WHAT DO WE FEEL ABOUT ARCHETYPES: SELF-REPORTS AND PHYSIOLOGICAL SIGNALS

Huang-Ming Chang [a, b], Leonid Ivonin [a, b], Marta Diaz [b], Andreu Catala [b], Wei Chen [a], and Matthias Rauterberg [a]

[a] Dept. Industrial Design, Eindhoven University of Technology
[b] CETpD Research Center, Universitat Politècnica de Catalunya

ABSTRACT

In research on emotion, presenting affective stimuli has been believed to be an effective and reliable technique for emotion elicitation. Instead of collecting stimuli for pre-defined emotions, we propose to develop stimuli based on their symbolic meanings. We adopted archetypal symbolism as a standard to edit eight movie clips of archetypes as a new set of affective stimuli. These stimuli were used in an experiment for emotion elicitation. Participants' emotional responses toward these stimuli of archetypes were measured by the self-report technique and the physiological measurement. The results of linear discriminant analysis show that physiological measurement is more robust than the self-report techniques in recognizing emotions toward stimuli of archetypes. However, it is still unclear which technique reflects the ground truth of human emotion. We discuss alternative implications of these results, and provide more research questions for future studies on emotion recognition and model development.

Index Terms— Emotion, affective stimuli, self-reports, physiological signals, archetype

1. INTRODUCTION

Human emotion has been a tough topic for scientists for decades. Since emotion is a psychological phenomenon that cannot be directly captured, research on emotion in the laboratory is usually done by means of emotion elicitation and emotion recognition. Researchers then are able to build models of emotion through interpreting the correlation between the input information and the output representation. Numbers of effective emotion elicitation techniques have been developed, e.g. hypnosis, affective guided imagery, and affective stimuli [1]. Hypnosis and imagery have high demands on experimenters' experience, and the procedures are time-consuming. In contrast, presenting affective stimuli is relatively simple and straightforward. More importantly, presenting affective stimuli is more reliable to reproduce on huge numbers of subjects, which allows researchers to collect data more efficiently. For this reason, presenting affective stimuli is considered to be a good technique for emotion elicitation under laboratory settings [2]. Although the procedure of this technique is relatively simple, the selection of the affective stimuli and the measurement of emotional responses are relatively crucial.

Traditionally, researchers develop affective stimuli for specific emotions, which might lead to some validity problem. In light of this, we propose to develop affective stimuli based on the symbolic meaning of their contents. Consequently, archetypal symbolism was adopted as the standard for developing new affective stimuli by analyzing symbolic meanings. To further determine the emotional responses toward these stimuli, an experiment was conducted by using two typical approaches of emotion recognition: self-report techniques and physiological measurement, which represent different aspects of human emotions. The data of the self-report technique depicts explicit emotions whereas the physiological data attribute to the implicit aspect of emotions. The main goal of this study is to compare the robustness of these two different techniques. This raises a further question: which aspect is the ground truth of human emotion? As a preliminary study in this new direction, we provide alternative interpretations based on the results of our experiment. In the final section, more new research questions for future work are discussed.

2. CONTENT-CENTERED STRATEGY

Traditionally, this selection process is subject to the pre-defined model of emotion. First, researchers adopt a hypothetical model that is expected to represent most of the common emotions in daily lives, and consequently select affective stimuli for these pre-defined emotions. Next, researchers recruit subjects to report their emotions toward these stimuli by using questionnaires [3]. Finally, the emotional qualities of the stimuli can be generalized by means of statistical analysis. Thus, these validated stimuli can be used in other studies to elicit target emotions. Eliciting specific emotions by presenting affective stimuli relies on an underlying premise, that the emotional quality embedded in the given stimuli is confirmed. The whole process of validation seems to be methodologically reasonable. However, there is no reliable standard for researchers to select potential stimuli before the validation process. However, this leads to a dilemma: how do researchers determine the emotional quality of each stimulus in the selection process?

Considering the selection process, it is reasonable for researchers to search for stimuli that better fit into their measurement to enhance the performance of their hypothetical model of emotion recognition. This measurement-centered strategy is particularly pertinent to research that aims at studying explicit emotions. Researchers tend to select stimuli embedded with commonly-known emotional qualities, e.g. pictures of happiness, sad music, or videos of fear, regardless the content of the stimuli. Although this would facilitate the validation of the pre-defined model and enhance the accuracy of the recognition on particular emotions, researchers might tend to filter out those stimuli embedded with emotions that are not covered by the pre-defined model. This strategy might lead to a biased selection process, and cause some limits, e.g. the *face validity* problem [4]. If this selection process only concerns how well the stimuli fit into the pre-defined model, researchers might lose the changes to discover new findings from the unselected affective stimuli.

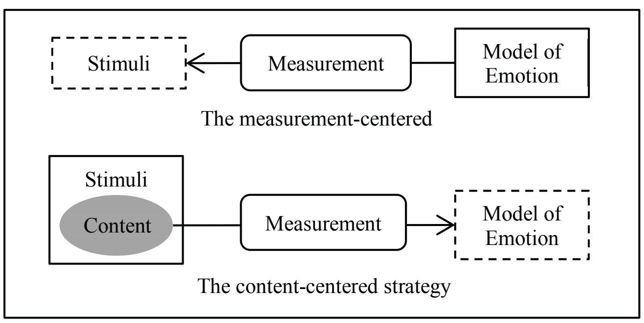

Fig. 1. The comparison between the measurement-centered strategy and the content-centered strategy.

Different from face validity, *construct validity* requires correct answers to not only “how to measure” but also “what the dimension we think we are measuring is”. Therefore, it requires more than one set of indicators to examine the procedure [4]. Thus, the selection process should not solely focus on how and what the emotions are to be measured, but also needs to investigate the correlation between the contents and the emotional qualities of the stimuli. The measurement-centered strategy looks for stimuli that fit its measurement based on a hypothetical model of emotion. In which case, the aim of this approach is to validate the pre-defined model, whereas the content-centered strategy develops stimuli based on the content and investigates the unknown emotions toward these normed contents (see figure 1). The latter approach does not set a pre-defined model in the selection process, but looks for new findings from the correlation between contents and emotional responses.

3. ARCHETYPAL SYMBOLISM

Recent years has seen a new trend of research on emotions. Appraisal theories [5] provides a different perspective of human emotion. It is argued that, while one is being chased by a tiger, it is not the ‘physical’ tiger that causes the fear, but the appraisal of this situation—being chased by a tiger—triggers this emotion of fear. Barrett [6] further propose an epistemological research question: how emotions become real? By saying this means that a physical stimuli itself is emotionally neutral (i.e. emotions are not real yet). Only when the stimuli are appraised by human, the symbolic meaning of the stimuli then elicits emotions (i.e. emotions become real). It is also argued that symbols play the role connecting the physical world and the psychological state toward it [7]. Therefore, it is humans’ appraisal of symbolic contents that makes emotional become real.

Exploring symbolic contents as affective stimuli that are universally valid is a challenging task. It is believed that humans share a kind of knowledge for emotion appraisals in a social way [6]. Nevertheless, there are still some symbolic contents that are found to be identical across time and space. Inspired by Jung’s theory of the collective unconsciousness and archetypes [8], archetypal symbolism is a discipline that investigates common symbolic contents in different cultures [9]. Jung argues that all humans unconsciously share identical patterns of thinking and behavior across time and space, which is called the collective unconsciousness. These patterns can be manifested in similar symbolic contents in dream, myth, and ancient artwork. Jung [10] further claims that archetypes are the fundamental components of all human experiences in their unconsciousness. It is also argued that archetypes are a kind of unconscious knowledge that defines how humans symbolically experience the physical world [11]. In this regard, archetypal symbolism provides a fruitful resource for studying common understandings about how people experience the world. We applied archetypal symbolism as the standard to investigate symbolic meanings of the content of affective stimuli. We cooperated with experts in the field of archetypal symbolism to develop a new set of affective stimuli based on the theory of Jung [8] and other related works [12]. Commercial movies were selected as the main recourse of our stimuli because movies often contain plenty of symbolic meanings, and it is easier for participants to get more immersed during the experiment.

4. EMOTION RECOGNITION TECHNIQUES

Except emotion elicitation, the other essential issue is emotion recognition. Emotional responses can be measured in at least three different manners: self-reports, physiological reactivity, and behavioral acts [13]. These techniques can generally be classified into subjective and objective techniques. Subjective techniques rely on self-reflection about psychological states, such as questionnaires and interviews, whereas the objective techniques are observational data from behaviors, including facial expressions, performance on a task, gestures, body stance or physiological measurement. When using stimuli presentation techniques in laboratories, researchers often apply Self-Assessment Manikin (SAM) [3] as a subjective technique , and physiological measurement as the objective technique.

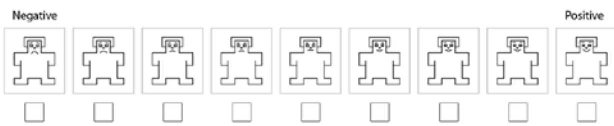

Fig. 2. The valence dimension in the SAM scale, adopted from [3].

SAM is broadly used by researchers who are interested in emotions, and becomes a very common standard of self-report technique for emotions. The dimensional model to represent human emotions is applied. By giving a score from one to nine for each of the three dimensions—valence, arousal, and dominance, emotions can be plotted into a three-dimension affective space (see figure 2 as an example for the dimension of valence). It is claimed that these three dimensions are capable of representing most of human emotions that are commonly known. The other technique directly monitors bodily changes in physiological states, which are driven by the autonomic nervous system [14], including heart rate, skin conductance, skin temperature, respiration rate, etc. Furthermore, in a previous study we also found that heart rate can be used to identify visual and audio stimuli of the archetype 'self' [15]. These two techniques are used to reflect human emotions in two different ways: explicit and implicit. Although both techniques have been well developed and are broadly used in experiments, the results of these two techniques cannot be completely mapped. It is still unclear which technique reflects the ground truth of human emotions. Therefore, it is reasonable to adopt both techniques in order to have a more comprehensive understanding on emotions toward stimuli with normed contents.

5. EXPERIMENT

An experiment was conducted to determine subjects' emotional responses toward the stimuli of archetypes. As a preliminary study, we prepared eight categories of stimuli of archetypes, including anima, animus, hero's departure, hero's initiation, hero's return, mentor, mother, and shadow. One short movie clip (approximately five minutes) for each category was edited from commercial movies. Because of this, the interpretation of this experiment has its limit. As only one stimulus was collected for each category, the results would lack statistical power to claim that each category of archetypal content induces a unique pattern of emotional responses. Although all the clips were verified by experts in archetypal symbolism, more stimuli for each archetype are required for future studies. In the present study, we are more focused on the comparison between self-report data and physiological data instead of looking for unique emotional responses for each archetype.

5.1. Participants

Twenty five healthy people were recruited for the experiment, consisting of 12 women and 13 men, Most of them were graduate students. 11 participants were from Europe, 10 participants were from Asia, 3 participants were from Middle East and one participant was from South America. The average age for the women was 23.0 years (Standard deviation = 1.9) and for the men 25.4 years (Standard deviation = 4.5). The limit of our study is that we applied convenience sampling for participants from the campus area of Eindhoven University of Technology. For other age groups, it still needs further investigation.

5.2. Procedure

The experiment followed a within-subject and double-blind test design. Only one participant took part in each session, and each participant viewed all eight film clips in a random order. The procedure was as follows: after reading and signing a consent form, the participant was asked to put on physiological sensors and to sit comfortably in the laboratory. After all the setting of the equipment was ready, an instruction with a sample film clip was presented to the subject. The introduction did not disclose any information about the experimental film clips. Meanwhile, the recording of the physiological signals was started throughout the session. Before presenting each experimental film clip, a short video demonstrating a breathing pattern was shown to guide the participant to adjust her respiration rate to a within-subject baseline. Immediately after each clip, the participant was asked to provide her emotional responses with a computer-based SAM interface. This iteration is repeated until the participant had viewed all eight film clips.

5.3. Results

Linear discriminant analysis (LDA) was performed separately with two datasets: the self-report data and the physiological data. The category of affective stimuli is assigned as grouping variables. The data of emotional responses were fed into LDA as independent variables. LDA allows us to observe how well these stimuli can be differentiated by the given data of emotional responses. Only cross-validated results are reported here because they indicate the capabilities of the model to deal with an unknown data sample. We proceed to the interpretation of the outcome of the classification. The model obtained from the SAM scale featured a classification rate of 28.5%, whereas the other model obtained from physiological data achieved a classification rate up to 79.5%.

Table 1
The confusion matrix of the model obtained from the discriminate analysis on the SAM scale [count (in %)]. The shades indicate the highest classification rate for each archetype. The bolded are correct predictions for each archetype.

Category	Predicted Group Membership (Self-Reports)							
	Anma	Anmu	Dep	Ini	Ret	Men	Mot	Sha
Anima	**4 (16%)**	3 (12%)	1 (4%)	4 (16%)	4 (16%)	6 (24%)	1 (4%)	2 (8%)
Animus	1 (4%)	**13 (52%)**	1 (4%)	3 (12%)	0 (0%)	1 (4%)	2 (8%)	4 (16%)
Departure	2 (8%)	0 (0%)	**10 (40%)**	1 (4%)	1 (4%)	3 (12%)	7 (28%)	1 (4%)
Initiation	3 (12%)	6 (24%)	3 (12%)	**6 (24%)**	4 (16%)	2 (8%)	0 (0%)	1 (4%)
Return	3 (12%)	4 (16%)	3 (12%)	3 (12%)	**5 (20%)**	4 (16%)	2 (8%)	1 (4%)
Mentor	0 (0%)	0 (0%)	4 (16%)	2 (8%)	6 (24%)	**11 (44%)**	2 (8%)	0 (0%)
Mother	4 (16%)	4 (16%)	6 (24%)	0 (0%)	0 (0%)	3 (12%)	**8 (32%)**	0 (0%)
Shadow	3 (12%)	10 (40%)	2 (8%)	2 (8%)	4 (16%)	2 (8%)	2 (8%)	**0 (0%)**

Notes. 28.5% of cross-validated grouped cases correctly classified. (Anma: anima, Anmu: animus, Dep: hero's departure; Ini: hero's initiation; Ret: hero's return, Men: mentor, Mot: mother, Sha: shadow)

Table 2
The confusion matrix of the model obtained from the discriminate analysis on the physiological signals, including heart rate, skin conductance, respiration rate, and skin temperature [count (in %)]. The shades indicate the highest classification rate for each archetype. The bolded are correct predictions for each archetype.

Category	Predicted Group Membership (Physiological Data)							
	Ani	Anm	Dep	Ini	Ret	Men	Mot	Sha
Anima	**23 (92%)**	1 (4%)	0 (0%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	0 (0%)
Animus	0 (0%)	**19 (76%)**	0 (0%)	2 (8%)	2 (8%)	1 (4%)	1 (4%)	0 (0%)
Departure	1 (4%)	0 (0%)	**21 (84%)**	0 (0%)	0 (0%)	0 (0%)	2 (8%)	1 (4%)
Initiation	0 (0%)	0 (0%)	0 (0%)	**22 (88%)**	1 (4%)	0 (0%)	0 (0%)	2 (8%)
Return	2 (8%)	1 (4%)	1 (4%)	1 (4%)	**17 (68%)**	1 (4%)	1 (4%)	1 (4%)
Mentor	2 (8%)	1 (4%)	0 (0%)	0 (0%)	3 (12%)	**17 (68%)**	2 (8%)	0 (0%)
Mother	2 (8%)	2 (8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**20 (80%)**	1 (4%)
Shadow	0 (0%)	0 (0%)	0 (0%)	2 (8%)	3 (12%)	0 (0%)	0 (0%)	**20 (80%)**

Notes. 79.5% of cross-validated grouped cases correctly classified. (Anma: anima, Anmu: animus, Dep: hero's departure; Ini: hero's initiation; Ret: hero's return, Men: mentor, Mot: mother, Sha: shadow)

Then we look into the classification rate of the two techniques for each stimulus. As can be seen table 1 and 2, we provide the confusion matrices obtained from LDA for the two datasets. In these matrices, each row demonstrates how well the obtained model can predict the category of the given stimulus. Only the correct predictions have statistical meanings for further discussion. Consequently, the diagonal cells in the two tables (top-left to bottom right) indicate the classification rate of correct predictions, whereas other cells show the classification rate of wrong predictions.

For the self-report data (see table 1), three archetypes cannot be correctly classified: anima, the hero's initiation, and shadow. Nevertheless, the stimuli of all the archetypes can be correctly classified based on physiological data (see table 2). For the self-report data, the highest classification rate of correct prediction is 52% (animus); the lowest is 0% (shadow). For physiological data, the highest classification rate of correct prediction is 92% (anima); the lowest is 68% (the hero's return and mentor).

6. DISCUSSION AND CONCLUSION

The first hypothesis of our experiment states that emotional responses induced by all kinds of affective stimuli can be measured by SAM (H1). Our results show that emotional responses toward the stimuli of archetypes are poorly represented by the three dimensions (28.5% of accuracy), which are substantially lower than the results of past studies using the same scale (at least 80%). Therefore, H1 is not supported. We proceed to define our auxiliary hypothesis (H2), assuming that physiological measurement can correctly capture the emotional responses. The results from the physiological data are similar to related studies (79.5% of accuracy), providing sufficient evidence that physiological signals we adopted can be used to differentiate the archetypal movie clips that were viewed (H2 is supported). In summary, the results show that the robustness of physiological data is considerably better than the self-report data collected from the SAM scale in differentiating one category from other categories. The implication of these results probably can be associated with Jung's theory of the collective unconsciousness, which suggests that emotions elicited by archetypal contents should be unconscious. However, the evidence is still insufficient to justify either self-report data or physiological data stands for the ground truth of human emotions. It is too bold to draw the conclusion that physiological signals can capture the unconscious emotions while the self-report technique cannot.

The discussion of our results can be twofold, depending on whether the elicited emotions are conscious or unconscious. First, if we assume that the emotions elicited by our stimuli are conscious, the low classification rate of SAM would lead to the suggestion that more dimensions are required to have a more comprehensive representation for the emotions elicited by archetypal contents. This might be because the current dimensional model was over simplified down to three dimensions, only focusing on prominent emotions that are commonly known, but not comprehensive enough to represent some trivial emotions. Another possibility is that the stimuli of some trivial emotions were filtered out at the selection process, which eliminated the opportunity to find new dimensions. As we demonstrated in the present study, focusing on the content of stimuli might remedy the validity issue of the selection process. On the other hand, our study also confirms the robustness of physiological signals in emotion recognition. Further analyses are required to identify the correlation between the pattern of physiological changes and the presented affective stimuli of archetype, e.g. what is the physiological pattern while a person feels like being a hero?

The other possible condition of the results is that the elicited emotions are unconscious. As a traditional self-report technique, SAM can only reflect conscious psychological states, which fails to differentiate the emotions elicited by archetypal contents. Nevertheless, recent studies have shown the possibility to capture unconscious thoughts through self-report techniques. Through manipulating time and other constraints, self-report techniques still can be used to investigate unconscious psychological states. Since no self-report techniques for recognizing unconscious emotions are currently available, it seems to be a promising direction for research on emotions. On the other hand, while the self-report data of our study did not gain promising results, the results of physiological data have reached a relatively high classification rate for recognizing unconscious emotions. However, as we mentioned in section five, more stimuli need to be included for each category to train the predictive model. A more challenging task is to identify and label these unconscious emotions since people are not consciously aware of them.

In the present study, we have illustrated the importance of exploring affective stimuli based on their symbolic meanings. According to the results of our experiment, we also provide a general comparison about the robustness between self-report technique and physiological measurement for recognizing human emotions. For the main research question about which of the two techniques reflects true emotions, the ultimate answer still requires further justification. We suggest keeping the discussion open for future studies in this direction.

7. ACKNOWLEDGEMENT

This work was supported in part by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments (ICE), which is funded by the EACEA Agency of the European Commission under EMJD ICE FPA n 2010-0012. We are also very grateful for the collaboration with the organization of ARAS, especially Dr. John Beebe, Dr. Baruch Gould, Dr. Iden Goodman, Dr. Tom Singer, and Allison Tuzo as being the great help with this project.

8. REFERENCES

[1] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, no. 1, pp. 87–108, 1995.

[2] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," in *Technical Report A-8*, Gainesville, FL: University of Florida, 2008, pp. 1–12.

[3] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[4] R. Kumar, *Research Methodology: A Step-by-step Guide for Beginners*, 3rd ed. Sage Publications Ltd, 2010.

[5] K. R. Scherer, "What are emotions? And how can they be measured?," *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.

[6] L. F. Barrett, "Emotions are real," *Emotion*, vol. 12, no. 3, pp. 413–29, 2012.

[7] J. Sabini and M. Silver, "Ekman's basic emotions: Why not love and jealousy?," *Cognition & Emotion*, vol. 19, no. 5, pp. 693–712, 2005.

[8] C. G. Jung, *The Archetypes and the Collective Unconscious*, 2nd ed. Princeton, N.J.: Princeton University Press, 1981.

[9] A. Ronnberg and K. Martin, *The Book of Symbols: Reflections on Archetypal Images*. Taschen, 2010.

[10] C. Jung, "Instinct and the unconscious," *The British Journal of Psychology*, vol. 10, no. 1, pp. 15–23, 1919.

[11] H.-M. Chang, L. Ivonin, M. Diaz, A. Catala, W. Chen, and M. Rauterberg, "Experience the world with archetypal symbols: A new form of aesthetics," in *DAPI/HCII 2013, LNCS 8028*, N. Streitz and C. Stephanidis, Eds. Springer, 2013, pp. 205–214.

[12] J. Campbell, *The Hero with A Thousand Faces*. Princeton University Press, 1973.

[13] P. J. Lang, "The mechanics of desensitization and the laboratory study of human fear," in *Behavior Therapy: Appraisal and Status*, C. M. Franks, Ed. New York: McGraw-Hill, 1969, pp. 160–191.

[14] S. D. Kreibig, "Autonomic nervous system activity in emotion: a review," *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.

[15] L. Ivonin, H.-M. Chang, W. Chen, and M. Rauterberg, "Unconscious emotions: Quantifying and logging something we are not aware of," *Personal and Ubiquitous Computing*, vol. 17, no. 4, pp. 663–673, 2013.