Mixed-initiative Creative Drawing with *webIconoscope*

Antonios Liapis

Institute of Digital Games, University of Malta
antonios.liapis@um.edu.mt

Abstract. This paper presents the *webIcononscope* tool for creative drawing, which allows users to draw simple icons composed of basic shapes and colors in order to represent abstract semantic concepts. The goal of this creative exercise is to create icons that are ambiguous enough to confuse other people attempting to guess which concept they represent. *webIcononscope* is available online and all creations can be browsed, rated and voted on by anyone; this democratizes the creative process and increases the motivation for creating both appealing and ambiguous icons. To complement the creativity of the human users attempting to create novel icons, several computational assistants provide suggestions which alter what the user is currently drawing based on certain criteria such as typicality and novelty. This paper reports trends in the creations of *webIcononscope* users, based also on feedback from an online audience.

1 Introduction

The creativity in human thought processes, design practices or engineering has been a topic of fascination since ancient times [17]. In recent years both philosophy and the cognitive sciences have allowed us to better understand and study the process of being creative. Creativity is no longer perceived solely as an activity of reclusive geniuses who conceptualize completely new theories or inventions, but also under the prism of an every-day, social form of creativity [5]. The last 30 years have seen a rise in popularity of the latter form of creativity (little-c creativity) both in the commercial innovation sector and in educational settings. *Lateral thinking*, i.e. the process of solving seemingly unsolvable problems or tackling non-trivial tasks through an indirect, non-linear, creative approach [4], is a skill that can be taught. The development of an educational curriculum around the collaborative, improvisational creativity of students in groups has gathered a strong support [21, 2]. According to a survey of European teachers [2], "schools promote a number of factors which favour creativity, such as learners' empowerment and open-mindedness, to rather a surprising extent" but "tend to promote other important creativity enhancing factors, such as risk-taking and mixing academic work and play, to a lesser degree".

To better integrate play into academic work, as noted above, the teaching process increasingly includes games. While commercial games in the right context can increase learner motivation and engagement [16, 26], it is also valuable to design games with the constraints of classroom use in mind, such as a limited play time or the need for short pauses for discussion. Such games are often aligned with an educational outcome or even an explicit topic. For instance, *Crystal Island* [20] tackles the topic of biology, as players interact with sick inhabitants of an island and attempt to find the solution to their ailments. In the realm of creativity support, many analog games promote creativity, e.g. *LEGO* bricks or the card game *Once Upon a Time* (Atlas Games 2012). However, there are few attempts at digital games designed to foster creativity in the classroom.

This paper describes *webIconoscope*, an online publicly available version of the Iconoscope game which was explicitly designed for fostering creativity within an educational setting [7]. In Iconoscope learners play in a group, attempting to draw icons which the other players will not be able to identify easily. Icons created via Iconoscope are evaluated by the other players, who vote which of the 3 possible concepts (described in words) is represented in the icon. The winner of a round of Iconoscope is the one with the most ambiguous icon, i.e. the icon with an equal number of correct and incorrect guesses. Ambiguous icons allow multiple interpretations from the viewers, and is an important disruptor which prompts lateral thinking [22]. Ambiguity results in cognitive dissonance between image and associated concept(s), or on the visual level itself (e.g. in optical illusions); this requires a creative, playful reading of the image. The Iconoscope game is played on Android tablets in groups of 4 or more players; each player uses a tablet to draw icons and passes it around the table during the voting phase. Iconoscope was deployed in educational institutions along with other creativity-oriented applications and games such as *4Scribes* [9]. Due to security and privacy concerns in educational settings, icons created with Iconoscope were only accessible to other users in the same educational institution. Instead, *webIconoscope* allows the anonymous use of both icon drawing and voting, making the icons publicly available and allowing anyone to engage with user-created content. This increases the application's publicity but more importantly allows for a broader evaluation of the user-created icons by a broader group of people. Compared to Iconoscope, *webIconoscope* lacks face-to-face interaction and feedback (relying instead on impersonal quantitative feedback such as the number of correct guesses) and the chance for immediate wins or losses. On the other hand, by redesigning the evaluation of ambiguity for potentially numerous votes and by enhancing the interface for audience feedback and presentation of results (e.g. as a leaderboard), *webIconoscope* transforms the short game sessions of Iconoscope into a broader, more social and public showcase of human (and human-computer) creativity.

2 Mixed-Initiative Co-Creation

With the pervasiveness of the digital world in every aspect of people's lives, diverse computer-aided design tools have emerged. Mixed-initiative tools are a special case of computer-aided design, where the computer takes on a more proactive role [27, 8]. Mixed-initiative tools rely on both a human initiative and a computational initiative to perform the creative tasks and take the creative decisions. Likening the creative process to a conversation, Novick and Sutton [14] identify three types of initiative: the *task initiative* (who introduces the problem), the *outcome initiative* (who decides whether the problem has been solved), and the *speaker initiative* (who decides whose turn it is to speak).

Mixed-initiative interaction has been extensively explored for game design tasks such as level creation. There is a breadth of level design tools with different degrees of computational initiative, for instance showing optional suggestions in *Sentient Sketchbook* [11] or guiding the creative process with some indirect human guidance in interactive evolution tasks e.g. in [10]. In other cases user creations act as a goal for the computer to approximate, such as recreating a user's rough sketch in higher resolution [12]. In *Tanagra* [24], the computer attempts to "fill in the gaps" left by the user, while obeying user-specified constraints.

Besides designing the functional properties of games, mixed-initiative tools have also been used for freeform creative tasks. Examples where such creative tasks are part of a game setting can be found in *Petalz* [18] and *Artefacts* [15], where the core game mechanic is interactive evolution [25] (IEC) of flowers and blocks respectively. In Petalz, evolved flowers are posted on one's public gallery (their "balcony"); players can view their Facebook friends' balconies, like and comment on specific flowers, and sell their flowers at a marketplace for in-game currency. Artefacts is a sandbox creation game where players evolve 3D blocks into interesting shapes, combining them into complex 3D "sculptures". Artefacts lacks the social mechanics of Petalz (e.g. ownership of a balcony, sharing and liking, marketplace) and is closer to IEC in evolutionary art and music where users select which pieces will evolve without an external purpose or motivation. Among such interactive evolutionary art projects, of special note is *PicBreeder* [23] and *DrawCompileEvolve* [28]. Both systems allow users to submit an evolved image to a common public gallery, thus inviting others to rate how much they like the image (using a 5-star rating scale) or evolve the icon further via IEC. Users evolving each others' images allows for shared ownership of the output as well as negotiations of an image's meaning; both factors are important for little-c creativity to emerge [3]. DrawCompileEvolve allows for more human initiative than traditional IEC, as users seed evolution from their own drawn images.

Both *webIconoscope* and its predecessor Iconoscope follow a mixed-initiative approach to user interaction, with the computational initiative presenting optional suggestions (similarly to *Sentient Sketchbook*) but at the user's request (similarly to *Tanagra*). The computational initiative appears as assistants with profile pictures, names and implied personalities, thus strengthening the analogy of a conversation with the computer. In terms of collaborative creativity, *webIconoscope* borrows from the principles of Petalz and PicBreeder, with a public gallery that allows users to engage with each others' work (although they can not edit them further). Finally, *webIconoscope* goes beyond freeform creative exploration projects such as Artefacts and PicBreeder as it motivates users to guess the concept represented in the image and has a leaderboard for the most ambiguous images. This gives more purpose to the interaction with existing artifacts and a clearer framing of the goals of the creative process of new artifacts.

(a) List of concept triplets, shown at the start of a drawing session.

(b) Drawing interface of *webIconoscope*, with the interface elements annotated.

Fig. 1. The interface of *webIconoscope*, embedded on a webpage as a Flash application.

3 The *WebIconoscope* Interface

At the highest-level, *webIconoscope* is a drawing tool, not unlike Microsoft Paint or OpenOffice Draw. However, *webIconoscope* has more clear-cut goals for the player, with a target concept that must be drawn and the aim of creating ambiguous icons that can represent more than one concept.

Typical Use Case of *webIconoscope*: In a sample use case of *webIconoscope*, a user starts by selecting a language of their choice, reading through the instructions page and inserting their username and an e-mail in case they want to participate in competitions[1]. Once this initial setup is completed, the user selects a triplet of concepts: each triplet is displayed on a post-it note (see Fig. 1a). When the user selects a triplet, they must also choose which concept they wish to draw among the three. The goal of the user is to create ambiguous icons which could be mis-interpreted as the other two concepts. The concepts are chosen by pedagogy experts to include thematically coherent but opposing ideas (e.g. “Protest”, “Conform”, “Sit on the Fence”) or concepts that are semantically similar (e.g. “Freedom”, “Play”,’ “Enthusiasm”). This way, the user is challenged to find the relationships on the semantic, thematic or visual level which can be exploited to create ambiguous icons. Once the player chooses a concept to draw, they are taken to the drawing interface where they can create their icon; once they are happy with their creation, or at the end of 5 minutes, the drawing is finished and uploaded to the database. The player has an option of choosing another concept to draw; if they do not, they are taken to the gallery page where they can survey their own and others’ creations, vote for which concept is represented in each icon and rate the icons in terms of appeal.

[1] The launch of *webIconoscope* was followed by a competition running for 3 months; similar competitions are planned for the future to increase the use of *webIconoscope*.

(a) A sample of the gallery interface.

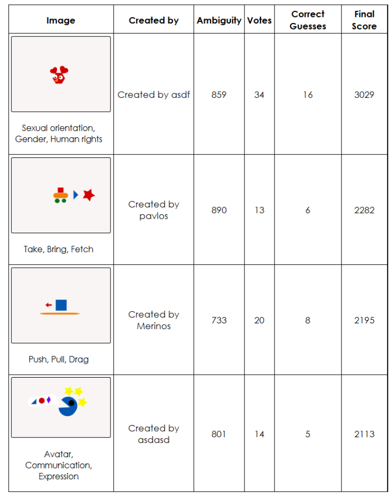


Image	Created by	Ambiguity	Votes	Correct Guesses	Final Score
Sexual orientation, Gender, Human rights	Created by asdf	859	34	16	3029
Take, Bring, Fetch	Created by pavlos	890	13	6	2282
Push, Pull, Drag	Created by Merinos	733	20	8	2195
Avatar, Communication, Expression	Created by asdasd	801	14	5	2113

(b) First 4 entries of the leaderboard.

Fig. 2. The website elements of *webIconoscope*.

Drawing Interface: The drawing interface of *webIconoscope* looks like a drawing table (see Fig. 1b), showing a clock and the concepts to be drawn on a post-it note (1,2 respectively in Fig. 1b). Drawing in *webIconoscope* is limited to the placement of pre-made abstract shapes, shown as a stencil (3). New shapes appear at the center of the screen in a neutral gray color: the user can then move, rotate, scale or recolor the shape as they desire. The shapes are mostly basic geometric shapes (squares, rhombi, circles, hexagons, triangles) and some more memorable shapes (star, heart). Shapes can be recolored via the palette (4): the palette has a small number of colors (mostly primary and secondary), as well as black and white. A button (5) allows users to end the drawing session before the five minutes are over. The interface includes portraits for all the computational assistants (6) which will be described in Section 4.

Website Interface: The main difference of *webIconoscope* from the multiplayer tablet-based digital game Iconoscope [7] is that the former is embedded in a website[2] which allows for many anonymous users to draw new icons and to survey previously created icons. All icons created through the *webIconoscope* interface are stored in a database alongside information on the concept represented and other interaction data. All icons in the database are shown in a gallery (see Fig. 2a), where any user can vote for which concept is represented by each icon (using the same concept triplet as the one used while drawing the icon). For each icon, the creator's name, the number of guesses by other users and the number of correct guesses are also displayed, as an invitation for the

[2] `http://iconoscope.institutedigitalgames.com/`

user to guess correctly. The user can also rate the icon from 1 to 5 stars, based on how much they 'like' the icon. Since the goal of the created icons is ambiguity rather than appeal, the rating interface (and stars) is smaller and underplayed. Once a user has provided feedback on an icon, their selection is highlighted and "locked". Similarly, once they provide a star rating, that section is replaced by the average rating for this icon; users can not change their votes or ratings.

To promote competition, a page on the website displays a leaderboard of the top 10 icons (see Fig. 2b) along with their creator's name, the concept triplets they could be representing, and metrics on user feedback. These metrics include the number of votes, the number of correct guesses, the ambiguity score and a final score used to rank the top 10 icons. The ambiguity score A (calculated via Eq. 1) rewards icons with an equal number of correct and incorrect guesses, and also rewards a balance between the two wrong options. The final score F (Eq. 2) rewards icons with high ambiguity but also favors icons with more votes; the rationale being that more votes not only denote popularity but are more difficult to "get right" in terms of balanced correct and incorrect guesses.

$$A = 1000 - 500 \cdot \left(|1 - 2 \cdot \tfrac{c}{t}| + |1 - 2 \cdot \tfrac{i_{max}}{i}| \right) \quad (1)$$

$$F = log(t) \cdot A \quad (2)$$

where t is the total number of votes; c the number of correct votes; i the number of total incorrect votes; i_{max} the number of votes for the wrong option with the most votes. If $t = 0$ (i.e. there are no votes) then both A and F scores are 0.

4 Computational Assistants

In order to provide a creative stimulus, a set of *computational assistants* were added to the solitary drawing task of *webIconoscope*. The computational assistants show the user mutations of their current icon; the user can choose one of the four suggested alternatives to replace their icon and continue drawing (see Fig. 4) or discard all suggestions. The user can ask a computational assistant for suggestions at any point during the icon drawing process via its portrait on the drawing screen. All assistants are shown and can be interacted with in the drawing interface (see 6 in Fig. 1b); to motivate the use of assistants, every few seconds a random assistant's portrait swings while a dialog balloon pops up.

When selected, all computational assistants in *webIconoscope* perform a short evolutionary sprint, starting from an initial population consisting of mutated copies of the user's icon. Mutation can clone an existing shape in the icon (moving it to a random position) or remove a random shape. Moreover, during mutation every shape in the icon has a chance to be moved, rotated, scaled, recolored, or changed into another shape (e.g. a circle changing into a square).

Each assistant has a unique name and portrait (see Fig. 3), and they search the space differently: *Chaotic Kate* merely performs 10 random mutations to 4 copies of the user's icon, *Mad Scientist* performs novelty search [6] to diversify the population, *Typical Tom* and *Progressive Petra* attempt to respectively approach and deviate from a typical icon for this concept.

(a) (b) (c) (d)

Fig. 3. Assistant profiles: Chaotic Kate (a), Mad Scientist (b), Typical Tom (c), Progressive Petra (d).

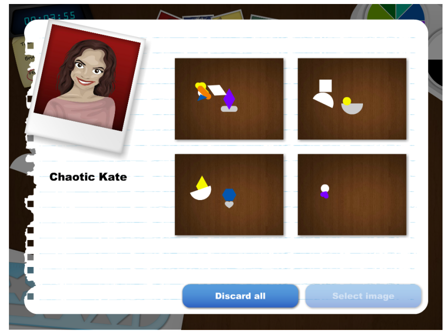


Fig. 4. Suggestions by Chaotic Kate.

Chaotic Kate performs the equivalent of a random walk and does not evaluate the quality of the content it produces in any way. The remaining three assistants choose the most promising individuals via fitness-proportionate roulette wheel selection. Icons evolve based on mutation alone and their fitness is computed based on the distance between phenotypes (icons). Since there is no obvious way of evaluating image diversity, a random distance metric based on difference in types, colors and positions of shapes is chosen: the five distance metrics are shown in Eq. (3)-(7) and illustrated in Fig. 5. Eq. (3) evaluates the number of colors that are not shared by both icons and Eq. (4) the number of shape types (e.g. circle, square) not shared by both icons. Eq. (5) evaluates the number of different shape types and colors between icons, penalized by the number of shapes that share both shape type and color in both icons. This assumes that different shapes and colors are both perceptually and semantically different, but the same shapes with the same color (e.g. a red star) are important in carrying the meaning from one icon to the next (regardless of size or number of shapes). Eq. (6) evaluates the average distance between all icons of one shape with all icons of the other shape: it largely rewards shapes placed in similar positions (also near each other) in both icons. Finally, Eq. (7) evaluates the difference in how "grouped" the shapes in each icon are.

$$d_c(i,j) = \frac{D_c(i,j)}{S_c(i,j)+D_c(i,j)} \tag{3}$$

$$d_s(i,j) = \frac{D_s(i,j)}{S_s(i,j)+D_s(i,j)} \tag{4}$$

$$d_{c,s}(i,j) = \frac{D_c(i,j)}{S_c(i,j)+D_c(i,j)} + \frac{D_s(i,j)}{S_s(i,j)+D_s(i,j)} + 10 \cdot S_{c,s}(i,j) \tag{5}$$

$$d_d(i,j) = \frac{1}{N(i) \cdot N(j)} \sum_{k=1}^{N(i)} \sum_{l=1}^{N(j)} d(\boldsymbol{p}_k, \boldsymbol{p}_l) \tag{6}$$

$$d_g(i,j) = |G(i) - G(j)| \tag{7}$$

where $D_c(i,j)$ and $S_c(i,j)$ the number of colors not common and common (respectively) in icons i and j; $D_s(i,j)$ and $S_s(i,j)$ the number of shape types not

(a) Icon 1 (b) Icon 2 (c) d_d calculation (d) d_g calculation

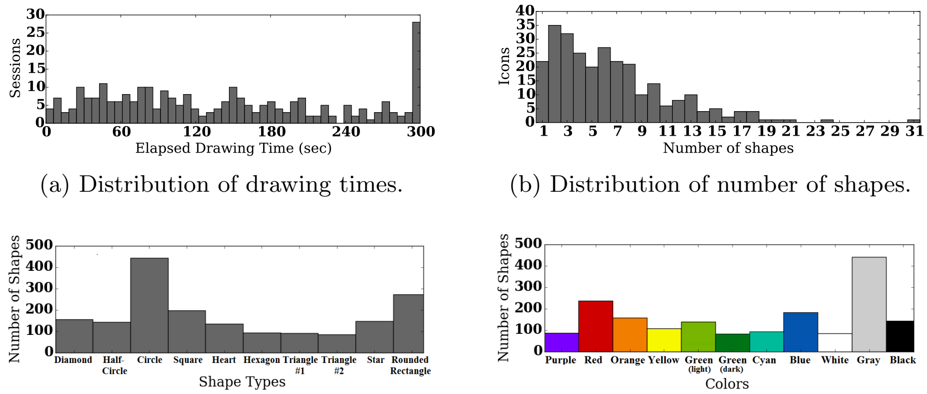
Fig. 5. Distance metrics calculation for icons 1 and 2 (Fig. 5a and 5b). The icons do not share colors so $d_c = 1$ ($D_c = 3$ and $S_c = 0$); they share the half-circle shape type so $d_s = \frac{3}{4}$ since $D_s = 3$ (circle, rectangle, diamond) and $S_s = 1$ (half-circle). Fig. 5c shows the calculation of d_d, which is the average distance of all shapes in icon 1 to all shapes in icon 2 (black lines). Fig. 5d shows grouping G of icon 1 (average distance of green lines) and icon 2 (average distance of blue lines); their absolute difference is d_g.

common and common (respectively); $S_{c,s}(i, j)$ the number of combinations of color and shape common in the two icons; $N(i)$ the number of shapes in icon i; $d(\boldsymbol{p}_k, \boldsymbol{p}_l)$ the Euclidean distance between the centers of shape k and shape l; $G(i)$ the average Euclidean distance of all shapes in icon i.

These distance metrics are very lightweight computationally, especially when used one at a time: in comparison, using pixel-based distance of two images with 1200 by 900 pixels (as the ones shown in Fig. 5) would be impossible to compute in real-time, not to mention use for evolution. Choosing one metric, however, comes at the cost of expressivity and accuracy of evaluations: for instance, using d_d means the shapes and colors will remain the same in all suggestions (excluding random mutations). The obvious benefit of one distance metric is the reduced effort which allows for almost real-time generation of suggestions. A welcome side-effect of a random distance metric, however, is that it is almost impossible for the user to anticipate the resulting artifacts since at times they feature different shapes, at times different colors, and at times different positioning.

In the case of novelty search [6] (performed by the Mad Scientist), fitness is calculated based on the average distance of the individual with other members of the population and a novelty archive. The novelty archive initially contains the user's icon and in every generation the fittest (most novel) individual in the population is added to it. By diverging from the novelty archive, evolution maintains a memory of where the search has been and attempts to deviate from both historical (via the novelty archive) and current (via the current population) areas of the search space. While novelty search [6] traditionally considers a subset of the population and archive when computing novelty, the small sizes of both the population and the archive allows us to consider all individuals.

In the case of typicality search (performed by Typical Tom and Progressive Petra), fitness is calculated based on the distance from a *typical* icon for this concept. Typical icons are inserted into the database by experts and include simple but characteristic icons (e.g. a red heart for "kindness" or green triangles for "nature"). If no typical icon is found for a concept, a random one is created and added to the database instead. Typical Tom attempts to minimize the distance between the evolving icon and the typical icon: depending on the distance



(a) Distribution of drawing times.

(b) Distribution of number of shapes.

(c) Distribution of shape types.

(d) Distribution of colors.

Fig. 6. Patterns of user interaction and final icons in *webIconoscope*.

function, this may mean that for "kindness" at least one shape should become a heart (for d_s) or red (for d_c). Progressive Petra attempts to maximize the distance between the evolving icon and the typical icon, and so for "kindness" it might eliminate all instances of red hearts from the icon (for $d_{c,s}$).

5 Results

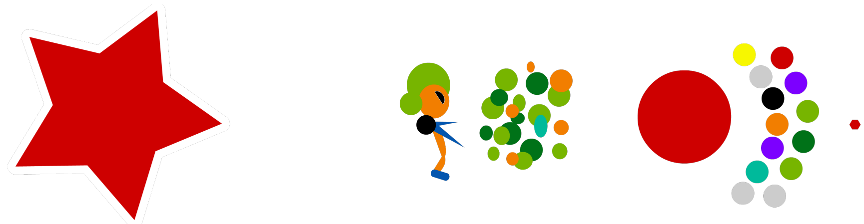
The results reported will discuss the degree of use of *webIconoscope* in terms of icons created via the drawing interface (in Section 5.1) and the response of the audience via the gallery (in Section 5.3). Moreover, the mixed-initiative aspect of *webIconoscope* will be evaluated in terms of the usefulness of its computational assistants (in Section 5.2).

5.1 Icons Created

Since its first launch in September of 2015 until the time of writing (October 2016), 275 valid icons have been created. Valid icons, in this case, have at least one shape and do not feature inappropriate content; the database was manually checked to ensure this. Of these 275 icons, most were created during 2015 due to a competition in schools and an extensive publicity push on social media.

Looking at patterns of *webIconoscope* use, most users chose English concepts (48%) with Greek concepts coming in a close second (45%) and German concepts at 7%. The primary reason for extensive use of Greek concepts was a publicity push in Greek schools. In general, the duration of drawing sessions in *webIconoscope* varied widely (see Fig. 6a) from 2 seconds to the full 5 minutes. The average drawing time was 142 seconds (standard deviation of 92 seconds).

Regarding the patterns of the final icons themselves, there was a large deviation in the number of shapes in each icon as shown in the distribution of



(a) Icon for the "Passive", "Aggressive" or "Assertive" German triplet.

(b) Icon for the "Nature', "Mankind" or "Technology" triplet.

(c) Icon for the "Sexual Orientation", "Gender" or "Social Rights" triplet.

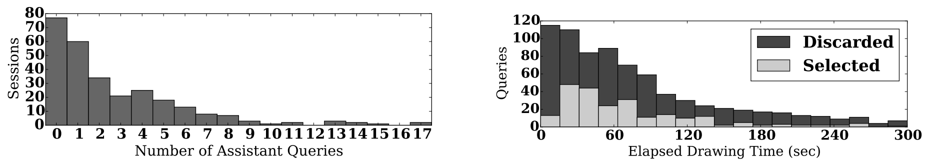
Fig. 7. Sample icons with the fewest shapes, the most shapes and the most colors.

Fig. 6b. While icons have an average of 6.4 shapes (standard deviation of 4.8), a non-trivial number of icons had only 1 shape (8%). This points to users who spend little time and effort drawing the simplest icons. On the other hand, certain icons were quite elaborate, with as many as 24 and 31 shapes. Regarding the types of shapes favored, Fig. 6c shows their distribution in all icons. Circles were the most popular shape; the thin rounded rectangle was interestingly the second most popular, as it seems to be used as a line[3]. It is surprising that shapes with more semantic associations (i.e. hearts and stars) were not used frequently. Finally, as shown in Fig. 6d the most prevalent color in the icons is gray, which is the default color (25% of all shapes), followed by red (13%). Again, gray shapes point to users that did not spend much time drawing[4]. In terms of color variety, most icons had one color (22%), three colors (20%) or two colors (18%), although some icons had as many as eight (2%) or nine (1%) colors.

To illustrate the variety in icons created via *webIconoscope*, Fig. 7 shows a sample of the icons, i.e. an icon with the fewest shapes (one), an icon with the most shapes (31) and an icon with the most colors (9). The large star of Fig. 7a plays with popular and historical symbols involving stars and uses a strong color (red) and a large size to show the "Aggressiveness" of the symbol; red color is often associated with passion and anger ("seeing red"). Fig. 7b uses simple shapes (primarily circles) to make a composite shape: that of a human with blue triangles acting as blades (possibly). Green circles hint at "Nature", the human figure hints at "Mankind" and finally the blue triangles could hint at "Technology" (as they seem to replace human hands). Fig. 7c is a multi-colored but highly abstract icon, juxtaposing a large red circle to a small hexagon. The size difference is obvious but the shape difference less so (making it intriguing). A group of multi-colored circles surround the large circle, perhaps in a threatening or idolizing fashion. It is not immediately obvious how this composition relates to any of the three concepts, although it could be a commentary on exclusion (for the "Social Rights" concept) or inequality (e.g. for "Gender").

[3] *webIconoscope* does not allow users to draw lines.

[4] Gray is not a color that users can pick, so any gray shapes were never re-colored.



(a) Assistant queries per session. (b) Timing of assistant queries.

Fig. 8. Assistant use.

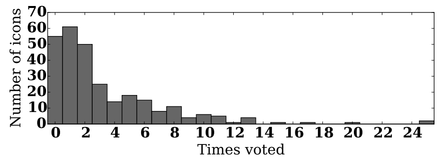
5.2 Assistant Contributions

Most drawing sessions of *webIconoscope* (201 out of 275) included one or more queries to computational assistants, with users viewing their suggestions. This finding is surprising, as in the tablet-based Iconoscope game users forgot to consult the assistants. There are two reasons for this shift in user behavior: (a) the assistants in *webIconoscope* are more animated, swaying and showing a speech bubble from time to time, and (b) drawing in *webIconoscope* is a single-user experience (often at home rather than a classroom) with fewer distractions.

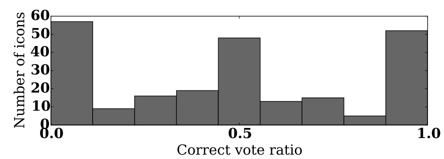
While assistants were queried at least once in 201 sessions, there were often more than one queries per session. With a total of 747 queries in those 201 sessions, this averages to 3.7 queries per session. As shown in Fig. 8a, in most sessions assistants were queried once, possibly as a test; however in some sessions they were queried extensively (up to 17 times). Due to the many sessions with only one assistant query, one can assume that users either liked the assistants' suggestions or didn't; those that queried an assistant once likely did not appreciate the suggestions and did not query it again.

There did not seem to be a big difference between queries to different assistants: 21% of queries were to Chaotic Kate, 30% to the Mad Scientist, 29% to Typical Tom and 20% to Progressive Petra. It is suspected that users chose whichever assistant was animating at the time (which was randomly chosen).

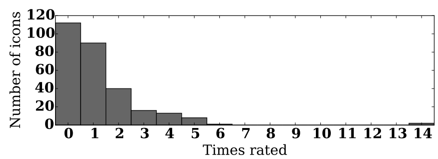
A relevant analysis for assistant queries is the timing when such a query was made: Fig. 8b shows the distribution of elapsed drawing time when assistants were queried. It is obvious that most users used assistants early in the drawing process, in the first 30-100 seconds; as in earlier findings in Sentient Sketchbook [11], the suggestions are often used as inspiration in early stages of the icon design when a blank canvas causes creative block. Assistants rarely get queried late in the process (in part due to the fact that few drawing sessions lasted the full 5 minutes). This is contrary to the use of Sentient Sketchbook suggestions, where many designers used the suggestions to fine-tune a design to e.g. reach perfect scores in game balance. This is likely because the evolutionary algorithms (which mostly reward divergence) and the high-level and context-agnostic distance metrics result in visually "noisy" suggestions which often break patterns of the users' icons. Thus, querying assistants at the final stages of the design process is almost guaranteed to "break" the user's nigh-final icon, which is undesirable. It is important to note that while assistants were queried often, their



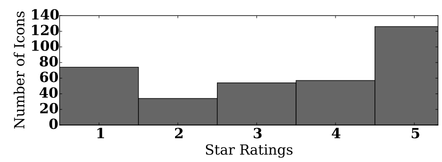
(a) Distribution of times each icon was voted for.



(b) Distribution of the ratio of correct votes over all votes of the same icon.



(c) Distribution of times each icon was rated.



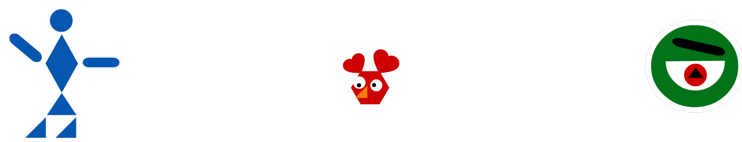
(d) Number of stars rated in each interaction with *webIconoscope*.

Fig. 9. Audience feedback to *webIconoscope* icons.

suggestions were selected to replace the user's icon in 102 of those sessions. However, in those 102 sessions, the assistants' suggestions were selected 230 times, i.e. more than twice per session on average. There was only a slight bias towards selecting suggestions from the Mad Scientist (29% of all selected suggestions); the least popular was Progressive Petra (21%). As noted above there was a divide between users who liked the suggestions and those who didn't. Users who queried assistants once actually selected a suggestion only in 32% of the sessions; by comparison, users who queried assistants 2 to 4 times selected a suggestion at least once in 53% of the sessions and those who queried assistants over 4 times selected a suggestion at least once in 67% of the sessions. Obviously, those who found the suggestions appealing queried the assistants more often and were more likely to take advantage of their suggestions. Interestingly, the ratio of selected suggestions versus queried assistants does not change regardless whether users make one, two or many queries: there is roughly a 1 in 3 chance that suggestions were found appropriate any time a user queried a computational assistant.

5.3 Audience Feedback

Since September 2015, there were 984 interactions with the feedback section of *webIconoscope*. Out of those, 935 were votes on which concept is represented while 347 rated an icon's appeal. The downplayed role of rating (with smaller stars and text located under the voting buttons) explains the lower engagement with rating; since the main goal of *webIconoscope* was to create ambiguity in icons, the large proportion of guessing interactions was the desired outcome. As with the drawing sessions of *webIconoscope*, the majority of interactions (89%) were made during 2015 due to dissemination in schools and a competition.



(a) Icon for the "Avatar", "Communication", "Expression" triplet.

(b) Icon for the "Sexual Orientation", "Gender" or "Social Rights" triplet.

(c) Icon for the "Danger", "Safety" or "Protecting the young" Greek triplet.

Fig. 10. Icons with the highest ambiguity score (Fig. 10a), most votes (Fig. 10b) and highest ratings on average (Fig. 10c).

As shown in Fig. 9a, many of the icons were voted for once (22% of all icons) or not at all (20%). Many of the icons, therefore, did not entice users to attempt to guess their concept, although the lower use of *webIconoscope* in 2016 could also explain the few votes: icons created in 2016 were not seen by many users. While most icons received few (if any) votes, 28% of the icons received 5 or more votes. The icon with the most votes (25) is shown in Fig. 10b; it also has the highest final score, largely due to this large number of votes, i.e. t in Eq. (2).

Regarding correct versus incorrect guesses in user's votes, Fig. 9b shows the distribution of that ratio. Since many icons received a few votes (i.e. one or two), in many cases all votes were incorrect (24% of all icons) or all votes were correct (22%); similarly, in many cases (21%) there was an equal number of correct and incorrect votes. This prevalence of all-correct or all-incorrect guesses, coupled with many icons being voted once or not at all, led to ambiguity scores of 500 (40% of icons) and 0 (19%) due to no votes. Only 3% of the icons reached ambiguity scores between 950 and 1000 (the maximum value). Among icons with maximum ambiguity, the icon in Fig. 10a has the most votes (8): this means that the icon has 4 correct votes and 2 votes for each of the wrong concepts.

In terms of ratings, Figure 9d shows the distribution of all ratings to icons in *webIconoscope*; as expected most ratings have extreme values (5 stars or 1 star) since people often give feedback when they really like or dislike something. As shown in Fig. 9c many icons were rated once or not at all, which is not surprising since many of the newer icons created during 2016 were not seen by many users. One icon was rated 14 times (shown in Fig. 10b). Among the icons with an average rating of 5 stars, the one with the most ratings is shown in Fig. 10c. The eye (formed by the half-circle) hints at surveillance (perhaps "Safety" or "Protecting the young"); the red color of the iris and the eyebrow (formed by the rounded rectangle) makes the expression angry which hints at "Danger".

6 Discussion and Future Work

The online deployment of *webIconoscope* allowed a large number of users to draw icons and leave them for other users to appraise. The 275 icons in the *webIconoscope* database feature a broad range of visual styles and topics (i.e. concept triplets). Icons were often fairly simple, with a few shapes and colors; however, even a few shapes (such as the six shapes in Fig. 10c) can be quite effective at creating a strong visual message. Some icons proved more popular than others, not only in terms of their average ratings but also in terms of the number of times other users interacted with them (either to guess the concept or to rate it). Unfortunately, audience interactions during 2016 was not proportional to the number of icons created during this period, leading to many of the latest icons having few if any votes or ratings. Another competition or publicity push could help increase attention and engagement with the recently created icons.

In terms of the computational assistants, their representation as human-like portraits and their periodic movement made them enticing to interact with. However, based on Section 5.2 the suggestions provided by the computational assistants were often not deemed appropriate and thus were not selected to replace the user's icon. The 'visual difference' metrics used to drive 3 of the 4 assistants were admittedly quite simplistic, favoring fast computation over accurate evaluation of perceptual differences. The role of assistants was primarily that of a random stimulus [1] which could break the user's frame of reference [22] and drawing practices. Under this prism, the fact that they were selected to replace a user's icon at least once in 102 of 275 sessions should be considered an achievement. On the other hand, there are many possible improvements in the generative processes of computational suggestions in order to increase the appeal, usefulness, and co-creative potential of *webIconoscope*. An obvious improvement could be the combination of all 5 distance metrics into a Euclidean distance: however this would increase the computational cost as well as introduce imbalances due to different value ranges in the distance metrics. Another improvement could be in the mutation operators, where rather than manipulating each shape individually, mutation can create groups (e.g. by color, by shape or by proximity) and apply scaling, recoloring, or cloning operators on all shapes in the group. This would increase the consistency and the semantic attributes of the user's icon, as a group of shapes is likely more than the sum of its parts.

While the analysis of the users' contributions (both in terms of icons drawn and votes on others' icons) is extensive in this paper, there is potential for more in-depth evaluation of the data collected via *webIconoscope*. The icons drawn can be analyzed via unsupervised machine learning techniques such as Non-Negative Matrix Factorization to find clusters with similar shapes and colors [13]. Moreover, the semantic associations of icons can be learned by applying supervised machine learning (e.g. Deep Learning) so that the system is able to predict the semantic association of an icon by its appearance alone. Finally, similar supervised learning methods can be used to learn a mapping between the icon and the average audience rating. Such a computational model of visual aesthetics for icons can then be used as an objective function for a computational assistant targeting value [19] (rather than typicality and novelty which are targeted currently), or used as a constraint on minimum predicted audience rating which tests all computational suggestions before they are presented to the user.

7 Conclusion

This paper presented the interface of *webIconoscope* and its outcomes on the first year of its deployment. Its online availability led to a frequent use of the creative aspect (with 275 user-created icons) as well as the evaluation aspect (with 984 instances of audience feedback) of *webIconoscope*. The included computational assistants were designed to act as a disruptor to the user's frame of reference while drawing; although they were frequently queried, the rather simplistic way in which they target visual diversity led users to often discard their suggestions. Overall, *webIconoscope* managed to collect a rich dataset of visual depictions of abstract concepts (see Fig. 7 and 10) which can be exploited further to analyze visual aesthetics, or learn computational models of icon ambiguity or appeal.

Acknowledgment

The author would like to thank Serious Games Interactive and the FP7 ICT project C2Learn (project No: 318480) for the implementation of Iconoscope. The ongoing research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No: 693150.

References

1. Beaney, M.: Imagination and creativity. Open University Milton Keynes (2005)
2. Cachia, R., Ferrari, A., Kearney, C., Punie, Y., Van, W., Berghe, D., Wastiau, P.: Creativity in schools in Europe: A survey of teachers. http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=2940 (2009), accessed: November 2016
3. Chappell, K., Craft, A.R., Rolfe, L., Jobbins, V.: Humanizing creativity: Valuing our journeys of becoming. International Journal of Education & the Arts 13(8) (2012)
4. De Bono, E.: Lateral thinking: Creativity step by step. Harper Collins (2010)
5. Jeffrey, B., Craft, A.: The universalization of creativity. In: Creativity in Education. Continuum (2001)
6. Lehman, J., Stanley, K.O.: Abandoning objectives: Evolution through the search for novelty alone. Evolutionary Computation 19(2) (2011)
7. Liapis, A., Hoover, A.K., Yannakakis, G.N., Alexopoulos, C., Dimaraki, E.V.: Motivating visual interpretations in iconoscope: Designing a game for fostering creativity. In: Proceedings of the Foundations of Digital Games Conference (2015)
8. Liapis, A., Yannakakis, G.N.: Boosting computational creativity with human interaction in mixed-initiative co-creation tasks. In: Proceedings of the ICCC workshop on Computational Creativity and Games (2016)
9. Liapis, A., Yannakakis, G.N., Alexopoulos, C., Lopes, P.: Can computers foster human users' creativity? theory and praxis of mixed-initiative co-creativity. Digital Culture & Education (DCE) 8(2), 136–152 (2016)
10. Liapis, A., Yannakakis, G.N., Togelius, J.: Limitations of choice-based interactive evolution for game level design. In: Proceedings of AIIDE Workshop on Human Computation in Digital Entertainment (2012)

11. Liapis, A., Yannakakis, G.N., Togelius, J.: Sentient sketchbook: Computer-aided game level authoring. In: Proceedings of the 8th Conference on the Foundations of Digital Games. pp. 213–220 (2013)
12. Liapis, A., Yannakakis, G.N., Togelius, J.: Sentient world: Human-based procedural cartography. In: Evolutionary and Biologically Inspired Music, Sound, Art and Design, vol. 7834, LNCS, pp. 180–191. Springer (2013)
13. Lim, C.U., Liapis, A., Harrell, D.F.: Discovering social and aesthetic categories of avatars: A bottom-up artificial intelligence approach using image clustering. In: Proceedings of the International Joint Conference of DiGRA and FDG (2016)
14. Novick, D., Sutton, S.: What is mixed-initiative interaction? In: Proceedings of the AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction (1997)
15. Patrascu, C., Risi, S.: Artefacts: Minecraft meets collaborative interactive evolution. In: Proceedings of the IEEE Conference on Computational Intelligence and Games (CIG) (2016)
16. Pirius, L.K., Creel, G.: Reflections on play, pedagogy, and World of Warcraft. EDUCAUSE Quarterly 33 (2010)
17. Plato: Ion. In: Plato: The Collected Dialogues. Princeton University Press (1961)
18. Risi, S., Lehman, J., D'Ambrosio, D.B., Hall, R., Stanley, K.O.: Combining search-based procedural content generation and social gaming in the petalz video game. In: Proceedings of the Artificial Intelligence and Interactive Digital Entertainment Conference (2012)
19. Ritchie, G.: Some empirical criteria for attributing creativity to a computer program. Minds and Machines 17(1), 67–99 (2007)
20. Rowe, J., Shores, L., Mott, B., Lester, J.: Integrating learning, problem solving, and engagement in narrative-centered learning environments. International Journal of Artificial Intelligence in Education 21(1-2), 115–133 (2011)
21. Sawyer, K.: Educating for innovation. Thinking Skills and Creativity 1, 41–48 (2006)
22. Scaltsas, T., Alexopoulos, C.: Creating creativity through emotive thinking. In: Proceedings of the World Congress of Philosophy (2013)
23. Secretan, J., Beato, N., D'Ambrosio, D.B., Rodriguez, A., Campbell, A., Folsom-Kovarik, J.T., Stanley, K.O.: Picbreeder: A case study in collaborative evolutionary exploration of design space. Evolutionary Computation 19(3), 373–403 (Sep 2011)
24. Smith, G., Whitehead, J., Mateas, M.: Tanagra: Reactive planning and constraint solving for mixed-initiative level design. IEEE Transactions on Computational Intelligence and AI in Games 3(3) (2011)
25. Takagi, H.: Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. Proceedings of the IEEE 89(9), 1275–1296 (2001), (invited paper)
26. Watters, A.: Legos for the digital age: Students build imaginary worlds. http://blogs.kqed.org/mindshift/2011/03/legos-for-the-digital-age-students-build-imaginary-worlds/ (2011), accessed: November 2016
27. Yannakakis, G.N., Liapis, A., Alexopoulos, C.: Mixed-initiative co-creativity. In: Proceedings of the 9th Conference on the Foundations of Digital Games (2014)
28. Zhang, J., Taarnby, R., Liapis, A., Risi, S.: DrawCompileEvolve: Sparking interactive evolutionary art with human creations. In: Evolutionary and Biologically Inspired Music, Sound, Art and Design, vol. 9027, LNCS. Springer (2015)