Assessing Usefulness of a Visual Blending System: *"Pictionary Has Used Image-making New Meaning Logic for Decades. We Don't Need a Computational Platform to Explore the Blending Phenomena", Do We?*

João M. Cunha, Sérgio Rebelo, Pedro Martins and Penousal Machado
CISUC, Department of Informatics Engineering
University of Coimbra
{jmacunha,srebelo,pjmm,machado}@dei.uc.pt

Abstract

Visual blending can be employed for the visual representation of concepts, merging input representations to obtain new meanings. Yet, the question arises whether there is any need for a computational approach to visual blending. We address the topic using different points of view and conduct two user studies to assess the usefulness of a visual blending system.

Introduction

The mere technical possibility to implement a given computational system can be regarded as insufficient to justify its actual implementation. In most cases, there is the question of how useful a given system would be to the user – e.g. is there really a need for a computational system for meaning-making through visual blending? In fact, this question can be analysed from multiple perspectives, for example by analysing existing research work and open issues.

First, by focusing on visual blending research, we observe that there is the pursuit of a system that is able to turn concepts into visual representations. Confalonieri et al. (2015) propose argumentation as a way to evaluate and refine the quality of blended computer icons. Xiao and Linkola (2015) present a semi-automatic system to produce visual compositions for specific meanings (e.g. *Electricity is green*). Ha and Eck (2017) train a recurrent neural network that generalises concepts and can be used for representing new concepts by interpolating between several concepts. Similarly, Karimi et al. (2018) describe a deep learning approach focused on conceptual shift and present the possibility of using it to aid humans in blend production.

In order to implement a general purpose visual blending system, a large repository of visual representations with adequate format is required. However, most available repositories only provide raster images, which demand complex computer vision techniques to be used in a blending process. On the other hand, the Emoji set fulfils both requirements due to its large conceptual coverage (2823 emoji in Emoji 11.0) and appropriateness of image format – e.g. Twemoji is composed of fully scalable vector graphics, appropriate for visual blending (Cunha et al. 2017).

The Emoji's suitability for blending is explored by Cunha et al. (2018b), who present a computational system that visually represents user-introduced concepts through visual blending. The system is aligned with research on emoji generation (Puyat 2017; Radpour and Bheda 2017) but has a different focus – it uses emoji as a mean and not as a goal.

The second version of the system includes an interactive evolutionary engine (Cunha et al. 2019), which allows the production of solutions that match the user preference. The current version of the system – the focus of this paper – brings yet another dimension into play by giving a co-creative nature to the user-system relation.

Computational approaches to co-creativity establish a collaboration between several agents, one of which is required to be artificial. It leads to a shared creative process where agents contribute to the same goal – e.g. drawing (Davis et al. 2016) or game level design (Yannakakis, Liapis, and Alexopoulos 2014). However, examples of co-creative approaches from the visual domain mostly focus on sketching (Davis et al. 2016; Karimi et al. 2018) or abstract icons (Liapis et al. 2015), and not on pictogram generation.

Despite providing evidence of value in three different fields, the analysis in terms of research interest presented in the previous paragraphs is, in our opinion, not enough to justify such a system – providing a strong argument requires putting the system in a real-world situation. In this paper, we focus on the assessment of the usefulness of a computational system for visual representation of concepts through visual blending (Cunha, Martins, and Machado 2018b). Our main contributions are: (i) the description of two user studies and (ii) an overall discussion using a multi-perspective approach on the usefulness of visual blending systems for new meaning making.

The System

This paper focuses on the assessment of the usefulness of a computational system that uses visual blending for the representation of concepts. The system uses an interactive evolutionary approach to produce visual representations for concepts introduced by the user (Cunha, Martins, and Machado 2018b; Cunha et al. 2019). As the system itself is not the focus of this paper, we will only describe it at a general level. The system integrates data from the following online open resources: (i) Twemoji 2.3; (ii) EmojiNet (Wijeratne et al. 2017); and (iii) ConceptNet (Speer and Havasi 2012).

The current version establishes a co-creative interaction between user and system. In this interaction, solutions are produced and the user is able to actively give feedback on their quality, leading to the evolution of better solutions. Briefly describing, the interaction could be said to have three agents: a solution generator (system), an evaluator (user) and an artificial evaluator (system). The latter agent is capable of selecting individuals based on its idea of quality and storing them in its own archive.

The solutions are produced using a process of visual blending, which consists in merging two input emoji (emoji A, the replacement, and emoji B, the base). Three different blend types can be used (Phillips and McQuarrie 2004): Juxtaposition (JUX) – two emoji are put side by side or one over the other; Replacement (REP) – emoji A replaces part of emoji B; and Fusion (FUS) – two emoji are merged together by exchanging parts.

Despite the existence of these three types of blend, in the previous versions of the system (Cunha, Martins, and Machado 2018b; Cunha et al. 2019) only juxtaposition and replacement were used. In this paper, the system tested already includes the fusion blend type.

Assessing Usefulness

As already mentioned, in this paper we delve into the usefulness of a computational system that uses visual blending of emoji to visually represent concepts. We identify three questions that we will address and aim to present evidence that points to possible answers. The three questions are: (Q1) Is visual blending effective in the visual representation of concepts? (Q2) Are all blend types equally adequate? (Q3) Is our system useful to users? In order to address the questions, we conducted two user studies. Study #1 focuses on Q1 and Q2, whereas study #2 mainly addresses Q3.

User Study #1: Visual Blending Effectiveness

The first user study is part of a larger study with single and double word concepts currently being conducted, in which preliminary results indicate that visual blending is not appropriate for one-word concepts, especially concrete ones (e.g. dog). As such, we chose to focus on two-word concepts. In spite of study #1 having several aspects that could be investigated, in the scope of this paper we mainly use it for two purposes: to investigate the effectiveness of visual blending in concept representation and to gather blends that represent a set of concepts, used in study #2 (see Fig. 1). Other aspects will be left for future work.

The study was conducted with 8 participants, who were asked to use the system to generate visual representations for a set of concepts. The concepts were selected from a list built by crossing a noun-noun compound dataset (Fares 2016) with a concreteness ratings dataset (Brysbaert, Warriner, and Kuperman 2014), which was divided into groups based on semantic concreteness and quantity of emoji retrieved by the system for each concept. For each participant, a set of five concepts was randomly built, aiming for variety and guaranteeing that each participant had at least one concept from each group. As the goal was to achieve maximum conceptual coverage, we decided to avoid concept repetition.

The participants were asked to use the system to evolve blends that, in their opinion, represented the concept and export the solutions which they considered the best, among the ones considered good solutions – i.e. good representations of the concept. In case no solution represented the concept, none was to be exported.

Table 1: Results in number of occurrences of a given type of blend in exported blends for each emoji quantity group – small ($\leq$5), medium (>5 and $\leq$15) and large ($\geq$25) – for juxtaposition (JUX), replacement (REP) and fusion (FUS). The "?" column refers to cases in which it was not possible to identify the type of blend and "hidden" to cases in which one of the emoji was hidden.

	JUX	**REP**	**FUS**	**?**	**hidden**
small	2	3	1	3	3
medium	8	6	0	1	2
large	4	9	2	0	2
	14	18	3	4	7

Results From a total of 40 concepts we obtained the following results: no solution was exported in 11; in 21 only one solution was exported; and in 8 more than one solution was exported. The fact that the process of visual blending was able to lead to good solutions for the majority of the concepts seems to indicate that it is a useful method for concept representation (Q1). Moreover, the results also show that the system is able to present the user with more than one good solution.

In order to further investigate the suitability of visual blending in the representation of concepts, we analysed a total of 39 blends exported by the participants in terms of blend type. The results show that juxtaposition and replacement are used in the majority of the exported blends and fusion is barely used – see Table 1 (Q2). In addition, in some cases, it was not possible to ascertain the type of blend, as one of the emoji was hidden. Another emoji hidden situation occurred in a fusion blend, in which the replacement emoji was not perceivable. We identified the cases in which one of the emoji was hidden in the blend.

User Study #2: Usefulness in Real World

The capability of a user to find a solution that visually represents a given concept with the system does not actually present much evidence of the usefulness of the system itself. For this reason, we conducted a study with the goal of comparing creative production by the user alone with results obtained with the system and assess the perception of quality by the user.

As such, we used the blends exported by the participants of study #1 – theses were considered as good visual representations. From all blends exported, we selected only one per concept, using the ones identified as the best when more than one had been exported. Then we excluded the ones in which one of the emoji was being hidden, as these could not be considered as visual blends. This resulted in a set of 22 concepts and corresponding visual representations (see Fig. 1). The set was divided into three groups, balanced in

Figure 1: Blends obtained in study #1 and used in study #2

terms of semantic concreteness. Then each group was given to participants and each concept was tested with a minimum of 15 participants. Due to participant availability, the first group of concepts was tested with 15 participants, the second with 19 and the third with 22. In total, 56 users with ages between 19 and 27 (*average* = 20.4 and *standard_deviation* = 1.6) participated in the study, all with background in graphic design. Each participant received a list of concepts and had to complete a survey for each concept. The survey was divided into two parts and was composed of five tasks. First, the participant was asked to conduct four tasks for each concept: T1 Do you understand the concept? T2 Draw the concept. T3 Describe the drawing in few words. T4 How well does the drawing represent the concept?

Tasks T1 and T4 required the participant to use a scale from 1 (not at all) to 5 (perfectly). In case the participant did not understand the concept (T1), the remaining tasks should be ignored. The participants were told to use a quick drawing style, similar to the one used in games such as "Pictionary". After conducting the four tasks for every concept, the generated blends of each concept were shown and the participant was asked to answer the following question for each concept, using the previously describe 1-5 scale: (T5) How well does the blend represent the concept?

Results The study resulted in a total of 414 concept tests – group 1 was composed of 7 concepts and was tested with 15 participants; group 2 had 7 concepts and was tested with 19; and group 3 had 8 concepts and was tested with 22. From the total of tests, 76 had to be excluded from the study due to invalid answering: in 3 no answer was given to any of the tasks; in 24 no answer was given regarding the familiarity with the concept (T1); in 40 the quality of the blend was not evaluated (T5); and in 9 a visual representation was drawn but not evaluated (T4).

In addition to these validity exclusions, for our analysis we only considered tests in which the participant reported to know the concept well or perfectly (T1 $\geq$4). We are aware that this procedure reduced the number of answers considerably but it would make no sense to analyse tests in which the concept was not known to the participant – such would invalidate the results. It is important to notice the notable difference between valid tests (V in Table 2) and valid tests in which the user understood the concept (A in Table 2). This may be due to two factors: concept complexity and participant language difficulties (the participants were not native English speakers).

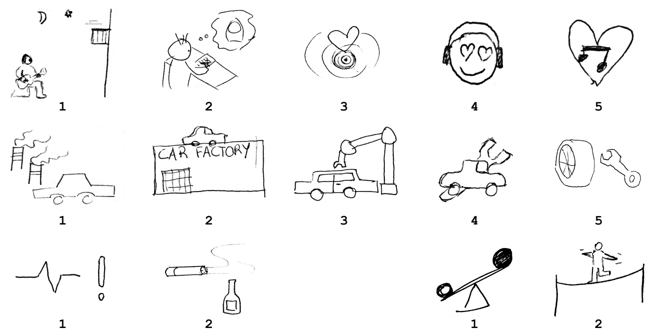

Figure 2: User drawings obtained in study #2 for *love song* (top row), *car factory* (middle row), *health risk* (bottom row, left side) and *balancing act* (bottom row, right side)

The results of the comparison between blend and drawing shown in Table 2 only consider the valid tests in which the user understood the concept (A). The cases in which no drawing was conducted are included in the value of better blend. Comparing the results from T4 (drawing) and T5 (blend) allow us to assess usefulness of the system.

In 8 out of the 9 concepts with high understanding rates – good understanding in the majority of the valid tests – the blend was considered better than the drawing by the majority of the participants (underlined in Table 2) and this majority was even absolute in 4 from these concepts. From the remaining 14 concepts, in 4 the blend was considered better by the majority of the participants, and in 2 the results obtained by the blend and the drawing were equal. Moreover, in two cases, the participant, despite knowing the concept, was not able to draw it and evaluated the blend as equal or better than good (T5$\geq$4). In contrast, the drawing was only better than the blend for the absolute majority of A in five concepts, four from which had very low understanding of the concept (less than 27% had T1$\geq$4). These results indicate that the system would be helpful to the user in 14 of the concepts, and its usefulness is particularly obvious in 8 from these concepts (36% of the 22).

When analysing the drawings made by the users, it is easy to observe how complex some of them are – e.g. drawings 1, 2 and 3 for *love song* in Fig. 2 were described by the users as "serenade", "writing a love song" and "sound waves". Yet, it is questionable whether these drawings are perceived as *love song*. Moreover, some drawings could even be more closely related to other concepts – for *car factory*, 1 could be perceived as a driving car, and 4 and 5 could be interpreted as the icon for a garage for fixing cars. In example 2, the user even included the label "car factory" to make it perceivable. In most of these examples, the blend obtained better quality

Table 2: Results of study #2 for each concept – number of tests conducted (T); mode (mo) and median ($\widetilde{x}$) for the tasks T1 (concept understanding), T4 (drawing quality) and T5 (blend quality); number of valid tests (V); number of tests analysed (A); percentage of A in which the blend was worse (B $<$ D) and better (B $>$ D) than the drawing (includes absence of drawing).

		T1		T4		T5					
	T	mo	$\widetilde{x}$	mo	$\widetilde{x}$	mo	$\widetilde{x}$	V	A	B $<$ D	B $>$ D
growth rate	15	4	4	4	4	5	5	14	10	20.0	60.0
flag carrier	15	1	2	1	2	4	3	10	1	0.0	100.0
peace accord	15	4	4	3	3	5	5	14	10	10.0	70.0
packaging product	15	5	5	3	3	4	3.5	13	9	33.3	44.4
power difficulty	15	1	2	1	1.5	3	2.5	11	3	100.0	0.0
risk disclosure	15	1	1	1	1	3	3	10	1	100.0	0.0
security house	15	3	3	3	3	5	4	11	6	33.3	50.0
cigarette market	19	5	4	3	3	4	3	17	9	33.3	44.4
failure risk	19	5	3.5	2	2	1	3	15	6	50.0	16.7
plane crash	19	5	5	5	4	5	5	13	12	25.0	50.0
vehicle operation	19	3	3	4	3	3	3	14	6	50.0	16.7
balancing act	19	5	3	3	3	1	1.5	14	7	71.4	14.3
future power	19	5	3	1	3	4	3	13	6	16.7	50.0
love song	19	5	5	5	4	5	5	16	15	6.7	40.0
car factory	22	5	5	2	2.5	5	5	20	16	6.3	75.0
health risk	22	4	4	2	2	5	4	20	18	11.1	77.8
rumor control	22	1	3	3	3	3	3	21	3	100.0	0.0
market depression	22	1	2.5	2	2	1	2	20	5	60.0	20.0
business information	22	3	3	1	2	4	4	18	2	0.0	100.0
university center	22	1	3	2	2	5	4	19	8	0.0	100.0
risk assessment	22	1	1	3	3	3	3	18	2	50.0	50.0
sugar harvest	22	1	1	1	2.5	4	4	17	3	33.3	33.3

results than the drawing. Despite this, it is interesting to see how some of the drawings are very similar to the blends (e.g. 5 of *love song*). On the other hand, the drawings 1 and 2 for *balancing act* were considered better than the blend, which shows that the system is not always capable of producing better solutions.

Discussion

The usefulness of a system can be assessed from several perspectives. In the introduction, we already addressed it using a research-interest point of view, showing the potential of the system in three different fields.

As already mentioned, our goal is not the generation of Unicode emoji. Despite this, we have shown the usefulness of the system in terms of complementing the emoji set (Cunha, Martins, and Machado 2018a), which still lacks in the coverage of several core concepts. For a list of 1509 core concepts, the system is able to produce concept-representative solutions for 1144 concepts, which is an improvement of 44.63% when compared to the results obtained by emoji set alone.

Our main goal is to produce visual representations of concepts. One question that arises is whether visual blending is suitable for concept representation (Q1). The high coverage of the core concept list is an argument in favour. However, when analysing the usage of the system by participants there are issues that point otherwise. First, the occurrence of emoji hiding – one of the emoji was partially or even totally hidden – which is an exploit of the system and does not make usage of visual blending. In study #1 hiding was observed in 7 out of the 39 exported blends. Another unfavourable result is the high occurrence of juxtaposition (Table 1), which we consider as a weak type of blending, having little advantage over a sequential positioning approach.

On the other hand, replacement as the most used type of blend is a good indication that the visual blending is beneficial. When analysing user drawings (obtained in a Pictionary-like task), users tend to draw existing objects and use a juxtaposition-based approach. Replacement often leads to metaphorical solutions, which normally require more complex reasoning from humans. As such, the system provides a quick way to present the user with solutions that require such reasoning. These topics are related to Q2 and provide an indication that different blend types have different advantages.

The two biggest advantages of the system (Q3) are that it provides the user with the possibility of choosing among different blend type solutions, often leading to more than one solution deemed good (study #1), and that it follows a multi-purpose approach, allowing the user to introduce any concept without requiring changes to the configuration or extra input data.

As far as co-creativity is concerned, one of the most used arguments in favour of such systems is the capability of fostering users creativity (Liapis et al. 2016; Cunha, Martins, and Machado 2018a; Karimi et al. 2018). The interaction with the system allows the user to evolve solutions that match his/her preferences and, at the same time, both the user and the system are constantly influencing the perception of one another, leading to novel ideas. The results obtained in study #2 provide evidence that this interaction leads to better solutions than the ones drawn by the user alone. The potential of the system is even clearer if we consider that in two cases participants who knew the concept were not able to draw it and afterwards considered the blend as a good representation. Despite these results, there is still work to be done in assessing the impact of the co-creative functionalities on the user (e.g. suggestions made by the artificial evaluator) and also in making them more adequate to the needs of the user.

Overall, we believe that we have demonstrated the usefulness of the system in terms of (i) research purposes, (ii) emoji set completeness, (iii) visual representation of concepts, and (iv) creativity aiding and fostering.

Acknowledgements

This research is partially funded by Fundação para a Ciência e Tecnologia (FCT), Portugal, under the grants SFRH/BD/120905/2016 and SFRH/BD/132728/2017. This work includes data from ConceptNet 5, which was com-

piled by the Commonsense Computing Initiative and is freely available under the Creative Commons Attribution-ShareAlike license (CC BY SA 4.0) from conceptnet.io.

References

Brysbaert, M.; Warriner, A. B.; and Kuperman, V. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods* 46(3).

Confalonieri, R.; Corneli, J.; Pease, A.; Plaza, E.; and Schorlemmer, M. 2015. Using argumentation to evaluate concept blends in combinatorial creativity. In *Proceedings of the Sixth International Conference on Computational Creativity*, 174–181.

Cunha, J. M.; Gonçalves, J.; Martins, P.; Machado, P.; and Cardoso, A. 2017. A pig, an angel and a cactus walk into a blender: A descriptive approach to visual blending. In *Proceedings of the Eighth International Conference on Computational Creativity*.

Cunha, J. M.; Lourenço, N.; Correia, J.; Martins, P.; and Machado, P. 2019. Emojinating: Evolving emoji blends. In *Proceedings of the Eighth International Conference on Computational Intelligence in Music, Sound, Art and Design*, 110–126. Springer.

Cunha, J. M.; Martins, P.; and Machado, P. 2018a. Emojinating: Representing concepts using emoji. In *Workshop Proceedings of the Twenty-Sixth International Conference on Case-Based Reasoning*.

Cunha, J. M.; Martins, P.; and Machado, P. 2018b. How shell and horn make a unicorn: Experimenting with visual blending in emoji. In *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018.*

Davis, N.; Hsiao, C.-P.; Singh, K. Y.; and Magerko, B. 2016. Co-creative drawing agent with object recognition. In *Twelfth artificial intelligence and interactive digital entertainment conference*.

Fares, M. 2016. A Dataset for Joint Noun-Noun Compound Bracketing and Interpretation. In *Proceedings of the ACL 2016 Student Research Workshop*, 72–79. Association for Computational Linguistics.

Ha, D., and Eck, D. 2017. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*.

Karimi, P.; Maher, M. L.; Grace, K.; and Davis, N. 2018. A computational model for visual conceptual blends. *IBM Journal of Research and Development*.

Liapis, A.; Hoover, A. K.; Yannakakis, G. N.; Alexopoulos, C.; and Dimaraki, E. V. 2015. Motivating visual interpretations in iconoscope: Designing a game for fostering creativity. In *Proceedings of the Foundations of Digital Games Conference*.

Liapis, A.; Yannakakis, G. N.; Alexopoulos, C.; and Lopes, P. 2016. Can computers foster human users' creativity? theory and praxis of mixed-initiative co-creativity. *Digital Culture & Education* 8(2):136–153.

Phillips, B. J., and McQuarrie, E. F. 2004. Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing theory* 4(1-2):113–136.

Puyat, M. 2017. Emotigan: Emoji art using generative adversarial networks. CS229: Machine Learning Course, Stanford University.

Radpour, D., and Bheda, V. 2017. Conditional generative adversarial networks for emoji synthesis with word embedding manipulation. *arXiv preprint arXiv:1712.04421*.

Speer, R., and Havasi, C. 2012. Representing general relational knowledge in conceptnet 5. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 3679–3686.

Wijeratne, S.; Balasuriya, L.; Sheth, A.; and Doran, D. 2017. Emojinet: An open service and api for emoji sense discovery. In *Proceedings of ICWSM-17*.

Xiao, P., and Linkola, S. 2015. Vismantic: Meaning-making with images. In *Proceedings of the Sixth International Conference on Computational Creativity, ICCC-15*.

Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2014. Mixed-initiative co-creativity. In *Proceedings of the Ninth Conference on the Foundations of Digital Games*.