9-2010

EXPLORING HOTEL SERVICE QUALITY EXPERIENCE INDICATORS IN USER-GENERATED CONTENT: A CASE USING TRIPADVISOR DATA

Elena García-Barriocanal
University of Alcalá, Spain, elena.garciab@uah.es

Miguel-Angel Sicilia
University of Alcalá, Spain, msicilia@uah.es

Nikolaos Korfiatis
University of Copenhagen, Denmark, nikolaos.korfiatis@econ.ku.dk

EXPLORING HOTEL SERVICE QUALITY EXPERIENCE INDICATORS IN USER-GENERATED CONTENT: A CASE USING TRIPADVISOR DATA

Elena García-Barriocanal, , elena.garciab@uah.es
Miguel-Angel Sicilia, msicilia@uah.es
University of Alcalá, Spain

Nikolaos Korfiatis, University of Copenhagen, Denmark, nikolaos.korfiatis@econ.ku.dk

Abstract

The new social and technological framework of Web 2.0 has resulted in the availability of significant volumes of user generated content. In some cases, user generated content describes existing services, as in the case of travel reviews, in which the users express their experiences and opinions about hotels among other aspects of travel experience. Existing approaches to measuring hotel quality from the customer perspective usually follow the expectation-experience gap model of SERVQUAL or some form of incident analysis. However, user generated content can be used as a complement to automatically gather user opinions in which the aspects covered are those spontaneously raised by customers. This paper reports an initial exploration of such approach on a small sample or reviews in Spanish gathered from TripAdvisor, using existing classifications of emotion types and eliciting conditions. Shallow natural language processing (NLP) techniques are applied to automatically extract simple expressions that can be used to obtain a profile of hotel quality. The results of the preliminary study were able to identify emotion types and eliciting conditions with a reasonable effectiveness which points out to the potential of the techniques to become a complementary tool for hotel evaluation.

Keywords: hotel service quality, user generated contents, emotional expressions, natural language processing.

1 INTRODUCTION

The perceived quality of hotel services has been subject to considerable attention in the literature, with two dominant approaches that can be roughly categorized as incident-based evaluation and attribute based evaluation. The former ones focus on the incidents that customers experience in service contact situations. The latter includes studies using the well-known SERVQUAL gap model (Parasuraman et al., 1998), adapting it to the specificities of the hotel industry (Juwaheer, 2004). Indeed, guest surveys found in many hotels ask for opinions about predefined service features, such as cleanliness of the room or attention by hotel staff, which are used regularly to get feedback on a systematic basis. However, such typical information has been pointed out as incomplete elsewhere (Barskey, 1992), which suggests that alternative ways of gathering the opinions of guests are worth being explored. For example, Desmet, Caicedo and van Hout (2009) have approached gathering reports of emotional experiences instead of opinions as a way to broad the scope of quality evaluation of conventional surveys and predefined assessment instruments.

The widespread use of the Web for travel planning (Barnett and Standing, 2001) points out to the Web as a complementary source of information for hotel quality. Particularly, travel sites that allow the users to express their opinions have been found to provide detailed, rich and relevant data for use by consumers in their travel planning (O'Connor, 2008). Such sites can be considered as "Web 2.0" systems. Web 2.0 is a term which refers to several things, including behaviours, technologies and ideals. It considers services and activities that foster a new kind of media consumer who is more engaged and active in creating and adding value to the content which is the basis for using the Internet (Allen, 2008). Thus, the key of Web 2.0 applications from the viewpoint of digital resource production is that they enable mass participation in social activities structured around the contents. User-generated content (UGC) consists mostly of micro-contents, e.g. reviews, posts or comments in different kinds of applications. Empirical analysis using UGC have started to be reported recently. For example, in the domain addressed in this paper, Stringam, Gerdes & Vanleeuwen (2010) identified four important aspects in the opinions of guests in Internet travel sites, namely: Hotel Service, Hotel Condition, Room Cleanliness, and Room Comfort, and found that hotel service and room comfort were the highest correlated with willingness to recommend.

This paper explores an approach to the automated gathering of hotel service opinions of some specific kinds from UGC sites. Concretely, it reports a concrete exploratory analysis carried out on data extracted from the Tripadvisor Web site[1], and focuses on emotion types and eliciting conditions for emotions. The main idea of the approach explored in this paper is that of exploiting data spontaneously contributed by guests in an attempt to extract useful information from guest experience reports. This data can be considered complementary to the application of questionnaires, as it has a very different nature. On the one hand, this data is not structured around variables or concerns as it happens with existing validated questionnaires, so that it cannot be used to provide a summative assessment of quality. But on the other hand, it has the potential to discover aspects of quality that are overlooked or given less importance in existing questionnaires. The approach to evaluation described here opens new possibilities for research, which are obviously not exhausted in this initial exploratory study.

The rest of this paper is structured as follows. Section 2 briefly reports related work. Then, in Section 3, the method for gathering and analyzing the data is described. Section 4 reports the results and discussion. Finally, conclusions and outlook are provided in Section 5.

2 RELATED WORK

The established incident- or attribute-oriented techniques for hotel service evaluation have started to be complemented by other approaches that use different sources of data. For example, Desmet, Caicedo and

[1] http://www.tripadvisor.es/

van Hout (2009) examined emotions experienced in response to hotel services using an on-line questionnaire.

There are several studies reporting on the actual usage of Web 2.0 sites for the process of travel planning and decision making. Dippelreiter et al. (2008) evaluated eight tourism communities with respect to Web 2.0, showing the different communication means used. Gretzel and Yoo (2008) found evidence that reviews are used mostly to inform accommodation decisions. Cox et al. (2009) reported how websites containing user-generated content (UGC) are used by consumers as a complementary source in their information search and travel behaviour. O´Mahony and Smith (2009) presented a machine learning approach to discriminate "helpful" reviews in UGC travel sites considering user and text attributes.

On-line travel sites have started to be used as sources of empirical data. For example, Briggs, Sutherland and Drummond (2007) used TripAdvisor reviews to examine service quality across small, medium and large hotels in Scotland with the aim of establishing management and customers' perceptions of service quality performance. Au, Buhalis and Law (2009) used TripAdvisor data to identify and analyze complaint categories for Hong Kong hotels. However, to date no automated extraction of service quality perceptions have been applied as a tool.

3 METHOD

Emotional expressions can be extracted from the text of the reviews in on-line travel sites. Such expressions can be identified by some lexical elements that denote emotions or affective information (Valitutti, Strapparava and Stock, 2004). The objective of our preliminary study was that of experimenting with a simple automated extraction procedure with some reasonable amount of error in the identification of positive/negative emotions from hotel reviews. In consequence, simple and non computational complex techniques were applied as a first step towards more complicated (but also more expensive in resources and complexity) solutions.

The study conducted was based on a random sample of review texts extracted from hotels classified in the TripAdvisor site. Only reviews in Spanish were selected as the natural language processing (NLP) tasks are language-specific and the quality of the tools used might vary across languages.

The overall architecture for the information extraction tool is depicted in Figure 1.

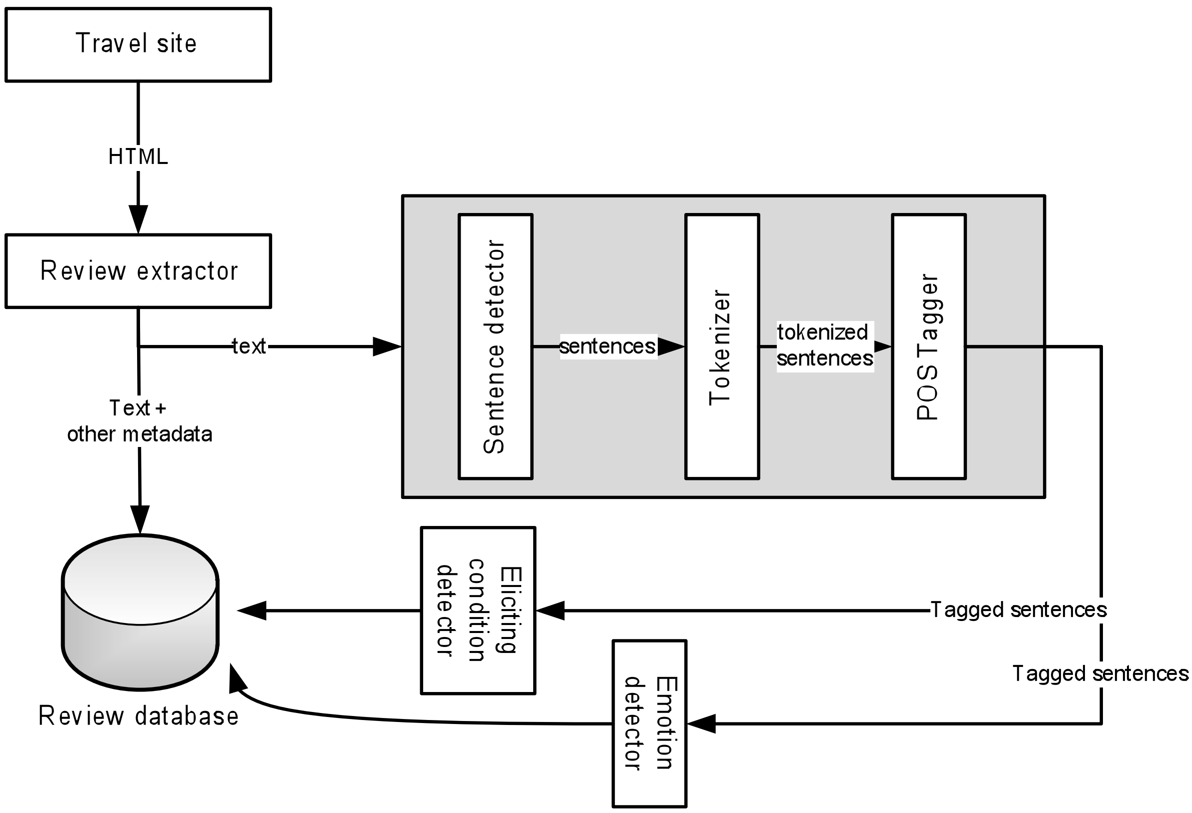

Figure 1. Overall architecture of the system processing the

The first component of the architecture is a crawler that gets via HTTP the reviews from the travel site. This component is highly specific of the travel site as it needs to tackle with the specific ways pages are marked up. The texts are stored in a database together with some additional metadata and each of the reviews is sent to NLP processing. That processing first breaks the sentences in pieces and then extracts the tokens for each of the sentences. Finally, Part-of-speech (POS) tagging is used to assign each word to a grammatical category. All the process so far was implemented using the OpenNLP[2] open source framework that includes these components for Spanish. No special training was applied before the processing. Then, the tagged sentences were sent to two components for detecting emotion types and eliciting conditions respectively. The results were stored in the database.

The emotion types are taken from Desmet, Caicedo and van Hout (2009), which in turn combined other typologies found in the literature. To make them operational, a number of expression types were identified per each of the types (examples are provided in Table 1). The same was done with eliciting conditions.

The emotion type detector was based on identifying positive or negative impressions regardless of the actual object that provokes them. For example, the following sentence "disfruté muchísimo con mi pareja" results in the following POS tagging[3]:

`[`**`VIS`**`, RG, SPS, DP, NC]`

In this case, a shallow pattern matching could simply search for the categorization `("disfruté", VIS`[4]`)`. It should be noted that the presence of the verb is in this case enough and no further pattern inspection is carried out (in some cases, the pattern could exclude negations which could be interpreted in the opposite sense). It should be noted that the lexical elements (verbs, nouns, adjectives) that are

[2] http://opennlp.sourceforge.net/
[3] The tag labels are described in the OpenNLP documentation.
[4] "Verb past"

denoting the emotion were for this case identified by inspecting a sample of texts and creating a simple dictionary. In the future, linguistic resources as WordNet Affect could be used instead together with some morphological analyzer that covers variations in number and inflection, providing some more flexibility to the approach.

As a noun, "disfrute" (enjoyment) appear with a different structure as in the sentence "Fue un disfrute total", which produces the following POS annotation:

`[VIS, DI, NC[5], AQ]`

These patterns are detecting emotion types based on some concrete words, which for some cases as the one described can appear in different grammatical categories.

The identification of eliciting conditions requires a different kind of structural patterns. Following the example, "Disfruté de la piscina climatizada" would result in the structure:

`[VIS, SPS, DA, NC, AQ]`

In this case, a vocabulary of hotel facilities and services is used to combine the verb indicating the experienced emotion with the concrete eliciting condition, in this case, the pool ("piscina" in Spanish). The representation of the pattern in this case would be `("disfruté", VIS)->(?x, NC)`. Regular expressions matching the possible syntactic structures are enough to detect the relationship condition-emotion in most cases. However, a more accurate version would require inspecting the syntactical structure of the sentence and implement more complex parsing mechanisms. Following the example, verb "disfrutar" can be transitive or intransitive, e.g. in its intransitive form as in "Disfruté con los animadores" it results in `[VIS, SPS, DA, NC]` where `SPS` is the preposition. Including such details in the parsing is not always providing some added value as in the example, for which the structure of `VIS` plus noun phrase (as in the previous example) still hold for the intransitive version.

The approach implemented is extensible in that it requires only extending the dictionaries to cover a broader spectrum of expressions, and the grammatical regular expressions matching positive/negative emotions can also be modified by editing a regular expression file. However, this still requires human inspection of the expressions, some kind of statistical learning alternative could be explored to make the process more automated and to allow more flexibility, however this would require the production of specific corpora.

4 RESULTS AND DISCUSSION

4.1 Extraction of reported emotions

A total of 412 texts were used for the analysis, from an initial set of 450 from which a shallow manual inspection was done to filter out texts that were poor from a grammatical viewpoint. The texts were describing 320 different hotels, with low frequencies of coincidence on reviews for the same hotel. As the objective was only to evaluate the potential effectiveness of the approach, such an sparse collection was considered adequate. However, it would be interesting in future studies to aggregate the reviews from the same hotel, so that a judgement of the overall positive or negative response of each reviewer could be obtained.

The results of the first task are showed in Table 1. It should be noted that the categories have been interpreted according to the concrete sample used for the study, and they may differ from the original interpretations given by Desmet, Caicedo and van Hout. Some of the categories were not included as there were no clear differentiating linguistic indicators for them (they are marked as "n.a." in Table 1).

[5] "noun (common)"

Kind of emotion	Example indicator	Frequency and %		% in Desmet, Caicedo and van Hout (2009)
Enjoyment	"Disfruté muchísimo", "Fue un disfrute"	142	23,5%	16
Satisfaction	"me gustó"	128	21,2%	15,1
Dissatisfaction	"no me gusto nada" "nos decepcionó"	67	11,1%	14,2
Aversion	"desagradable"	32	5,3%	12,5
Boredom	"después de un par de días se vuelve muy aburrido"	29	4,8%	7,1
Admiration	"El diseño era admirable en cada detalle..." "impresionante" "increíble"	82	13,6%	6,8
Attraction	"La piscina era muy apetecible."	56	9,3%	5,3
Sadness	"La localización es un poco triste"	17	2,8%	5,0
Fascination	"Lo más impresionante de todo, sin embargo, es el personal"	43	7,1%	4,7
Fear	n.a.		-	4,2
Hope	n.a.		-	4,2
Pride	"Michelangelo estaría orgulloso"	3	0,5%	2,1
Shame	"Es una vergüenza que..."	4	0,7%	1,8
Contempt	n.a.	-		0,9

Table 1. Emotion types, example expressions and frequencies found in the sample, contrasted with a previous study.

The figures reported in Table 1 show a different distribution of emotions than in the previous study Desmet, Caicedo and van Hout (2009), but still they have a 0.82 statistically significant Pearson correlation. However, sampling and analysis methods are broadly diverging so that no strong conclusions can be drawn from the contrast. Anyway, the results show that similar datasets of emotions can be identified with the simple automated approach used.

The emotions found were revised manually after the automatic extraction by reading the original text. Incorrectly extracted emotion types were less than 20%, which represents a good figure of error for the simple and straightforward approach used. It should be noted that the NLP techniques used are in a rather preliminary status, so that it can be expected that better figures would be obtained after refining the patterns used and extending the vocabulary.

4.2 Extraction of eliciting conditions

The approach to extract eliciting conditions differs significantly from the one reported by Desmet, Caicedo and van Hout. In our case, the main aim of the extraction process is that of identifying elements of the hotel experience that are attributed to particular aspects of the hotel service and facilities. It should be noted that the categories found are hierarchically structured, for example "hotel staff" subsumes the category of "cleaning staff". However, the hierarchical relationships were not considered in this initial study. Table 2 shows the percentages of occurrence of pleasant and unpleasant emotions expressed for a selection of the eliciting conditions reported by Desmet, Caicedo and van Hout.

Pleasant emotions		Unpleasant emotions	
Eliciting condition	Percentage	Eliciting condition	Percentage
Hotel staff	15,9%	Price	15,7%
Bed	15,2%	Hygiene	12,7%
Food	12,4%	Hotel staff	11,4%
Space	10,3%	Delays	10,8%
Price	9,7%	Decoration	10,2%
Decoration	8,3%	Bed	9,0%
Hygiene	8,3%	Food	9,0%
Location	6,2%	Space	7,2%
Maintenance	6,2%	Maintenance	6,6%
Lighting	5,5%	Location	4,2%
Delays	2,1%	Lighting	3,0%

Table 2. Pleasant and unpleasant emotions found grouped by eliciting condition

It should be noted that the eliciting conditions reported in Table 2 are not necessarily covering the emotion types in Table 1, as the set of conditions analyzed is restricted, and some of them are not captured by the analysis of emotion types carried out. For example, *Price* were captures either by adjectives as "expensive" or "cheap" or by sentences as "[something] was expensive". This example shows also the additional problem of overlap of some eliciting conditions, as in "dinner was expensive" for which it is a matter of interpretation to classify it either on "*Food*" or in "*Price*". That kind of problem is pointing out that a clearer terminology or ontology for the elements present in hotel experiences is required.

4.3 Limitations

The present study is of an exploratory nature and the results cannot be generalized without further studies that analyze much larger samples. Also, the NLP techniques used are limited to simple sentences following some structural patterns, which are excluding a significant part of the information contained in reviews. This is a serious challenge, as using more complex NLP techniques would also entail increased computational complexity and in general decreased performance in processing tasks, which would affect the scalability of this kind of evaluation. Also, it is apparent that many reviews are written with poor style and subject to abundant grammatical errors. This may hamper the usefulness of the techniques reported here and it would call for some flexibility in the NLP processes with regards to errors.

The selection of expressions for categorizing emotion types cannot be done in all the cases directly from the Spanish translation of the kind categories, as some of them may lead to erroneous results. For example, admiration in Spanish language is expressed as “admirable” as an adjective and with the verb “admirar” among others. However, these appear most frequently describing the surrounding places (monuments, natural environments) and not the hotels themselves. These have been removed manually in the present study, but further work should attempt to apply semantic techniques to filter out expressions not describing the hotel.

A potential limitation of the kind of analysis presented here lies in the influence of cultural factors that are known to be significant in the perceptions of hotel service quality (Mok and Armstrong, 1998), alomg with differences in the effectiveness of the technique depending on the language addressed.

Another potential limitation is the problem of false reviews. However, this is not a problem of the approach presented here but in general of the trustworthiness of UGC sites. However, O’Connor (2008) pointed out that the belief that *TripAdvisor* is compromised by false reviews posted to enhance a hotels reputation or tarnish that of competitors is unfounded.

5 CONCLUSIONS AND OUTLOOK

Hotel quality models have focused in the past in customer perceptions and expectation gaps that capture the subjective evaluation of different elements of the service and its context. However, the widespread use of Web 2.0 sites that collect hotel comments and ratings by their users provide a complementary source of information to questionnaire-based studies, which has the potential of better capturing the emotional response to staying experiences. Shallow NLP techniques can be used to detect sentences in on-line hotel reviews and tag them to a level that allows identifying simple opinions that provide some useful feedback for management. This paper has reported a preliminary study on such an approach, focusing on emotional responses and their associated eliciting conditions. The results are comparable with previous ones in which the analysis was done solely on the basis of human effort in analyzing text. This constitutes a promising direction in automating information extraction from travel sites, which would constitute an inexpensive mechanism for hotel managers to get feedback from the social context of Web 2.0 applications.

However, the study is limited in sample size and selection method, and the techniques used are straightforward and account only for part of the potential emotion-expressing kinds of sentences. In consequence, future work should advance in making NLP processing more elaborated to account with more diverse kinds of expressions. For example, WordNet Affect (Valitutti, Strapparava and Stock, 2004) can be used as a more reliable and comprehensive source of lexical information for detecting emotional expressions in the text. In another direction, there are additional sources of information that could be exploited and combined with the data extracted from text. Of particular importance are ratings (Stringam, Gerdes & Vanleeuwen, 2010), which have demonstrated to be a successful approach for recommendation in e-commerce (Sarwar et al., 2000), and also opinions on other user’s ones, which have the effect of a filtering mechanism.

References

Allen, M. (2008). Web 2.0: An argument against convergence. First Monday 13(3).

Au, N., Buhalis, D. & Law, R. (2009). Complaints on the Online Environment – The Case of Hong Kong Hotels. In Hopken, W., Gretzel, U. and Law, R. (Eds.), Information and Communication technologies in Tourism 2009. Springer-Verlag Wien-New York, pp. 73-85.

Barnett, M., Standing, C. (2001), Repositioning travel agencies on the Internet, Journal of Vacation Marketing 7(2), pp.143-52.

Barsky, J. (1992). Customer satisfaction in the hotel industry: meaning and measurement. Hospitality Research Journal, 16(1), pp. 51-73.

Briggs, S., Sutherland, J. and Drummond, S. (2007) Are hotels serving quality? An exploratory study of service quality in the Scottish hotel sector, Tourism Management, 28(4), pp. 1006-1019.

Buhalis, D. and Law, R. (2008) Progress in information technology and tourism management: 20 years on and 10 years after the Internet--The state of eTourism research, Tourism Management, 29(4), pp. 609-623

Cox, C., Burgess, S, Sellitto, C & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior', Journal of Hospitality Marketing and Management, 18(8), pp. 743-764.

Desmet, P.M.A., Guiza Caicedo, D., & Van Hout, M. (2009). Differentiating Emotional Hotel Experiences. In: T. Kokko (Ed.), Proceedings of the International Hospitality & Tourism Educators EuroCHRIE conference, Helsinki (Finland), 22-24 October 2009.

Dippelreiter, B., Grün, C., Pöttler, M., Seidel, I., Berger, H., Dittenbach, M., Pesenhofer, A.(2008) Online Tourism Communities on the Path to Web 2.0: An Evaluation. Information Technology & Tourism, 10(4), pp. 329-353(25)

Gretzel, U. & Yoo, K. H (2008). Use and Impact of Online Travel Reviews. In O'Connor, P., Höpken, W. & Gretzel, U. (Eds.). Information and Communication Technologies in Tourism 2008, 35-46. Vienna, Austria: Springer.

Juwaheer T.D. (2004) Exploring international tourists’ perceptions of hotel operations by using a modified SERVQUAL approach: a case study of Mauritius, Managing Service Quality 14 (2004) (5), pp. 350–364.

Mok, Connie, Armstrong, Robert W. (1998) Expectations for hotel service quality: Do they differ from culture to culture? Journal of Vacation Marketing 1998 4, pp. 381-391

O'Connor, P. (2008). User-Generated Content and Travel: A Case Study on TripAdvisor.com. In O'Connor, P., Höpken, W. & Gretzel, U. (Eds.). Information and Communication Technologies in Tourism 2008, pp. 47-58. Vienna, Austria: Springer

O'Mahony, M. P. and Smyth, B. (2009). Learning to recommend helpful hotel reviews. In Proceedings of the Third ACM Conference on Recommender Systems (New York, New York, USA, October 23 - 25, 2009). RecSys '09. ACM, New York, NY, pp. 305-308.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1988) SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64 (1988) (1), pp. 12–40.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. In Proceedings of the 2nd ACM Conference on Electronic Commerce (Minneapolis, Minnesota, United States, October 17 - 20, 2000). EC '00. ACM, New York, NY, pp. 158-167

Stringam, B., Gerdes, J. & Vanleeuwen, D. M. (2010). Assessing the Importance and Relationships of Ratings on User-Generated Traveler Reviews. Journal of Quality Assurance in Hospitality & Tourism, 11(2), pp. 73-92.

Valitutti, A., Strapparava, C. and Stock, O. (2004). Developing affective lexical resources. Psychology: 2 (1), pp. 61 - 83