B^{\sim}_{vc:c} < C^{\sim}_{vc:d}$	$B^{\sim}_{vc:c} < A^{\sim}_{vc:e}$ $B^{\sim}_{vc:c} < C^{\sim}_{vc:d}$

(d) Post-hoc test after (c)

Fig. 8. Diagrams of the statistical tests performed on comparing types of puzzles (a, b) pairwise (Mann-Whitney rank test) and (c) group-wise (Kruskal-Wallis H-test), for completion time (t) and number of clicks (k). Bonferroni correction is applied with $\alpha^* = 0.05/m$, for $m = 48$ accounting for the all pair-wise hypotheses (24) and the two dependent variables (t and k). Note: ** means p-value is very low (lower than 0.0001), thus representing statistically significant differences (d) Ranking puzzles from pair-wise post-hoc tests (Mann-Whitney rank test). Note: the notation X < Y represents that significant differences exists between X and Y and puzzle X has lower mean for the corresponding metric than puzzle Y.

D. Instant Emotional Feedback

Beyond the performance metrics, we are also interested in finding which additional insights can be gained from the qualitative instant emotional feedback (IEF). Around 90% of the users who completed the game used some emoji to provide their reaction to particular puzzles (Table X), in similar amounts for users in ST and VC groups, just slightly higher for the latter. We analyze which emojis were used globally in section 1, and per-puzzle in section 2.

TABLE X. Usage of the Instant Emotional Feedback (Emojis)

Condition	Users (%)	# Emoji	# Emojis/user (avg.)
ST	88.3	478	6.2
VC	92.9	517	7.3
Total	89.9	995	6.7

E. Global Analysis

It was found that the VC group completed the game with more time and clicks, which we can associate to a generally higher difficulty of the game under the VC condition. Now, based on the emojis chosen by the users (Table XI), it can be observed that emotion easy is used more than hard in the ST group, and the opposite happens in the VC group, which reinforces such association. Besides this increment in the perceived difficulty in the VC group, emotion `entertaining`, although decreasing, maintains the first place, while the rest of positive emotions (`surprising` and `exciting`) increase. In other words, by considering the group of the three emotions (`entertaining`, `surprising`, and `exciting`) as a whole, they are used similarly in ST (56.9%) and VC (53.4%), which can be interpreted that, overall, emotions with positive valence are similarly expressed in both conditions, which support our hypotheses regarding the game modulation and diversification through VC. Similarly, although hard is used more in VC than in ST, other emotions with negative valence (dull, frustrating and boring), not only are used the least in both conditions, but also with the same percentage (17.8% in total).

TABLE XI. Percentage of Emotions Usage (With Respect to the Total Emotions Used Group-Wise) in Both Groups, Sorted in Descending Order

ST		VC	
Emotion	%	Emotion	%
entertaining	31.8	entertaining	21.5
surprising	14.4	hard	19.1
easy	14.0	surprising	17.0
hard	11.3	exciting	14.9
exciting	10.7	easy	9.7
dull	10.0	dull	6.6
frustrating	6.1	frustrating	6.4
boring	1.7	boring	4.8

When comparing the usage proportion of each emotion (Table XII), it turns out that ST users found the game more entertaining, while VC users found it not only harder (which brings further statistical support to previous analysis) but also more boring and more exciting at the same time. Although this result clearly indicates that a per-puzzle analysis is called for, it is also worth noting that the result for `boring` is possibly less relevant, since this emotion is used little (less than 5%). Also, note that after Bonferroni correction, only the differences regarding `alike` are found significant.

TABLE XII. Emotions Whose Usage Proportions are Found Statistically Different (Z score, A = 0.05). The Symbol + Represents in Which Group (ST Or VC) the Corresponding Emotion is More Used, Whose % of Usage is Also Given

ST	VC	% usage	p-value	z
+ entertaining		31.8	0.046	1.997
	+ **hard**	19.1	**	-4.517
	+ exciting	14.9	0.011	-2.528
	+ boring	4.8	0.016	-2.404

Note: ** means p-value is very low (lower than 0.00001). After Bonferroni correction with $\alpha^* = \alpha/m$, for the m = 8 emotions tested, the difference in proportion is significantly different only for hard, which is boldfaced.

F. Per-puzzle Analysis

When the emoji usage is analyzed pair-wise (ST vs VC) per puzzle (Table XIII), it is found that significantly more emojis were used in VC than in ST for all puzzles: up to 9 emotions are found to be significantly more used in VC than in ST, while only 2 emotions were more used in ST, which additionally occurs in just two puzzles (A and A˜)[2]. This seems to imply that the VC condition generally rouses more emotions in a wider range of conditions. Another interesting observation is that although entertaining was found to be predominant in ST in a global sense (Table XII), its significance actually only occurs in a single puzzle ($A_{\tilde{st}}$), which suggests that it cannot be generalized the fact of ST being more entertaining than VC. A final remark is that, as noted above, no significant difference was found in the time required to solve B˜ between ST and VC (Fig. 8a). However, according to the IEF, B˜ is found easier in VC, which not only further supports the idea that distracting images produces higher difficulty, but also that VC modulates this difficulty and, very interestingly, by lowering it in this particular case. Notice that this analysis changes after Bonferroni correction.

TABLE XIII. Emotions Used Significantly More (Z score, α = 0.05) Between the ST and the VC Versions of Each Puzzle. After Bonferroni Correction With $\alpha^* = \alpha/M$, for the M = 48 (6 Puzzles and 8 Emotions), the Difference in Proportion is Significantly Different Only for the Emotions in A, Which are Boldfaced

	A	z	p-value	A˜	z	p-value
ST	+ **easy**	+3.319	0.001	+ entertaining	+2.458	0.0140
VC	+ **hard**	-3.616	0.0003	+ hard	-2.619	0.0088
	+ **frustating**	-3.436	0.001			
	B			B˜		
VC	+surprinsing	-2.156	0.0311	+ easy	-2.000	0.0456
	C			C˜		
VC	+ hard	-2.438	0.0148	+ hard	-2.155	0.0312
	+ exciting	-2.000	0.0456	+ exciting	-3.000	0.0026

G. Opinion Questionnaire

Finally, the participants opinions are analyzed mainly to find out how each game condition (ST or VC) was perceived (Sect. 1) and then to compare subjective perceptions with actual performance metrics (Sect. 2).

1. Participants' Perception

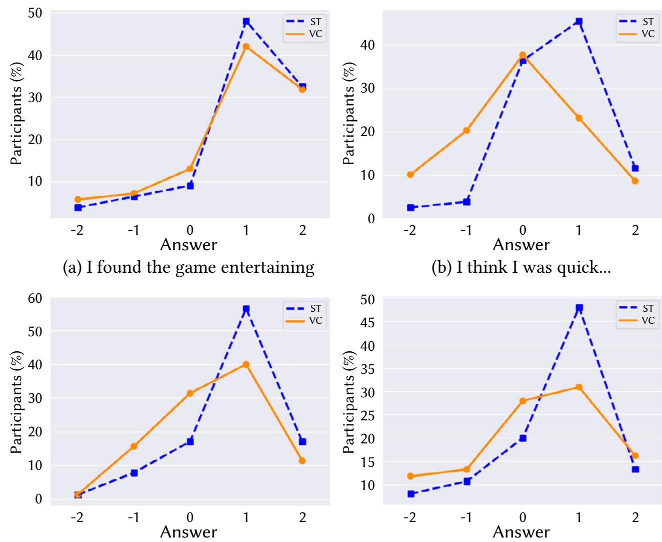
The responses of the questionnaire (Figs. 9–11) were compared pair-wise (ST vs VC) per question. No statistically significant differences were found in most cases. A general qualitative observation of the responses (Fig. 9) reveals that most users found the game entertaining (Fig. 9a), and easy to complete (Fig. 9c), and they agreed that they would play the game again (Fig. 9d). Most of the users in the ST group also agreed they were quick at solving the puzzles (Fig. 9b).

2. Subjective Perception Vs Actual Effort

Regarding the overall experience (Fig. 10), both groups mostly found it entertaining, compared to boring (Fig. 10a), and surprising, compared to dull (Fig. 10b), but neither exciting nor frustrating (Fig. 10c). However, ST users found it mostly simple, while VC users did not (Fig. 10d). This difference is statistically significant ($\chi 2$ = 13.934, p = 0.001) and reinforces the conclusions obtained in the previous sections. Therefore, this can be seen as that users' subjective perception matches objective data. Interestingly, this increase in difficulty does not generally seem to translate into less entertainment or more frustration.

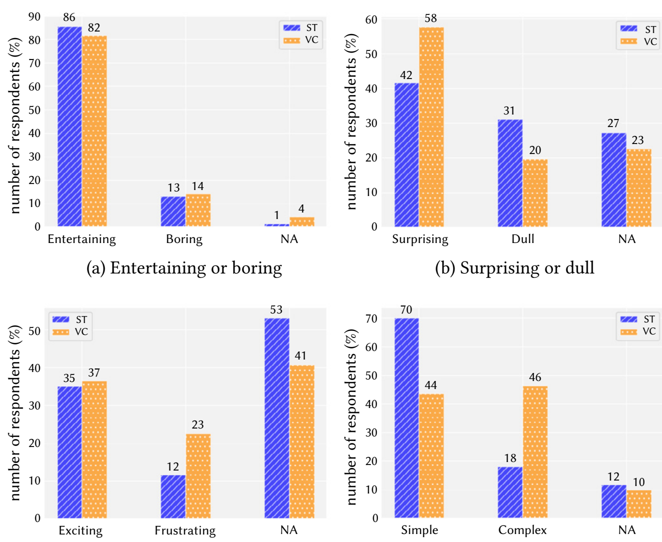
The results of the most and the least preferred puzzles (Fig. 11) are also similar in both groups, despite the actual differences observed regarding actual effort. It is worth noticing how, regardless of the group (ST or VC), puzzles B˜ and C˜ are liked twice and three times

[2] We remind the reader that these are not necessarily the first puzzles presented to the player, since the order is chosen randomly for each participant, as discussed in A.



(a) I found the game entertaining
(b) I think I was quick...
(c) I found easy to complete...
(d) I would play the game again

Fig. 9. Distribution of replies to Questions 1–4 of the questionnaire. Please, note that a numerical scale has been used for the agreement scale, as given below, for convenience.



(a) Entertaining or boring
(b) Surprising or dull
(c) Exciting or frustrating
(d) Simple or complex

Fig. 10. Distribution of replies regarding the overall game experience. We include the percentage of users that did not select (NA) any of the adjectives of each pair.

as much as their B and C counterparts, respectively, even though they include the additional challenge of distracting images. It is interesting to note that a similar proportion of VC and ST users liked $C^{\sim}$ the most, in spite of $C^{\sim}_{vc:d}$ being particularly hard. The puzzle selected as the least liked was A in both ST and VC. When comparing the most and least liked puzzle-wise, it can be observed that some puzzles are judged as the least liked as much ($B^{\sim}_{vc:c}$), half as much ($C^{\sim}_{vc:d}$), or three times as much ($A^{\sim}_{vc:e}$) as they are preferred. These observations lead to two main conclusions: players exhibited a wide variety of skills and tastes, and a notable amount of them seems to have enjoyed some of the most challenging puzzles.

It is important to find out whether users' perception of effort and difficulty align with actual collected metrics. Two questions in the questionnaire (namely, whether they think they were quick, and they found the game easy) relate to the two metrics collected (completion time and number of clicks). Then, it is possible to explore how the

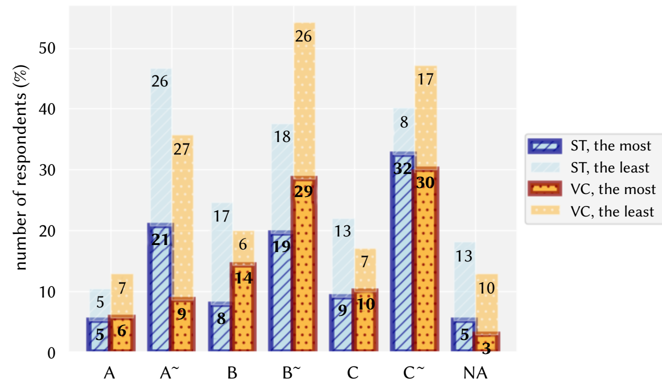
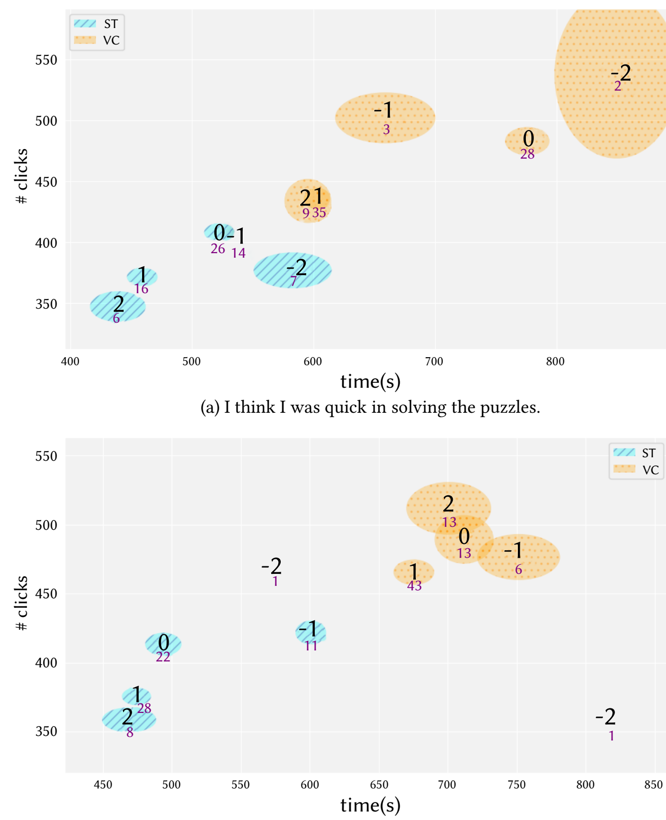


Fig. 11. Which puzzle participants liked the most and the least. Both groups, ST and VC, liked the puzzle $C^{\sim}$ the most, and both liked the puzzle $A^{\sim}$ the least. For some puzzles, there are noticeable differences worth studying (see text for details).



(a) I think I was quick in solving the puzzles.

(b) I found easy to complete the game.

Fig. 12. Means and standard errors (represented as the lengths of the corresponding ellipses' axes) of completion time and number of clicks for VC and ST users, grouped by their Likert replies to whether they (a) think they were quick, and (b) found the game easy to complete. For quicker reading of the plots, the agreement scale is coded as positive (for agreement), negative (for disagreement) and 0 (for neutral), which is the number in the center of the ellipses. The number below this agreement scale is the number of subjects for each particular response.

answers to these questions distribute in the time-clicks space. Regarding the effort, it can be observed that users' perception strongly correlates with the actual times (Fig. 12a). This happens in both user groups, which is interesting because the time and click ranges are significantly different among these groups. In other words, VC users who felt they were quick, took shorter, on average, than VC users who believed they were not, but longer than ST users in general, even than those ST users who felt they were not quick. It can also be noticed that time correlates better with the user perception of time than the number of clicks.

As for the second question, the perception of difficulty correlates generally better with the number of clicks than with completion times (Fig. 12b). In this case, however, the results are notably different for the VC and ST users. On the one hand, ST users who perceived the puzzles as difficult took generally longer and used more clicks. On the other hand, although the pattern for VC users is possibly less clear, it is intriguing that those VC users who used the most number of clicks, strongly agreed that the game was easy. It can be very tentatively argued that this may partly relate with a state of flow: those users might have been enjoying the game *despite* making more mouse clicks, since for their skills the challenge was at a good difficulty balance. On average, this user subgroup took also longer than those who less strongly claimed to find the game easy.

VI. Discussion

The overall results of this study strongly support the idea that a variety of simple visual ingredients can be used to modulate the game difficulty while preserving the core gameplay. Interestingly, the analysis of user perception through the instant emotional feedback and opinion survey reveals that more challenging puzzles do not necessarily lead to boredom or frustration, but they may also be found entertaining and enjoyable. Although the evidence for claiming that a flow state was achieved by some participants is certainly weak, and not directly pursued at design stage, a door is certainly open regarding this desirable possibility.

Regarding the applicability of this study, puzzles and card games seem the most straightforward choice, which is not little considering that a wide range of these games exist, either as complete games or as mini games. Beyond these genres, however, other possibilities where some variation of this work is feasible includes games whose backgrounds or other visual elements (enemies, non-playable characters) can be modified to modulate the gameplay. It can even be speculated that the concept may inspire not only variations of existing games, but also new games.

The preferences for some puzzles observed in the questionnaire may suggest that solving these puzzles are intrinsically more enjoyable. However, an alternative hypothesis is that these preferences might have to do with the particular images being used. Further work might explore the use of the same images across different visual concepts to remove this potential bias, and making sure that the questions are understood as meant. Regarding the different visual computing elements considered, they produce a variety of effects and it is therefore hard to generalize how exactly each contributes to modulate the difficulty. All in all, however, the combination of abstract shapes and the dynamic effect (rotational motion) led to more participants to surrender, and therefore this visual effect should be used only when the highest challenge is advisable. The color-based modification has possibly contributed to lower the difficulty to the extent that using distracting images neutralized its effect. This suggests that the same visual concept may have a notably different impact depending on which images it is applied to. The edge-like effect was possibly the least popular and among the hardest ones, both with similar and dissimilar distracting images. Overall, the challenge that a particular visual element will induce has to be carefully studied, and it is possible hard to predict. One possibility to address this is relying on the body of knowledge of the human cognition and visual perception, to leverage on its strengths and weaknesses. Another possibility would be to automatically learn the implications of each of a large set of visual effects through large-scale long-term usage.

User feedback is provided when a row or column of cubes are correctly completed. Indirectly, this feedback acts as a facilitating mechanism and, while this is applied to all puzzles, it makes the hardest puzzles easier, which complicates the derivation of accurate conclusions on how hard the visual concept itself is, since the feedback may be interfering. In addition, this feedback might also promote some players to apply some form of strategic playing. Therefore, these issues would require further examination and experimentation.

It is interesting to note that the number of clicks and the elapsed time to complete individual puzzles correlate well with the particular game version that users played. This implies that these metrics may have some potential predictive power of how hard players are finding a game or part of it. However, the time distributions (Fig. 13) reveal two important observations. First, despite having statistically different means, there is a notable distribution overlap; second, the time distribution for VC players is right long-tailed (in fact, this distribution is not normal whereas that of ST is). These observations mean that many VC users actually took shorter than the average VC user, with times close to ST users, and that some VC others took significantly longer. Therefore, although VC is *generally* and intrinsically harder, user skill is also a critical factor in explaining the observed performance.

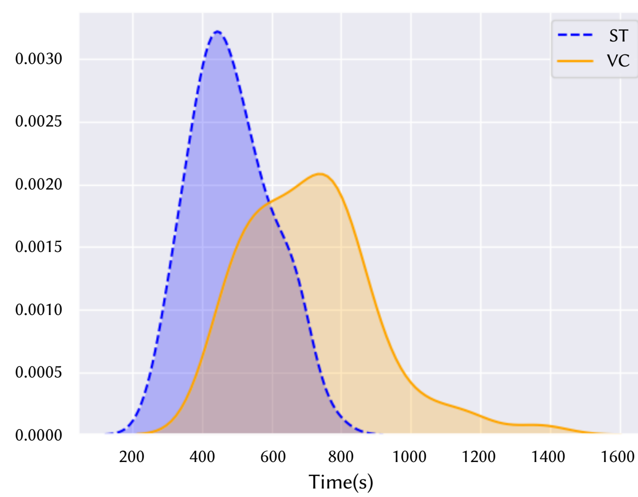


Fig. 13. Density estimation of time for game completions in VC and ST. The distributions for the number of clicks are similar.

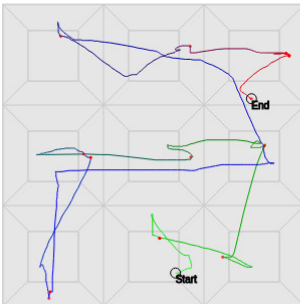
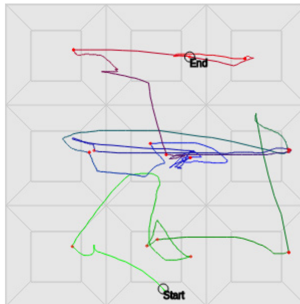
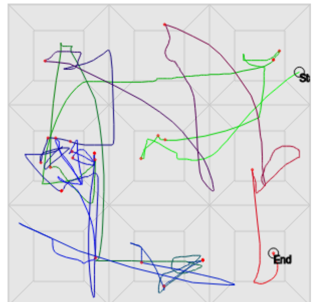
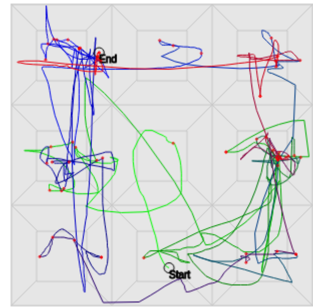
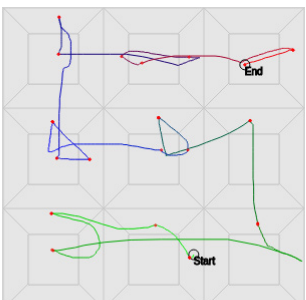
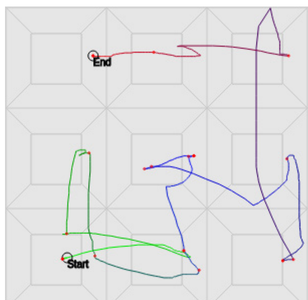
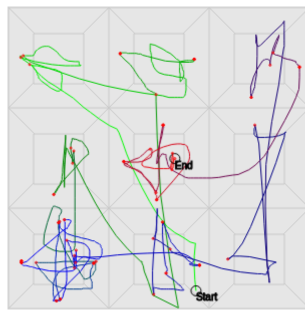
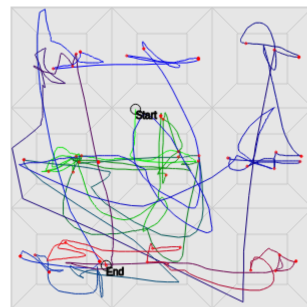
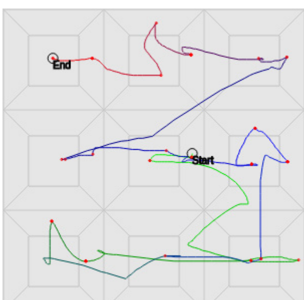
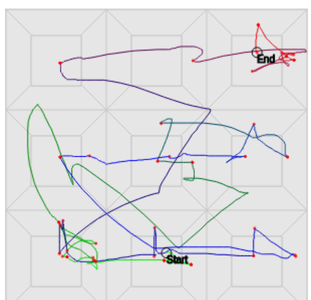
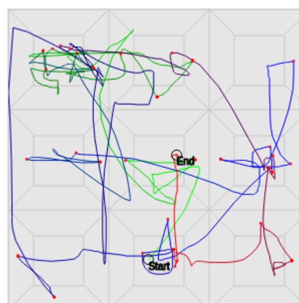
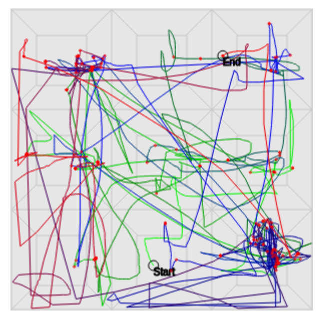
The considerations above suggest that additional data, besides times and clicks, can be useful for player modeling and predicting their skills. For instance, mouse traces (Table XIV) could bring an additional metric of effort (e.g. length of mouse trajectories) and, more generally, they can be very insightful into the different solution strategies employed by users, or for guessing which sense of control over the game a particular user is having.

Beyond manual inspection of these visualizations, or deriving heuristic metrics, disciplined machine learning techniques can turn out to be very useful for characterizing players and, in turn, dynamic game adaptation. In fact, a first approach to this problem has recently been explored [85].

VII. Conclusions

This work has empirically shown that simple visual computing techniques can be introduced to modulate the play experience *without* any other change in the gameplay. In particular, the conducted between-subjects user study revealed significant performance differences between users who played the visually-modified game and those who played a standard, unmodified game version. Interestingly,

TABLE XIV. Mouse Traces and Clicks for Each of the Six Puzzles as Solved by Two Participants, One Under ST and Another Under VC Who, Respectively, Took the Least and the Most to Complete the Game. For Each Puzzle, the Start and end Cursor Points are Labeled, and the Percent of Elapsed Time is Represented With the Trace Color (From Green to Blue to Purple to Red). The Nine Pieces (Cubes) and the Five Areas Per Cube Used to Rotate it are Illustrated in the Background as a Visualization Guide. This Representation Provides an Idea of the Effort Devoted to a Particular Puzzle by a Particular User, and it Lends Itself to Per-Puzzle, Per-User, and Per-Condition Analyses

	User under ST		User under VC	
	distinct	alike	distinct	alike
A				
B				
C				

the user feedback in terms of in-game emojis and a final questionnaire, suggest that the more challenging parts of the games are not necessarily always found frustrating, but can also be found enjoyable by some users.

The work has used three visual concepts (edge, color mapping, and rotating motion), plus distracting images looking similar to the target image, on a cube puzzle web game prototype. Further work may explore other visual computations and alternative games to keep exploring this research theme. The findings of this first study can inform the design of games where these kind of visual modifications can be easily introduced, as well as guide further related research. We believe that through data obtained from players' interaction, machine learning techniques can be leveraged to model users and, eventually, be able to adapt dynamically the game to match each individual's skills. In the long term, this would lead to more enjoyable game experiences.

Acknowledgment

We are very grateful to all participants in the study. The contribution of Francisco Andújar De Andrés in the implementation of the mouse traces is acknowledged.

References

[1] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, J. N. Pidruzny, "The development of the game engagement questionnaire: A measure of engagement in video game-playing," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 624–634, 2009.

[2] E. Knekta, C. Runyon, S. Eddy, "One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research," *CBE life sciences education*, vol. 18, no. 1, 2019.

[3] D. Johnson, M. J. Gardner, R. Perry, "Validation of two game experience scales: The player experience of need satisfaction (PENS) and game experience questionnaire (GEQ)," *International Journal of Human-Computer Studies*, vol. 118, pp. 38–46, 2018.

[4] A. Denisova, C. Guckelsberger, D. Zendle, "Challenge in digital games: Towards developing a measurement tool," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, 2017, pp. 2511–2519.

[5] A. Denisova, P. Cairns, C. Guckelsberger, D. Zendle, "Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS)," *International Journal of Human-Computer Studies*, vol. 137, 2020.

[6] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience.* Harper Perennial, 1991.

[7] B. Cowley, D. Charles, M. Black, R. Hickey, "Toward an understanding

of flow in video games," *Computers in Entertainment*, vol. 6, July 2008.

[8] K. Jalife, C. Holmgård, "Cognitive components of flow states and applications to video game design: A brief framework," in *Proceedings of the 2019 ACM SIGCHI Conference on Creativity and Cognition, C&C 2019, San Diego, CA, USA, June 23-26, 2019*, 2019, pp. 570–577, ACM.

[9] P. Sweetser, P. Wyeth, "Gameflow: A model for evaluating player enjoyment in games," *Computers in Entertainment*, vol. 3, July 2005.

[10] K. Procci, A. R. Singer, K. R. Levy, C. Bowers, "Measuring the flow experience of gamers: An evaluation of the DFS-2," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2306– 2312, 2012.

[11] C. Shu-Hui, W. Wann-Yih, J. Dennison, "Validation of EGameFlow: A self-report scale for measuring user experience in video game play," *Computers in Entertainment*, vol. 16, Sept. 2018.

[12] P. Sweetser, D. Johnson, P. Wyeth, A. Anwar, Y. Meng, A. Ozdowska, "Gameflow in different game genres and platforms," *Computers in Entertainment*, vol. 15, Apr. 2017.

[13] M.-Y. Hwang, J.-C. Hong, Y. wei Hao, J.-T. Jong, "Elders' usability, dependability, and flow experiences on embodied interactive video games," *Educational Gerontology*, vol. 37, no. 8, pp. 715–731, 2011.

[14] M.-J. Tsai, L.-J. Huang, H.-T. Hou, C.-Y. Hsu, G.-L. Chiou, "Visual behavior, flow and achievement in game-based learning," *Computers & Education*, vol. 98, pp. 115–129, 2016.

[15] Y.-H. Hsieh, Y.-C. Lin, H.-T. Hou, "Exploring the role of flow experience, learning performance and potential behavior clusters in elementary students' game-based learning," *Interactive Learning Environments*, vol. 24, no. 1, pp. 178– 193, 2016.

[16] Y.-C. Yeh, S.-Y. Chen, E. M. Rega, C.-S. Lin, "Mindful learning experience facilitates mastery experience through heightened flow and self-efficacy in game-based creativity learning," *Frontiers in Psychology*, vol. 10, 2019.

[17] M. T. F. Mak, M. Wang, K. W. S. Chu, "Effects of a gamified learning platform on elementary school students' flow experiences in leisure reading," *Proceedings of the Association for Information Science and Technology*, vol. 56, no. 1, pp. 454–458, 2019.

[18] M. Zheng, H. A. Spires, *Fifth Graders' Flow Experience in a Digital Game-Based Science Learning Environment*, pp. 1433–1450. I. Management Association, 2015.

[19] X. Fang, S. Chan, J. Brzezinski, C. Nair, "Development of an instrument to measure enjoyment of computer game play," *International Journal of Human–Computer Interaction*, vol. 26, no. 9, pp. 868–886, 2010.

[20] A. Corcos, "Being enjoyably challenged is the key to an enjoyable gaming experience: an experimental approach in a first-person shooter game," *Socioaffective Neuroscience & Psychology*, vol. 8, no. 1, 2018.

[21] A. Touati, Y. Baek, "What leads to player's enjoyment and achievement in a mobile learning game?," *Journal of Educational Computing Research*, vol. 56, no. 3, pp. 344– 368, 2018.

[22] Y. Baek, A. Touati, "Exploring how individual traits influence enjoyment in a mobile learning game," *Computers in Human Behavior*, vol. 69, pp. 347–357, 2017.

[23] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, 2008.

[24] J. Chen, "Flow in games (and everything else)," *Communications of the ACM*, vol. 50, pp. 31–34, Apr. 2007.

[25] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*. Prentice Hall, 2018. 4th. edition.

[26] W. Thompson, R. Fleming, S. Creem-Regehr, J. Kelly Stefanucci, *Visual Perception from a Computer Graphics Perspective*. A K Peters/CRC Press, 2011.

[27] A. M. Barry, *Visual intelligence: perception, image, and manipulation in visual communication*. Albany: State University of New York Press, 1997.

[28] M. S. El-Nasr, S. Yan, "Visual attention in 3D video games," in *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 2006, p. 22–es, ACM.

[29] R. Dubey, P. Agrawal, D. Pathak, T. Griffiths, A. A. Efros, "Investigating human priors for playing video games," in *International Conference on Machine Learning (ICML)*, July 2018, pp. 1348–1356.

[30] C. Green, D. Bavelier, "Action-video-game experience alters the spatial resolution of vision," *Psychological Science*, vol. 18, no. 1, pp. 88–94, 2007.

[31] Y.-H. Kim, D.-W. Kang, D. Kim, H.-J. Kim, Y. Sasaki, T. Watanabe, "Real-time strategy video game experience and visual perceptual learning," *Journal of Neuroscience*, vol. 35, pp. 10485–10492, July 2015.

[32] W. Boot, D. Blakely, D. Simons, "Do action video games improve perception and cognition?," *Frontiers in Psychology*, vol. 2, no. SEP, 2011.

[33] R. L. Achtman, C. S. Green, D. Bavelier, "Video games as a tool to train visual skills.," *Restorative neurology and neuroscience*, vol. 26 4-5, pp. 435–446, 2008.

[34] I. Spence, J. Feng, "Video games and spatial cognition," *Review of General Psychology*, vol. 14, no. 2, pp. 92–104, 2010.

[35] D. Kersten, "Perception, computer graphics, and video games," *Journal of Vision*, vol. 7, no. 15, 2007.

[36] D. Dziedzic, W. Włodarczyk, "Approaches to measuring the difficulty of games in dynamic difficulty adjustment systems," *International Journal of Human–Computer Interaction*, vol. 34, no. 8, pp. 707–715, 2018.

[37] O. Sotamaa, "When the game is not enough: Motivations and practices among computer game modding culture," *Games and Culture*, vol. 5, no. 3, pp. 239–255, 2010.

[38] J. Schell, *The Art of Game Design: A Book of Lenses*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2008.

[39] "Mirror's Edge." https://en.wikipedia.org/wiki/ Mirror's_Edge.

[40] J. Juul, *A Casual Revolution: Reinventing Video Games and Their Players*. The MIT Press, 2009.

[41] K. Hicks, P. Dickinson, J. Holopainen, K. Gerling, "Good game feel: An empirically grounded framework for juicy design," in *Proceedings of the 2018 DiGRA International Conference: The Game is the Message*, July 2018, DiGRA.

[42] E. Buckthal, "Juiciness in citizen science computer games: Analysis of a prototypical game," Master's thesis, 2014.

[43] J. Juul, B. Jason, "Good feedback for bad players? A preliminary Study of 'juicy' Interface feedback," in *First joint FDG/DiGRA Conference*, 2016.

[44] J. S. Prook, D. P. Janssen, S. Gualeni, "The negative effects of praise and flattery," in *Foundations of Digital Games*, 2015.

[45] J. Korinek, S. L. Fagerli, "Juiciness — a study of visual effects in games," Master's thesis, Aalborg University, 2017.

[46] K. Hicks, K. Gerling, G. Richardson, T. Pike, O. Burman, P. Dickinson, "Understanding the effects of gamification and juiciness on players," in *IEEE Conference on Games (CoG)*, 2019.

[47] K. Hicks, K. Gerling, P. Dickinson, V. Vanden Abeele, "Juicy game design: Understanding the impact of visual embellishments on player experience," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '19, 2019, pp. 185–197.

[48] D. Kao, "The effects of juiciness in an action RPG," *Entertainment Computing*, vol. 34, 2020.

[49] F. James, "Crash 4 n. verted mode explained: How it works and what to expect with the inverted version of levels," Oct. 2020. https://www.gamesradar.com/crash-4-nverted-mode.

[50] "Mirror mode (Mario Kart Racing Wiki)." https:// mariokart.fandom.com/wiki/Mirror_Mode.

[51] D. Kao, D. F. Harrell, "Toward understanding the impact of visual themes and embellishment on performance engagement, and self-efficacy in educational games," in *American Educational Research Association Conference*, 2017.

[52] D. Kao, D. F. Harrell, "Exploring the impact of avatar color on game experience in educational games," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, 2016, pp. 1896– 1905.

[53] "Colorblind mode (Battlefield wiki)." https:// battlefield.fandom.com/wiki/Colorblind_Mode.

[54] A. Rodríguez-Benítez, I. Boada, S. Thió-Henestrosa, M. Sbert, "Cprforblind: A video game to introduce cardiopulmonary resuscitation protocol to blind people," *British Journal of Educational Technologies*, vol. 49, no. 4, pp. 636–645, 2018.

[55] M. Magdics, C. Sauvaget, R. J. García, M. Sbert, "Post-processing NPR effects for video games," in *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '13, 2013, pp. 147–156, Association for Computing Machinery.

[56] "Non-photorealistic rendering." https://en.wikipedia.org/wiki/Non-photorealistic_rendering.

[57] D. Gravina, A. Khalifa, A. Liapis, J. Togelius, G. N. Yannakakis, "Procedural

content generation through quality diversity," in *IEEE Conference on Games (CoG)*, 2019, pp. 1– 8.

[58] M. Zohaib, "Dynamic Difficulty Adjustment (DDA) in computer games: A review," *Advances in Human-Computer Interaction*, vol. 2018, 2018.

[59] T. Constant, G. Levieux, "Dynamic difficulty adjustment impact on players' confidence," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2019, ACM.

[60] A. Denisova, P. Cairns, "Player experience and deceptive expectations of difficulty adaptation in digital games," *Entertainment Computing*, vol. 29, pp. 56–68, 2019.

[61] T. A. Sanocki, K. W. Bowyer, M. Heath, S. Sarkar, "Are edges sufficient for object recognition?," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 24, pp. 340–349, Jan. 1998.

[62] M. A. Webster, K. Halen, A. J. Meyers, P. Winkler, J. S. Werner, "Colour appearance and compensation in the near periphery," *Proceedings of the Royal Society B: Biological Sciences*, vol. 277, Feb. 2010.

[63] D. K. M. G. Bloj, A. C. Hurlbert, "Perception of three-dimensional shape influences colour perception through mutual illumination," *Nature*, vol. 402, pp. 877–879, 1999.

[64] L. Teichmann, T. Grootswagers, T. A. Carlson, A. N. Rich, "Seeing versus knowing: The temporal dynamics of real and implied colour processing in the human brain," *NeuroImage*, vol. 200, pp. 373–381, 2019.

[65] J. Pan, G. Bingham, "With an eye to low vision: Optic flow enables perception despite image blur," *Optometry and Vision Science*, vol. 90, pp. 1119–1127, Oct. 2013.

[66] P. Lewis, R. Rosén, P. Unsbo, J. Gustafsson, "Resolution of static and dynamic stimuli in the peripheral visual field," *Vision Research*, vol. 51, no. 16, pp. 1829–1834, 2011.

[67] J. Maarseveen, C. L. Paffen, F. A. Verstraten, H. Hogendoorn, "Representing dynamic stimulus information during occlusion," *Vision Research*, vol. 138, pp. 40–49, 2017.

[68] F. Ghayour, D. Cantor, *Real-Time 3D Graphics with WebGL 2: Build Interactive 3D Applications with JavaScript and WebGL 2 (OpenGL ES 3.0), 2nd Edition.* Packt Publishing, 2018.

[69] S. K. Bhatia, V. Lakshminarayanan, A. Samal, G. V. Welland, "Human face perception in degraded images," *Journal of Visual Communication and Image Representation*, vol. 6, no. 3, pp. 280–295, 1995.

[70] M. A. Webster, D. I. A. MacLeod, "Visual adaptation and face perception," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, June 2011.

[71] D. P. Katharina Dobs, Leyla Isik, N. Kanwisher, "How face perception unfolds over time," *Nature Communications*, vol. 10, 2019.

[72] D. Patterson, "The human face in play based, shared, digital learning experiences," in *Proceedings of the Australasian Computer Science Week Multiconference*, 2019, ACM.

[73] T. Bachmann, *Perception of Pixelated Images.* Academic Press, 2016.

[74] N. Bowman, T. Wulf, "Finding nostalgia in the pixelated video games of decades past," Aug. 2018. https:// theconversation.com/finding-nostalgia-in-the-pixelated-video-games-of-decades-past-98518.

[75] T. Wulf, N. D. Bowman, J. A. Velez, J. Breuer, "Once upon a game: Exploring video game nostalgia and its impact on well-being," *Psychology of Popular Media Culture*, vol. 9, no. 1, pp. 83–95, 2018.

[76] C.-C. Carbon, "Understanding human perception by human-made illusions," *Frontiers in Human Neuroscience*, vol. 8, July 2014.

[77] R. M. Williams, R. V. Yampolskiy, "Optical illusions images dataset," *CoRR*, vol. abs/1810.00415, 2018.

[78] L. E. Nacke, "An introduction to physiological player metrics for evaluating games," in *Game Analytics, Maximizing the Value of Player Data*, 2013, pp. 585–619.

[79] G. Chanel, C. Rebetez, M. Bétrancourt, T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Transactions on Systems, Man, and Cybernetics -Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052–1063, 2011.

[80] P. Dragicevic, "Fair Statistical Communication in HCI," in *Modern Statistical Methods for HCI*, Springer, 2016, pp. 291– 330.

[81] J. B. Vornhagen, A. Tyack, E. D. Mekler, "Statistical significance testing at CHI PLAY: Challenges and opportunities for more transparency," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '20, New York, NY, USA, 2020, pp. 4–18, Association for Computing Machinery.

[82] P. Dragicevic, *Fair Statistical Communication in HCI*, pp. 291–330. Springer International Publishing, 2016.

[83] P. D. Ellis, *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results.* Cambridge University Press, 2010.

[84] K. Jung, J. Lee, V. Gupta, G. Cho, "Comparison of bootstrap confidence interval methods for GSCA using a Monte Carlo simulation," *Frontiers in Psychology*, vol. 10, 2019.

[85] X. Anadón, P. Sanahuja, V. J. Traver, A. Lopez, J. Ribelles, "Characterising players of a cube puzzle game with a two-level bag of words," in *Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2021, Utrecht, The Netherlands, June 21-25, 2021*, 2021, pp. 47–53, ACM.

Jose Ribelles

He is Associate Professor of the Computer Languages & Systems department and member of the Institute of New Imaging Technologies, both at the Universitat Jaume I of Castellon, Spain. He earned a PhD in Computer Science in 2000. His teaching covers subjets as Game Character Modeling, Computer Animation, or Image Synthesis. His research has focused on Computational Photography, Computer Graphics, and, more rencently, Videogames and Human-Computer Interaction.

Angeles Lopez

She has a PhD on Computer Science (Universitat Jaume I de Castellón, 1999) and is an Associate Professor in the Departament of Engineering and Science of Computers since 2002. She is a member of the Visual Engineering research group, integrated in the Institute of New Imaging Technologies (INIT). Her current research interests include depth estimation from stereo pair, single image or spherical panorama, and their applications in Computer Vision, Virtual Reality and Computational Photography.

V. Javier Traver

He is a member of the Computer Languages & Systems department at Universitat Jaume I, where he lectures at several undergraduate degrees (Computer Science, Videogame Design and Development, Intelligent Robotics) and at master level (Intelligent Systems). He is also affiliated with the Institute of New imaging Technologies (INIT), and his research interests include a broad range of topics within Computer Vision, Machine Learning, Human-Computer Interaction, and Videogames.