HAL
open science

Would you hire me ? Selfie portrait images perception in a recruitment context

Filippo Mazza, Matthieu Perreira da Silva, Patrick Le Callet

To cite this version:

Filippo Mazza, Matthieu Perreira da Silva, Patrick Le Callet. Would you hire me ? Selfie portrait images perception in a recruitment context. SPIE HVEI XIX, Feb 2014, France. pp.90140X, 10.1117/12.2042411 . hal-00952704

HAL Id: hal-00952704

https://hal.science/hal-00952704v1

Submitted on 27 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Would you hire me? Selfie portrait images perception in a recruitment context

F. Mazza[a], M.P. Da Silva[b] and P. Le Callet[b]

[a]LUNAM, Ecole Centrale de Nantes, IRCCyN UMR CNRS 6597, Nantes, FRANCE
[b]LUNAM, Université de Nantes, IRCCyN UMR CNRS 6597, Nantes, FRANCE

ABSTRACT

Human content perception has been underlined to be important in multimedia quality evaluation. Recently aesthetic considerations have been subject of research in this field. First attempts in aesthetics took into account perceived low-level features, especially taken from photography theory. However they demonstrated to be insufficient to characterize human content perception. More recently image psychology started to be considered as higher cognitive feature impacting user perception. In this paper we follow this idea introducing social cognitive elements. Our experiments focus on the influence of different versions of portrait pictures in context where they are showed aside some completely unrelated informations; this can happen for example in social networks interactions between users, where profile pictures are present aside almost every user action. In particular, we tested this impact on resumes between professional portrait and self shot pictures. Moreover, as we run tests in crowdsourcing, we will discuss the use of this methodology for these tests. Our final aim is to analyse social biases' impact on multimedia aesthetics evaluation and how this bias influences messages that go along with pictures, as in public online platforms and social networks.

Keywords: image aesthetics, content perception, portrait images, social bias, social networks, crowdsourcing

1. INTRODUCTION

During last ten years, multimedia quality evaluation evolved from the mere technical parameters analysis to the consideration of human factors in content perception. Later on, image aesthetic appeal has been introduced to deepen the analysis. Research in image aesthetics showed good results, allowing to build automatic aesthetic assessment systems[1] evaluating low-level features. These features were mainly taken from photography literature, like contrast, composition and colors. Many researches on color as factor on human behaviours have been carried out; a critical review on color psychology is present in Whitfield's review.[2] Higher level features considering perceptual differences and technical details regarding photography - notably from camera settings - have also been proposed.[3] However, a gap is still present, as low-level computable features alone are not enough to express completely the impact of multimedia on users' experience. Moreover, there is a large variability between people in aesthetic canons.

High level human-oriented semantics are required.[4] These should take into account more complex mechanisms in human content perception. While content semantics are not new, as for example it has been used with content recognition in automatic image tagging[5] , in it's use there is more than the simple content perception. Content is not only perceived but processed by deeper mechanisms in human brain involving many factors, like memories, personal experiences, belief, and cultural background; we refer to this fact as "content cognition". First steps in taking into account this element have been done: image psychology has been suggested as a potential focus for further research by Fedorovskaya[6] and recently familiarity within image context[7] has been addressed.

In this paper we consider image psychology elements proposed by Fedorovskaya focusing on social factors as particular cognitive biases may be triggered. Cognitive biases are deeply investigated in psychology due to their impact on the decisional process. Their importance is underlined by the fact that they occur at an unconscious level, so that people are even not aware of being influenced. Between cognitive biases, an important category

HVEI XIX, 2 - 6 February 2014 San Francisco, California, United States of America.
Special session on Cognition, Emotion, and Aesthetics in Image Quality

is regarding the social elements impacting our perception; these are called social biases.[8] We think that how an image is perceived is impacted by this bias by different social clues. In this work we start to explore how social clues conveyed inside an image impact the perception of the subject depicted in the image, as the message conveyed changes radically with elements not related to the objective image quality itself. This is especially true when people are present in the image, as more elements can be understood from the context regarding the person depicted. Many times these guessed elements impact in any case our perception of the person and change the message conveyed by the content itself and any other interaction with the person in the picture. This is the case for example of profile pictures on social networks where this image will be aside every comment or action the user does. In that case, people could evaluate his/her interactions influenced also by the picture aside. In case a person is depicted in the picture, they will associate a face to a message. Another example is regarding resumes, of course in case attaching pictures is not forbidden. In fact, candidate image may influence selection process as choice may be based not only on competencies but also on many other elements conveyed by the image - i.e. if it's clearly understandable sex, race, age, beauty. As we cannot change our face traits from shot to shot, mastering photo elements would be of great value: with this knowledge we can think of biasing hidden message to convey with the picture as desired. This for example can be useful to underline different personality traits of a person depending on the context of the image. These elements are related more to social interactions than image aesthetics.

Recently selfie portraits have become really popular, both between the mass but also between public figures and celebrities. These shots are self shot pictures, usually holding the camera in front of a mirror or more frequently pointing it to the subject holding it with one hand.[9] Technically a selfie can also be shot via professional cameras with timers or remote controllers; however in this research we refer to normal selfies shot in consumer conditions: picture subject take a shot holding a consumer grade camera. Self shot pictures are not new in photography, but their use and production is recently growing hugely due to the spread of consumer cameras - now present also in most new mobile phones - and internet pervasiveness. More and more people are taking selfies in a wide range of contexts and upload them on social networks directly from hand-held devices. These pictures are often adopted as profile pictures, appearing in every interactions users do online. As this kind of pictures convey spontaneous highly variable situations, it becomes worth to see if and how these elements impacts on subject perception. This impact can be seen in the broader context of social biases. This study can be particularly important for social networks, as they can host different kind of interactions: with friends, for work or in public announcements. Different messages conveyed by the picture can affect also the message given aside it. In this work we adopted our previous example of resumes' pictures to evaluate the bias given by different picture portraits. Considered the growth of selfie pictures, we adopted this kind of portrait against professional portraits, commonly conceived more suitable in working contexts. Our research question is if a professional portraits gives more importance to a message, even unconsciously, respect a selfie one. We expect professional portraits to give more importance to message associated with them, as this kind of portraits are usually conceived as more valuable. To study this behaviour we will propose participants stimuli which message can be 'reinforced' by pictures, through our resume example previously said. Our analysis will also consider if this effect is affected by the content itself. Moreover we will investigate validity of crowdsourcing, a novel methodology already adopted in similar image aesthetic researches, for this kind of study. Next section is dedicated to experiment description, stimuli creation and crowdsourcing. Data section will follow this part, just before the conclusions and discussion of results.

2. METHODS

To evaluate picture bias on candidate choice, we set up our experiment as a resume selection process for a working position. We proposed the test as a game where we are an informatics company hiring developers. For this task participants have to select the best candidate for a proposed position, choosing between two resumes shown each time. Test adopts a pair comparison methodology. Proposed resumes are invented and have of course a picture attached, either a selfie portrait or a professional one.

To formalize the problem, three different factors are considered for our stimuli, as summarized in the table 2: the portrait typology, the resume informations and the model in the picture. Using this notation, each stimuli is noted as combination of these three. As example the stimuli having first portrait subject, professional typology

and first version of resume informations is noted A_p^{PhD}. The same, with selfie typology would be noted as A_s^{PhD}. These three factors have been controlled during our test. Pairs showed to participants are made on possible stimuli combinations regarding these factors. A total of $6*2*2 = 24$ stimuli originate from these. Some constraints have been enforced on pairs generation. We took care about which pairs to display each participant in order avoid people to get the real scope of the test, as this would have maybe biased the answers. In fact, providing to a participant different informations regarding a portrait subject, both in terms of resume or picture, would have made the participant more aware of the real purpose. To this extent, we followed these constraints:

- the same participant must not see different portrait typology for the same portrait subject;

- the same participant must not see different resume informations for the same portrait subject.

Following notation adopted, the first constraint implies that a participant can observe either X_p^{PhD} or X_s^{PhD}, and the second one imposes either X_p^{PhD} or X_p^{M}. These constraints inevitably lead to much more participants needed, as we cannot propose to evaluate all our stimuli to each participant. It has to be noted that participants can be faced instead with two portraits having same typology in one pair, as there is no problem with this configuration. Pictures have been created ad-hoc as described in next section.

Name	Versions	Notation	Description
Portrait subject	6	[A,B,...,F]	The model in the portrait image attached to the resume
Typology	2	[p,s]	The typology of the portrait image, either professional portrait (p) or selfie (s)
Resume informations	2	[PhD / M]	One of the two resume versions for each portrait

Table 1. Factors of influence in our experiment

2.1 Stimuli creation

2.1.1 Portraits

As no database like this already exists at the best of our knowledge, we constructed a small database of stimuli ad-hoc. Eight male subjects from different countries (and face traits) and aged between 23 and 40 participated in the photo shooting as models. Female portraits have not been taken due to the lack of sufficient different models.

Professional portrait pictures have been shot with professional equipment. A high grade DSLR camera and objective have been adopted. Subjects have been shot in a photo studio with controlled diffused light. They were asked to take a natural posture but supervised to get a professional portrait. White balance has been controlled and then retouched digitally, as photos have been corrected in postprocess. Contrast has been enhanced, as well as unwanted shadows removed. Minimal retouches have been manually done on details, to enhance skin tone and softness, being careful to preserve face traits. Pictures have been taken at a resolution of 16 Mpx, then manually cropped to adjust the composition and cut empty space around subject.

A similar procedure has been adopted for the selfies shots, but in a spontaneous context and with only a consumer camera as equipment. Subject have been asked to take a self portrait with a mobile phone, without the help of the photo studio lighting, as they would do in a real case. Device is equipped with 2 Mpx front camera. Although subjects were free to take shots, they had a minimal supervision in the process to have correct portraits. Postprocess have been done to resize images as previously said. Minimal modifications on brightness and contrast have been done in order to conceal unwanted shadows effects. Unrecognizable and blurred shots have been discarded.

While on selfies - but also on some professional portraits - subjects depicted have an happy or smiling expression, we adopted here more serious portraits. This has been done in order to avoid biases from facial expressions. As we are not checking if a candidate is preferred to another in general we did not adopted a standard to follow regarding glasses or hairs. We suppose that other normal elements, as long hairs versus short hairs, or glasses, give no bias in this simulated selection process. Figure 2.1.1 shows some examples of portraits shot. In the end, regarding image aesthetics, most people will agree on the fact that professional shots look much

(a)

(b)

Figure 1. Some portraits created; from top to bottom professional shots and selfies.

sharper, brighter and more defined respect selfies shots. Moreover, as social clue, professional shots convey also the sense of a higher value photo where subject put more effort and took care of the details by purpose.

We kept only six from these selecting the best professional shots in terms of face expression and posture. We chose this value to limit the number of possible combinations to be evaluated while having enough different portraits.

2.1.2 Resumes

As said before, resumes excerpts have been proposed with created portraits. To simplify this part, we told participants that we were at the end of the selection process and informations on resumes were summarized. We did this in order to have minimal resumes informations, simple enough for us to deal with - create, modify, propose in comparisons - but also for participants to evaluate. Informations proposed in the resume are candidate age and instruction level. Informations are invented on the proposed context of an informatic company seeking employees. Each created portrait has been associated with two resume's versions, both with same specialty but different degrees: one is PhD and one Master Degree. Each participant saw either one version or the other, to avoid making him aware of our setup, as specified by our previously mentioned contraints. Ages are between 24 and 31 years old, to comply with picture models' ages, and coherent with the instruction degree in the resume. Studies courses names for the degree are taken from some known study courses and specializations related to computer science. Figure 2.1.2 shows a proposed candidate selection with our resumes excerpts.

Adopting a pair comparison methodology, participants are faced with a couple of resumes each time from which they have to select either the left or the right one. As said, we proposed this as the choice of a candidate for a particular position. However, differently from a normal pair comparison configuration, we had more constraints regarding which stimuli to show to a particular observer. In fact, we designed our experiment taking care to divert participants attention from the real purpose; experiment was presented as a different task then image

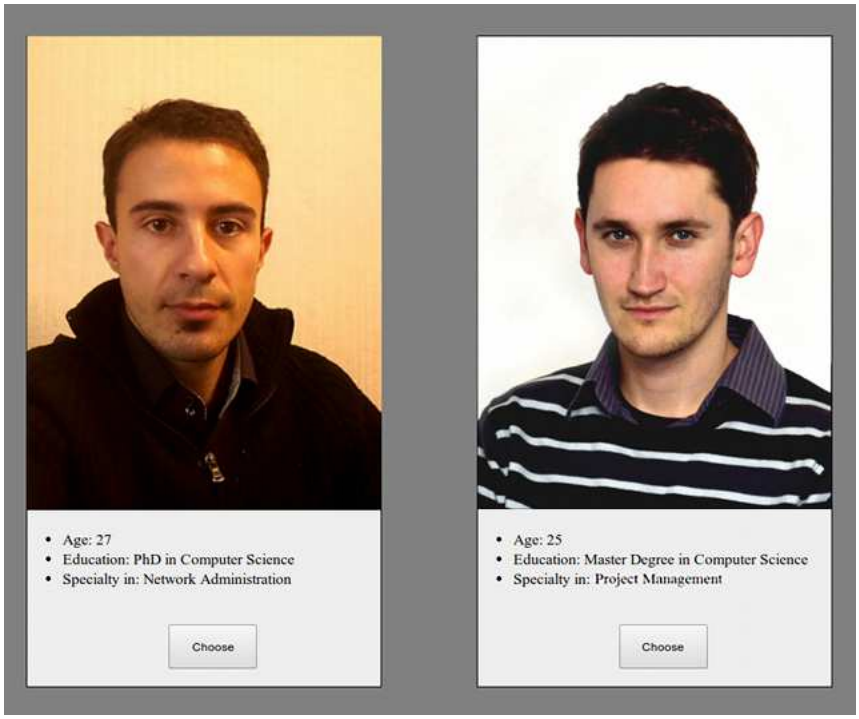


Figure 2. A comparison proposed in the experiment. Invented curriculum excerpt are visible under the pictures.

assessment. Previously formalized constraints regarding our experiment setup were taken into account during pair comparison implementation; these inevitably led us to need more participants.

Lot of data from a large panel of heterogeneous participants can help mitigating the effect of personal opinions' biases on evaluations. As our task is small and well defined - chose the best candidate between the two proposed each time - we prepared a web interface to allow people to participate remotely, exploiting the power of crowdsourcing. This allowed us to gather relatively fast a large amount of evaluations as explained in next section.

2.2 Crowdsourcing the experiment

Crowdsourcing is a recent practice adopted in many fields to outsource to a crowd of people a task commonly done in a controlled context. In particular regarding multimedia quality research field, crowdsourcing is used to collect evaluations from a crowd of people through the internet. Outsource is done through an open call usually via online commercial platforms that collect small tasks provided by employers and gather huge crowds of people willing to participate. Examples of well known platforms are Microworkers* or Amazon Mechanical Turk †. Thanks to the web, crowdsourcing can quickly provide large number of workers - in our case tests participants - at a cost cheaper than the one to conduct laboratory experiments. This experiment could have been run in a controlled laboratory environment, however with this strategy we can have a much higher number of participants, as it is much less expensive and faster to run the test. Times and costs are cut by the fact that participants need no supervision and can participate remotely without moving from home or office. Moreover, crowdsourcing can provide heterogeneity between participants, as the crowd is spread all over the world. This demographic diversity of users, mostly in terms of geographical distributions but also of ages and education, would have been much more difficult to achieve with a normal laboratory setup. This element can be valuable in our research as social biases can be influenced by participants' social background. In our experiment, to gather data that can be considered later, we included a small anonymous questionnaire asking for participant nationality, job and education level.

*www.microworkers.com

†https://www.mturk.com/mturk/

Crowdsourcing has already been exploited in multimedia quality experiments many times, not only for still images[10] but also for video assessments.[11] However, this methodology still poses many challenges.[12] In fact, while in laboratory participants have a well defined environment and follow precise rules, remotely the crowd is anonymous and uncontrolled under many aspects: the context and hardware adopted to participate can vary greatly from user to user and some environments can be unsuitable to participate. This is the case for example of users participating with small screens unable to display our contents. Used devices are often out of control; some parameters can be detected and checked but for others we need to rely on other elements, as a posteriori controls on results useful to detect outliers. Moreover many subjects can provide incorrect results because they lack supervision or they do not understand the task to be done.

2.2.1 Our platform

To deal with all these problems, expanding our possibilities and propose our experiment, a new crowdsourcing software framework has been developed. This framework has the purpose of instructing users regarding the experiment, proposing the experiment itself and monitoring the users. While some platforms are already available for this purpose, like QualityCrowd[13] some features are missing and a modular platform as the one created allows greater flexibility. A wide range of features has been designed allowing future reuse for other tests and fast customizations. With this framework, we were able to have a greater control on many aspects, notably experiment work flow. In order to have a better understanding of our participants we proposed a small anonymous demographic questionnaire before the actual experiment. Information asked were the age, the Country of residence, the education level and if they previously already participated in similar test for research purposes. After the questionnaire we provided clear instructions in the first page, explaining to select a candidate each resume pair provided. Then, at the first pair provided, we provided an help screen highlighting interface buttons to chose a candidate. This screen disappears after clicking anywhere on the screen, as mentioned also in this help screen itself.

In our platform we implemented also a set of functions to allow the permutation of image metadata and captions. These are important for our experiment regarding resumes and pictures, as we need to handle metadata in profile shots. For example, we can track which elements were present in pictures proposed to a participant. As said, crowd is uncontrolled and unreliable; this is why functions to check participants reliability have also been added. Mainly this goal has been fulfilled adding content questions regarding stimuli proposed, sometimes called challenges or 'honey pots'[14] . These questions aim to check if users are really aware of what they are doing or they do proposed task as fast as possible just to earn money. In our case we proposed questions in the context of resume selection plot we designed; participants are told from time to time during the evaluations that the personnel chief is worried about some some informations about chosen candidate. Randomly, participant is asked if chosen subject was wearing glasses, or if he had long hair, or regarding his resume informations. Answers are stored with metadata functions previously told and shuffled with other stimuli answers, so that participants are asked randomly something different every time. Answers are checked on the spot and an error message is displayed to the user as feedback on his answer; again the answer is given taking advantage of proposed game context, saying that the answer 'did not convince' the chief personnel. Participants' reliability has been controlled also implicitly. Answers statistics and response times have been monitored to identify outliers; avoiding people selecting repeatedly too many times stimuli either right or left - just to speed up the experiment.

A problem detected with crowdsourcing regards the time limit participants are tolerating before quitting prematurely the experiment. Previous researches done in the field with similar methodology analyse participants' withdrawal behaviour, and their results suggested us a maximum limit of around five minutes.[12] We adopted this duration for designing our experiment; to achieve this with the honey pots provided in our configuration, we chose to limit to 15 the number of evaluations proposed each participant, considering a maximum of ten seconds needed for each choice. However this consideration does not guarantee that all participants complete the test; participants' withdrawal can make problems as some pairs are left unevaluated and at the end of the whole experiment have less quotes than the others. However, we took this problem into account while making our experiment and checked that every possible image pair in our setup has been evaluated by at least 20 people. With the considerations done, we decided to pay 0.50 each participation, looking also at current prices found in Microworkers for similar survey tasks and experiment durations. Before launching the experiment on a broad scale, we started a smaller one with fewer subjects to test our framework again. Crowdsourcing platform has

been targeted to allow users from all over the world, taking care of limiting acceptance speed. In fact time zones difference must be taken into account in order to avoid accepting only people from a particular Country, active at the moment the experiment is launched.[15]

Crowdsourcing the experiment worldwide imposes some challenges also regarding huge diversity of participants' devices and connections. Without taking care about these details, test can be impaired by different technical factors, i. e. wrong visualizations on some displays or network impairments due to poor connections. To mitigate these problems, all pictures have finally been resized to suit our online resume scenario; they have to fit a normal screen in a pair comparison methodology, that is to say two images must fit in screen side by side. To be sure that images would have not resizing on participants' browsers, we resized them to 400 x 533 px, considering internet users screen resolutions as done in.[11] At present more than half internet users have a screen resolution equal or bigger than 1024*768px.[16] Internet connection speed is another factor to consider.[17] The best solution is to reduce needed bandwidth as much as possible compatibly with our stimuli: files size must be kept as small as possible. To shrink file size we converted pictures to JPEG file format, setting quality to 85%. This quality setting was found as the best tradeoff allowing small filesize and minimal almost unnoticeable compression artifacts. However we can't be sure that stimuli will be presented completely unaffected by network delays or other network impairments; we will monitor these effects looking at answers times previously mentioned.

3. DATA ANALYSIS

Crowdsourcing methodology allowed us to gather data from 978 participants in 6 days, from 47 different Countries worldwide. Regarding their demographics, half of them are aged between 20 and 30 years old; majority of participants (74%) has a Bachelor or Master Degree and around 20% of them has a high school diploma. Data analysis on these results will be carried out in our future works regarding crowdsourcing. Here will focus on participants reliability.

Analysing participant answers to content questions data, we discovered that around half of participants answered wrong two or more "honey pots" on the five we proposed in total per participant. While the best scenario would be that no error is made on these content questions, we considered one error to be normal due to memory and fatigue influence. For this reason we removed from our first analysis users that made more than one error. Interestingly, from content questions asked it emerged that participants made more errors regarding resumes informations than resumes pictures - around 40% versus 60%. We underline that questions were prepared by our framework, randomly assigned and uniformly distributed on pairs informations, both the resume and the picture. As participants where unaware that test regarded portraits differences, we expected them to be more focused on resumes informations. This result may point to the fact that people are more likely to pay attention to or remember better profile pictures even in contexts where it should be of no importance - as in resumes selection.

Regarding participations, even simplifying test at maximum and making it last at maximum 5 minutes, around 17% of participants quit the experiment before the regular end (15 pairs to evaluate). Deeper analysis investigating this withdrawal behaviour can be carried out and it is part of our future works specific on crowdsourcing. To check evaluation reliability we analysed also evaluation timings. Measurements have been done with our crowdsourcing system, monitoring page loading times and evaluations timestamps. We found that 12% of evaluations have been given before stimuli page was fully loaded in participants' web browsers. This behaviour can be due to poor network connections, slowing stimuli download from our platform, or to unreliable participants that try to give evaluations as fast as possible just to finish earlier the test. However, as both cases cannot be considered reliable evaluations, we excluded these from successive data analysis.

3.1 Analysis of influent factors

Many analysis are possible with our proposed setup. Resume informations, portrait typology and demographic informations can be considered as factor of influence in candidate choice. In this work we focus on the influence of portrait typology, either selfie or professional portrait. To this extent we check if the choice between two different typologies is statistically significant, all the other conditions being equal, for every possible face-resume combination. In other words we will check if candidate choice "can be ascribed to chance" respect portrait

typology: this can be done with a Barnard's exact test.[18] This non-parametric test was claimed more powerful then Fisher's exact test on contingency tables 2x2.[19] This statistical test checks the independence of rows and columns in contingency tables, returning a p-value under the null hypothesis of pure chance. On 95% confidence a p value< 0.05 implies significant difference between the two test scenarios. The only assumption required to be respected is that responses are independent between them; this hypothesis is not violated in our case as every evaluation is independent between different participants as well as between different comparisons.

Running Barnard test on all pairs between the two conditions, as said either professional portrait or selfie, we obtain in total a out of N significant differences. N in our case is given by the number of possible combinations after respecting previously mentioned constraint regarding stimuli - here equal to $10 * 24$. To test if factor under test is a factor of influence on these results, we conducted a permutation test and checked if significance ratio a/N is sufficiently large. Algorithm followed is the same as in Li's work.[20] This methodology implies the random permutation of some participants quotes between the two conditions and recomputing Barnard tests for all pairs as said before. As in a Monte Carlo simulation, with a sufficiently large number of permutations this distribution can be estimated. Our simulation took into account 1000 iterations. Results shows that significant ratio before permutation ($a = 23$) belongs 95th percentile, pointing out that portrait typology is an influent factor even if not for all the possible combinations.

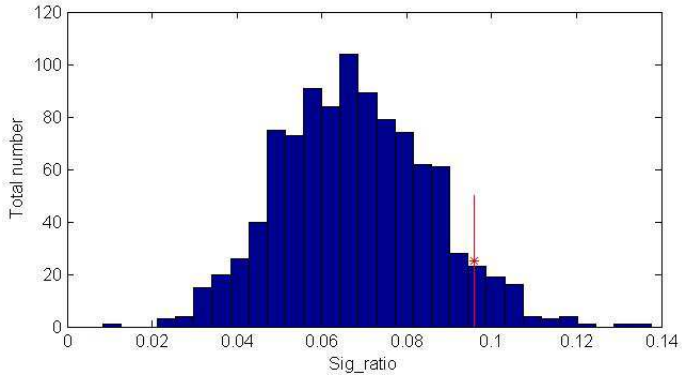
```
Data: Raw comparisons, influence factor to test, Loop_num number of loops
Result: Vector Sig_ratio containing number of significant pairs every iteration
Sig_ratio ←
initialization;
Group_1, Group_2 ←
Divide pairs in two groups based on condition to study
for n ← Loop_num do
    foreach possible evaluated pair e do
        if n ≠ 1 then
            randomly swap m quotes between the two conditions;
        end
        p(n) ←
        Barnard's test p-value;
        if p(n) <0.05 then
            Sig_ratio(n) ← Sig_ratio(n) + 1
        end
    end
    Sig_ratio(n) ← Sig_ratio(n)/N
end
```

Algorithm 1: Permutation test algorithm

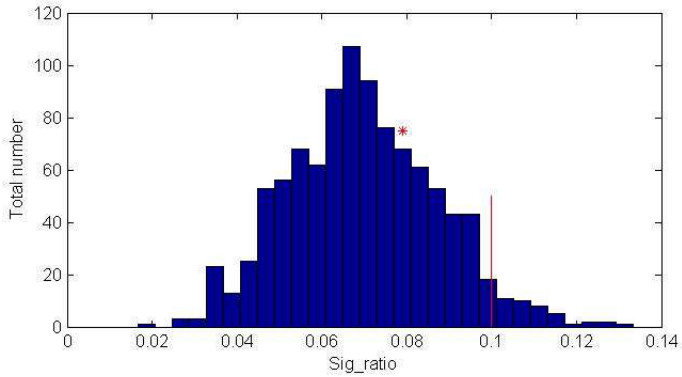
As previously said, participants with more than one honey pot mistaken have been removed from first analysis as considered unreliable. To evaluate our strategy we reintroduced those in our dataset and repeated the analysis as before. The final distribution of significant comparisons reveals we are now slightly more than the percentile of 55, indicating no evidence of portrait typology influence. This fact tend to depict those evaluations as noise in our analysis.

4. RESULTS AND DISCUSSION

In this paper we started to analyze image psychology biases given by image contents. We analyzed the bias given by the difference of self shots and professional portraits through a simulated resume selection for a work profile. We conducted the experiment online through a crowdsourcing platform build ad hoc, proposing a pair comparison experiment showing invented resumes. Crowdsourcing methodology has been evaluated outlining some problems and adopting control strategies; outlier participants have been screened based on answer timings analysis and content questions during the experiment. Factors of influence have been outlined and controlled; in



(a)



(b)

Figure 3. Permutation test results; significant ratio distribution considering data screening (a) and without data screening (b). Red line indicates 95th percentile and a star indicates significant ratio before permutations.

this work we focused on portrait typology. Barnard exact test has been adopted to check statistical significance of this factor for each pair. Permutation tests have been run to check the overall influence on all the stimuli proposed. Evidence of preference for professional portraits has been shown to be an influent factor for part of proposed stimuli. Our research also outlined that crowdsourcing is a suitable strategy for this kind of picture evaluation, although a careful planned experiment is needed as well as reliability measures. Many analysis are still possible with the data gathered; these concern demographic data collected with our pre-test questionnaire proposed to participants but also analysis regarding the impact of resume itself. These deeper analysis is part of our future works.

However, only this data is not sufficient to derive a conclusion and deeper investigation is needed. All these elements underlined that researching social biases impact for multimedia quality evaluation is interesting from multiple perspectives, as if better understood it can be considered for building improved quality evaluation models. Moreover mastering those biases is important especially for evaluations done remotely via internet: nowadays social networks are a huge resource and their social engagement features - e.g. "Likes", "Share" or friends suggested elements - are always present. How this factor can be exploited and/or corrected has not been dealt at the moment in our knowledge. Furthermore from a practical point of view it can open the path to further research in product advertising as it can influence social impact.

REFERENCES

[1] Datta, R. and Wang, J., "ACQUINE: aesthetic quality inference engine-real-time automatic rating of photo aesthetics," *Proceedings of the international conference on . . .* , 1–4 (2010).

[2] Whitfield, T. W. and Wiltshire, T. J., "Color psychology: a critical review.," *Genetic, social, and general psychology monographs* **116**, 385–411 (Nov. 1990).

[3] Ke, Y., Tang, X., and Jing, F., "The Design of High-Level Features for Photo Quality Assessment," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)* **1**, 419–426.

[4] Datta, R., Li, J., and Wang, J., "Algorithmic inferencing of aesthetics and emotion in natural images: An exposition," *Image Processing, 2008. ICIP 2008. ...* , 8–11 (2008).

[5] Fu, H., *Semantic Image Understanding : From Pixel to Word*, PhD thesis (2012).

[6] Fedorovskaya, E. a. and De Ridder, H., "Subjective matters: from image quality to image psychology," in [*SPIE Proceedings Vol. 8651*], Rogowitz, B. E., Pappas, T. N., and de Ridder, H., eds., **8651**, 86510O–86510O–11 (Mar. 2013).

[7] Chu, S. L., Fedorovskaya, E., Quek, F., and Snyder, J., "The effect of familiarity on perceived interestingness of images," **8651**, 86511C–86511C–12 (Mar. 2013).

[8] Wang, C. A., Zhang, X. M., and Hann, I.-h., "Social Bias in Online Product Ratings : A Quasi-Experimental Analysis," *WISE 2010, St. Louis, MO* (February 2010), 1–35 (2010).

[9] Krause, J., [*Photo Idea Index*], HOW Books (2005).

[10] Ribeiro, F., "Crowdsourcing subjective image quality evaluation," *Image Processing (ICIP), ...* , 3158–3161 (2011).

[11] Gardlo, B., *Quality of Experience Evaluation Methodology via Crowdsourcing*, doctoral thesis, University of Zilina (2012).

[12] Keimel, C., Habigt, J., and Diepold, K., "Challenges in crowd-based video quality assessment," *Quality of Multimedia ...* , 13–18 (2012).

[13] Horch, C., Keimel, C., and Diepold, K., "QualityCrowd. Crowdsourcing for Subjective Video Quality Tests," tech. rep., Technische Universität München (2011).

[14] Kazai, G., Kamps, J., and Milic-Frayling, N., "An analysis of human factors and label accuracy in crowdsourcing relevance judgments," *Information Retrieval* **16**, 138–178 (July 2012).

[15] Hirth, M., Hoß feld, T., and Tran-Gia, P., "Human Cloud as Emerging Internet Application-Anatomy of the Microworkers Crowdsourcing Platform," *... of Computer Science, Am Hubland, D- ...* (February) (2011).

[16] NetMarketShare, "Internet users screen resolutions statistics," (2013).

[17] Hoß feld, T., "Modeling YouTube QoE based on Crowdsourcing and Laboratory User Studies," tech. rep., COST QUALINET STSM report (2011).

[18] Barnard, G. A., "Significance Tests for 22 Tables," *Biometrika* **34**(1/2), 123–138 (1947).

[19] Barnard, G. A., "A New Test for 2 2 Tables," *Nature* **156**, 177–177 (Aug. 1945).

[20] Li, J., Barkowsky, M., and Callet, P. L., "Subjective assessment methodology for preference of experience in 3DTV," in [*IVMSP Workshop, 2013 IEEE ...*], (2013).