Beyond Text: Leveraging Multi-Task Learning and Cognitive Appraisal Theory for Post-Purchase Intention Analysis

Gerard Christopher Yeo[1], **Shaz Furniturewala**[2], **Kokil Jaidka**[3]
[1]Institute of Data Science, National University of Singapore,
[2]Birla Institute of Technology and Science,
[3]NUS Centre for Trusted Internet and Community, National University of Singapore
e0545159@u.nus.edu, mohammadshaz529@gmail.com, jaidka@nus.edu.sg

Abstract

Supervised machine-learning models for predicting user behavior offer a challenging classification problem with lower average prediction performance scores than other text classification tasks. This study evaluates multi-task learning frameworks grounded in Cognitive Appraisal Theory to predict user behavior as a function of users' self-expression and psychological attributes. Our experiments show that users' language and traits improve predictions above and beyond models predicting only from text. Our findings highlight the importance of integrating psychological constructs into NLP to enhance the understanding and prediction of user actions. We close with a discussion of the implications for future applications of large language models for computational psychology.

1 Introduction

Natural language processing (NLP) tasks involve predicting outcomes from text, ranging from the implicit attributes of text to the subsequent behavior of the author or the reader. Recent research suggests that user-level features can carry more task-related information than the text itself (Lynn et al., 2019), but these experiments have been conducted in a limited scope. Other studies have explored how the linguistic characteristics of text, such as its politeness or the use of discursive markers, may predict subsequent user behavior (Danescu-Niculescu-Mizil et al., 2013; Niculae et al., 2015). Yet, these studies offer unimodal perspectives of users through the text they author and lack rich annotations of other user attributes, such as their cognitive and psychological traits. Such data would be especially useful in applied NLP tasks, such as in the context of online reviews, to better contextualize and predict outcomes related to purchase behavior and product recommendations.

In this study, we focus on **Cognitive Appraisal Theory**, one of the primary theoretical frameworks in Psychology to understand emotional experiences and how they are elicited (antecedents). Central to Cognitive Appraisal Theory is the proposition that emotions are not merely spontaneous reactions but are the result of intricate cognitive evaluations conducted across multiple dimensions of psychological motivation that are of personal significance to one's well-being, as discussed by seminal works in the field (Lazarus and Folkman, 1984; Ortony et al., 2022; Scherer et al., 2001; Smith and Ellsworth, 1985). People interpret— or appraise— situations along various dimensions, and the specific manner in which people appraise their situations characterizes the particular emotions they feel. For example, if a consumer evaluates a restaurant experience as slow (goal inconduciveness), the server was specifically being rude to them (unfair), and blames the waiter for such an experience (accountability-other), then the consumer might feel an emotion like *anger*. Our empirical investigation specifically targets the nuances of purchase behavior, guided by a focus on two critical dimensions as illuminated by Cognitive Appraisal Theory:

- **Cognitive appraisals**: The multifaceted evaluative processes through which consumers engage with and interpret their interactions with products, including, but not limited to, the novelty and pleasantness of the consumer-product encounter (Yeo and Ong, 2023).
- **Emotions**: The range of emotions consumers may experience during product usage. Emotions such as anger and disappointment are pivotal, as they color the immediate consumer experience and influence subsequent behaviors and attitudes towards the product (Ruth et al., 2002).

Setup and Motivation: This study predicts post-purchase behavior as the outcome of emotions and their antecedents. Prior work has reported that the myriad of emotions experienced by consumers interacting with a product/service (Richins, 1997) can influence post-consumption behaviors (PCB)

Findings of the Association for Computational Linguistics ACL 2024, pages 12353–12360
August 11-16, 2024

like future purchases and likelihood to promote the product to others (Folkes et al., 1987; Lerner et al., 2015; Nyer, 1997; Watson and Spence, 2007). Although previous studies have demonstrated that language models capture emotionally relevant features (Acheampong et al., 2021; Deng and Ren, 2021), these studies do not relate such features to other relevant psychological traits such as cognitive appraisals in the understanding of user behavior. Modeling cognitive appraisals and emotions in language models not only aids in predicting behavioral intentions but also explains *why* people have different behavioral intentions after interacting with a product or service. For example, if a person does not want to recommend the product, this could be attributed to appraisals such as low goal-conduciveness or unfairness and emotions such as disappointment and anger.

We evaluate a series of multi-modal and multi-task learning setups that apply Cognitive Appraisal Theory, as reported in Figure 1. The models proposed here are not constructed by merely combining additional information such as emotion and cognitive appraisals but based on theoretical proposals on how such information is related to one another, which is something that previous research has not done in the context of behavioral outcomes (Liu and Jaidka, 2023). The following are our contributions:

- A multi-task learning framework incorporating emotional and cognitive appraisal variables in a theoretical manner to predict PCB.
- An exploration of the empirical association of PCB with cognitive appraisals, emotions, and the text authored by the consumer.

2 Dataset and Variables

We used the PEACE-Reviews Dataset (Yeo and Jaidka, 2023), a dataset of 1,400 author-annotated product reviews describing people's emotional experiences of using an expensive product/service. To our knowledge, this is the only reviews dataset annotated with a large number of first-person emotional and behavioral intentions variables. Existing emotion text datasets are usually annotated with only a subset of these variables, without the inclusion of any behavioral intentions ratings (Scherer and Wallbott, 1994). Most importantly, existing emotion text datasets are typically annotated with third-person annotations (i.e., raters rate text written by other people), where such annotations might not correspond to the writers' first-hand experiences (Mohammad et al., 2018). In the PEACE-Reviews Dataset, each review was annotated with first-person emotions, cognitive appraisals, and PCB ratings, which makes the dataset exceptionally relevant in comprehensively modeling consumers' first-hand emotional experiences and behavior intentions. Our multi-task framework incorporates the following inputs:

- **Review text**. The review text comprises detailed descriptions of consumer-product interactions and specific aspects of the product/service that explain why consumers feel a particular emotion. The mean length of the reviews is 190.2 tokens, which makes them substantively longer than other review datasets (Maas et al., 2011).
- **Cognitive appraisals**. Each review is annotated with 20 appraisal dimension ratings that measure how consumers evaluate the consumer-product interactions relevant to their emotional experiences (Yeo and Jaidka, 2023). Each dimension is rated on a 7-point Likert scale, assessing the extent to which participants appraised their consumption experience in a particular manner (see Table 2 in Appendix A). For example, suppose a participant rated a particular appraisal dimension such as *novelty* as high; it means that they evaluated the product/service usage as a new experience they have never encountered before.
- **Emotions**. Each review was also annotated on a 7-point Likert scale measuring the intensity for 8 emotions: *anger*, *disappointment*, *disgust*, *gratitude*, *joy*, *pride*, *regret*, and *surprise*, adapted from the common emotions experienced in a consumption context (Richins, 1997). Unlike current emotion recognition datasets where each text is labeled with only one emotion (Mohammad et al., 2018; Scherer and Wallbott, 1994), the presence of multiple emotion ratings in this dataset is more consistent with real-life situations where consumers typically experience more than one emotion in a consumption context (Ruth et al., 2002).
- **Post-consumption behaviors (PCBs)**. These are the primary outcome variables in our study. Two variables in the dataset assessed the likelihood of engaging in different post-consumption behaviors-*intention to repurchase* and *intention to promote*. They are both measured on a 7-point Likert scale. These variables are indicative of whether real actions might be taken in the future (Engel and Roger, 1995).

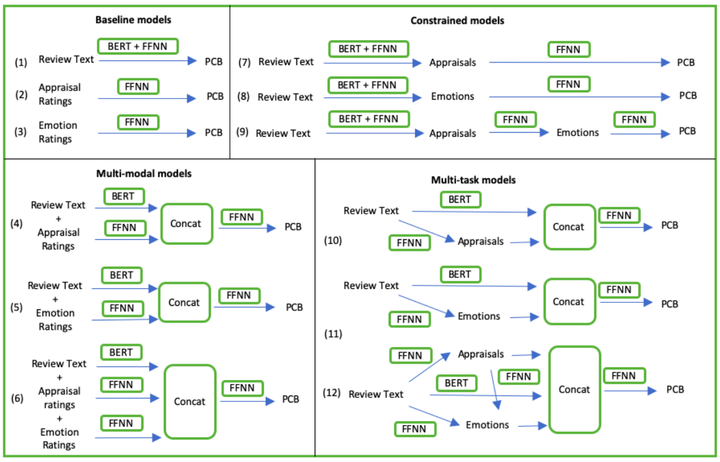

Figure 1: Models implemented in our study. Model (12) is the theoretical model.

3 Experiments

See Figure 1 for a visual representation of all models. We fine-tuned the BERT-base model (Devlin et al., 2018) for models requiring input text. We trained feed-forward neural networks (FFNN) for models that require appraisal and emotion ratings as inputs. Since PCBs are rated on a 7-point Likert scale, we segment each rating into low (1-2), moderate (3-5), and high (6-7) and define it as a three-way classification task (see Appendix B for the distribution of classes). For multi-task models where appraisals and emotions are outcome variables, we defined a multi-label binary classification task for emotion ratings, where we segment each rating into low (1-4) and high (5-7). This segmentation represents the presence or absence of emotions experienced by the participant in the situation, where only emotions that are felt with high intensity are considered to be present. We define a multi-output classification task for appraisals where we segment each rating into low (1-2), moderate (3-5), and high (6-7). The segmentation of appraisal ratings in this manner is typical in emotion research (Smith and Ellsworth, 1987). We conducted 5 repetitions for each model and obtained the means and standard deviations of the accuracy and F1 scores. Implementation details are in the Appendix C.

Baseline models. Three models serve as the baselines. We run separate models to predict PCBs for each modality M_i, where M = [text, appraisals, emotions]. We would like to observe which modality performs best in predicting PCBs.

Constrained models. We implemented three models. The first two models use the BERT model fine-tuned on the reviews to predict either the appraisal or emotion ratings, and the resulting embeddings are then used to predict PCBs. The third model uses the BERT model fine-tuned on the reviews to predict appraisals, subsequently uses these appraisal embeddings to predict emotions, and finally uses the resulting emotion embeddings to predict PCBs. According to emotion theory, this follows where appraisals are deemed to be antecedents to emotions, resulting in behaviors (Watson and Spence, 2007). They are termed *constrained* because the intermediate variable (appraisals or/and emotions) serves as a bottleneck that reduces the textual dimensions to a much lower dimension in predicting PCBs, compared to directly predicting PCBs from text.

Multi-modal models. We implemented three models. The first two models predicted PCBs using review text + M_i, where M = [appraisals, emotions]. The third model predicted PBs from all three modalities. The embeddings of the modalities are concatenated to predict PCBs. This modeling approach is chosen for its capacity to assimilate psychological variables alongside linguistic features. The results allow us to compare whether ratings combined with review text help improve performance predicting PCBs.

Multi-task models. We implemented three models. For the first two models, the review texts are used to predict the PCBs and R_i, where R = [appraisals, emotions], simultaneously. Moreover, the embeddings of R_i are used to predict PCBs by concatenating with the text embeddings. The final model, termed 'Theoretical model', uses the review text to predict appraisals, emotions, and PCBs. The resulting embeddings from each modality are then

	Intent to repurchase		Intent to promote	
Model	**Accuracy**	**F1**	**Accuracy**	**F1**
Baseline				
Text -> PCB	70.1 (0.29)	0.61 (0.01)	72.4 (0.57)	0.66 (0.01)
Appraisals -> PCB	73.4 (0.29)	0.70 (0.01)	75.9 (1.05)	0.74 (0.01)
Emotions -> PCB	67.7 (0.53)	0.66 (0.01)	73.3 (0.35)	0.72 (0.01)
Constrained				
Text -> Appraisals -> PCB	69.0 (0.35)	0.58 (0.01)	69.6 (0.57)	0.58 (0.01)
Text -> Emotions -> PCB	68.6 (0.45)	0.58 (0.01)	69.1 (0.29)	0.58 (0.01)
Text -> Appraisals -> Emotions -> PCB	67.3 (0.83)	0.57 (0.01)	68.3 (0.31)	0.58 (0.01)
Multi-modal				
Text + Appraisals -> PCB	68.0 (0.34)	0.68 (0.02)	72.6 (0.97)	0.70 (0.01)
Text + Emotions -> PCB	72.0 (0.21)	0.66 (0.01)	70.0 (0.44)	0.69 (0.02)
Text + Appraisals + Emotions -> PCB	72.0 (0.24)	0.72 (0.01)	72.0 (0.23)	0.70 (0.02)
Multi-task				
Text -> PCB + Appraisals	69.3 (0.45)	0.58 (0.01)	71.7 (0.32)	0.64 (0.03)
Text -> PCB + Emotions	69.1 (0.32)	0.61 (0.02)	73.6 (0.58)	0.67 (0.01
Theoretical model	69.3 (0.58)	0.60 (0.02)	73.4 (0.64)	0.69 (0.02)

Table 1: Results of three-way (high, medium, low) post-consumption behavior (PCB) classification across models, for intention to promote and intention to repurchase. Values without and within the parentheses represent the means and standard deviations across 5 runs.

concatenated to predict PCBs. Additionally, we also used the appraisal embeddings to predict emotions. Overall, this model is based on consumer and psychological theories. We would like to validate whether such a computational model consisting of the variables and their theoretical links has predictive utility in the context of language.

4 Results

Table 1 presents the results for the different models in predicting the two PCBs. Among the baseline models, models trained directly on appraisals were the most accurate. The Emotions -> PCB model only outperformed the Text -> PCB model in predicting intentions to promote, but not for intentions to repurchase. Despite this, the Text -> PCB model's performance was still competitive, suggesting that large language models can capture pertinent linguistic features, including those beyond emotional content.

The poorest results came from constrained models, likely due to the reduction of text embeddings to a lower dimensional appraisal and emotion feature space and resulted in the lost of information in predicting PCB.

The integration of different modalities (multi-modal models) did not enhance the performance as expected, indicating that unique information from each modality may not be additive for PCB prediction. Nevertheless, some multi-modal models offer a slight edge in accuracy and F1 scores compared to the baseline Text -> PCB model as observed for the results of the Text + Appraisals -> PCB model in predicting intention to promote, and Text + Emotions -> PCB and Text + Appraisals + Emotions models in predicting intent to repurchase.

For the multi-tasks and theoretical models, for the prediction of intent to promote, the theoretical model and the Text -> PCB + Emotions models outperformed the constrained, multi-modal and baseline Text -> PCB models. In predicting intent to repurchase, the performances of the multi-tasks and theoretical models are similar to the constrained and the baseline Text -> PCB models but did not perform better than the multi-modal models. This suggests that combining appraisals and emotions based on theory might not be optimal in predicting intention to repurchase compared to merely combining the features of the text, appraisals, and emotions.

In general, for intention to promote, the multi-task and theory-informed models performed modestly better than the rest of the models (except for the Appraisals -> PCB model), likely due to their structured integration of appraisal and emotional constructs. However, for intention to repurchase, the two multi-modal models (excluding Text + Appraisals -> PCB) performed the best (except for the Appraisals -> PCB model). Overall, our results affirm that incorporating appraisal and emotional considerations generally enhances PCB prediction and supports the validity of Cognitive Appraisal Theory in informing multi-task learning approaches.

Word attributions and explainability. We implemented the Integrated Gradients method to obtain the word attributions to explain the predictions (Sundararajan et al., 2017). The visual depictions in Figure 2 showcase word attributions corresponding to high and low instances of intentions to promote or repurchase, respectively, predicated upon our baseline Text -> PCB model. The word attributions underscore the integral role of the emotionally-charged lexicon — 'enjoyment,' 'disappointing' — and cognitive appraisal terms — 'unexpected,' 'important,' and 'consistent' — in influencing the predictive outcomes of our BERT-based model.

The first two rows indicate that the model's reliance on affective language is pronounced, indicating a robust association between sentiment-laden words and positive intention to promote. In contrast, the word- and phrase- associations with intention to purchase illustrate a less pronounced correlation. We can infer that emotionally reso-

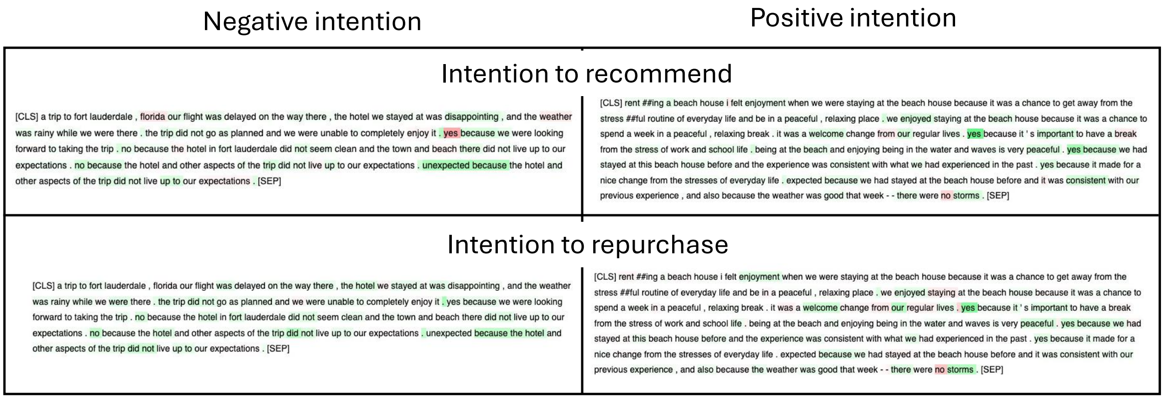

Figure 2: Word attribution of two samples that scored high and low in PCBs based on the baseline text -> PCB model, respectively.

nant words seem more decisive in predicting the intention to promote, while a blend of cognitive appraisal and emotional language informs purchase intentions. This distinction may be crucial for refining the predictive efficacy of sentiment analysis models in consumer behavior contexts.

Finally, the figures highlight the errors in how non-cognitive, non-emotional words (e.g., 'Florida,' and 'hotel') are correlated with PCBs. Overall, our results are consistent with the findings that emotions and appraisals have significant links to PCBs (Nyer, 1998). The analysis of word attributions in our models sheds light on the cognitive processes underpinning specific emotional reactions and behavioral tendencies. Therefore, fine-tuning transformer models with appraisal and emotional variables and identifying linguistic features of such variables can potentially improve the prediction of PCBs. Future studies could implement models that learn these variables simultaneously in a multi-task framework, thereby predicting PCBs.

5 Conclusion

Many NLP tasks focus on predicting user behavior, and enriching text-based models with user and social contexts is increasingly necessary. This work emphasizes the increasingly prominent role of cognitive and emotional signals in behavioral prediction. Consumption emotions act as adaptive signals of how we evaluate how the use of products/services affects our well-being, which subsequently triggers future actions to either promote positive emotions (e.g., repurchasing or promoting to others) (White, 2010) or reduce negative emotions (e.g., complaint behaviors) (Stephens and Gwinner, 1998). To our knowledge, the current work is the first to construct models grounded on psychological theory to model real post-consumption decision-making processes, and we find empirical support for these associations. More broadly, our study offers a novel methodological approach to study psychological variables in the context of empirically validating theoretical relationships within review texts—a domain previously unexplored beyond the confines of traditional survey methods. Our work finds variance in the importance of these appraisals across tasks, raising important practical considerations for designing future approaches to behavioral prediction.

Limitations

This study used a dataset primarily curated to study emotional responses in review text in the context of using expensive products/services. Although we have established that emotional constructs are important in modeling PCB intentions, one limitation is that the current results might not generalize to other review datasets and contexts. One research direction we would like to pursue is to analyze whether the results from fine-tuning models on the PEACE-Reviews dataset can generalize to other public review datasets with different emotional content, length, contexts, and product/service types. Moreover, since typical review datasets only contain ratings of sentiments and helpfulness, to establish the criterion validity of our models in measuring PCBs, we can estimate the correspondence between predicted PCB scores of our models with other ratings like sentiment and helpfulness. This can further solidify the case that emotion and appraisals are important variables in modeling consumer experiences and behaviors.

Another limitation is that the dataset only provides ratings for 8 emotional experiences. Although we mentioned that these emotions are typically experienced during consumption, they might not comprehensively capture all emotional experiences (Richins, 1997). Despite that, we accounted for the observation that consumers might experience multiple emotions in a situation and also used appraisal dimension ratings to model emotional experiences. Since cognitive appraisal theory posits a one-to-one mapping between appraisal profiles and emotional experiences (Ellsworth and Scherer, 2003), modeling the 20 appraisal dimensions could mitigate the issue of not comprehensively capturing a wide range of emotional experiences.

Ethics Statement

Since we did not collect any data from human subjects but instead used an existing dataset that a review board has reviewed, we do not foresee any potential harm in the methodology of the current study. Moreover, no personal information that could identify individual human participants was in the dataset which can cause privacy issues.

Extensive literature corroborates the significant impact of cognitive appraisals and emotions on consumer behavior. Our study's objective, to model consumer behavior through emotional variables in review texts, is anchored in a vision of advancing product design and business strategies. Note that the intention of this study is not to manipulate emotional and psychological traits to influence consumer behaviors, but rather to understand and predict consumer behaviors more accurately, thereby contributing valuable insights for an informed decision-making process in business practices. The empirical results and models offered in this study can have potential positive managerial implications such as informing marketing strategies, business decisions, and product engineering. Therefore, users of our models should tailor them to their use cases to aid in understanding consumer behaviors in their specific domain. Furthermore, the current work also adopted the Integrated Gradients method to explain the models' predictions to improve the transparency and interpretability of models to better shape users' decisions. This ensures that decisions are supported by linguistic features in reviews that have theoretical links with PCBs.

References

Francisca Adoma Acheampong, Henry Nunoo-Mensah, and Wenyu Chen. 2021. Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 250–259, Sofia, Bulgaria. Association for Computational Linguistics.

Jiawen Deng and Fuji Ren. 2021. A survey of textual emotion recognition and its challenges. *IEEE Transactions on Affective Computing*, 14(1):49–67.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Phoebe C Ellsworth and Klaus R Scherer. 2003. Appraisal processes in emotion.

James F Engel and D Roger. 1995. Blackwell (1982), consumer behavior. *New York: Holt, Renehard, and Winston*.

Valerie S Folkes, Susan Koletsky, and John L Graham. 1987. A field study of causal inferences and consumer reaction: the view from the airport. *Journal of consumer research*, 13(4):534–539.

Richard S Lazarus and Susan Folkman. 1984. *Stress, appraisal, and coping*. Springer publishing company.

Jennifer S Lerner, Ye Li, Piercarlo Valdesolo, and Karim S Kassam. 2015. Emotion and decision making. *Annual review of psychology*, 66:799–823.

Xuan Liu and Kokil Jaidka. 2023. I am psyam: Modeling happiness with cognitive appraisal dimensions. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1192–1210.

Veronica Lynn, Salvatore Giorgi, Niranjan Balasubramanian, and H Andrew Schwartz. 2019. Tweet classification without the tweet: An empirical examination of user versus document attributes. In *Proceedings of the third workshop on natural language processing and computational social science*, pages 18–28.

Andrew Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pages 142–150.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation*, pages 1–17.

Vlad Niculae, Srijan Kumar, Jordan Boyd-Graber, and Cristian Danescu-Niculescu-Mizil. 2015. Linguistic harbingers of betrayal: A case study on an online strategy game. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1650–1659.

Prashanth U Nyer. 1997. A study of the relationships between cognitive appraisals and consumption emotions. *Journal of the Academy of Marketing Science*, 25(4):296–304.

Prashanth U Nyer. 1998. The effects of satisfaction and consumption emotion on actual purchasing behavior: An exploratory study. *The Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 11:62–68.

Andrew Ortony, Gerald L Clore, and Allan Collins. 2022. *The cognitive structure of emotions*. Cambridge university press.

Marsha L Richins. 1997. Measuring emotions in the consumption experience. *Journal of consumer research*, 24(2):127–146.

Julie A Ruth, Frederic F Brunel, and Cele C Otnes. 2002. Linking thoughts to feelings: Investigating cognitive appraisals and consumption emotions in a mixed-emotions context. *Journal of the Academy of Marketing Science*, 30:44–58.

Klaus R Scherer, Angela Schorr, and Tom Johnstone. 2001. *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Klaus R Scherer and Harald G Wallbott. 1994. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66(2):310.

Craig A Smith and Phoebe C Ellsworth. 1985. Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, 48(4):813.

Craig A Smith and Phoebe C Ellsworth. 1987. Patterns of appraisal and emotion related to taking an exam. *Journal of personality and social psychology*, 52(3):475.

Nancy Stephens and Kevin P Gwinner. 1998. Why don't some people complain? a cognitive-emotive process model of consumer complaint behavior. *Journal of the Academy of Marketing science*, 26(3):172–189.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR.

Lisa Watson and Mark T Spence. 2007. Causes and consequences of emotions on consumer behaviour: A review and integrative cognitive appraisal theory. *European Journal of Marketing*, 41(5/6):487–511.

Christopher J White. 2010. The impact of emotions on service quality, satisfaction, and positive word-of-mouth intentions over time. *Journal of marketing management*, 26(5-6):381–394.

Gerard Yeo and Kokil Jaidka. 2023. The peace-reviews dataset: Modeling cognitive appraisals in emotion text analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2822–2840.

Gerard Christopher Zheng Jie Yeo and Desmond C Ong. 2023. A meta-analytic review of the associations between cognitive appraisals and emotions in cognitive appraisal theory.

A Measures of Appraisal Dimensions

Table 2 provides the items measuring the 20 appraisal dimensions in the PEACE-Reviews Dataset. Each item is measured on a 7-point Likert scale, indicating the endorsement of the particular appraisal dimension used in the evaluation of the person-product interaction.

B Distribution of PCB classes

Table 3 provides the distribution of the three PCB classes for the whole dataset.

C Model Details and Implementation

We split the dataset up into training, validation, and test sets using 80:10:10 configuration. Since the primary task of predicting PCB is a three-way classification task, we implemented cross-entropy loss for all models to predict PCBs. We used binary cross-entropy loss for appraisal and emotion prediction in multi-task models. Adam optimizer was used with a learning rate of 0.00001. A linear scheduler was also implemented during training. This setting was applied in all models. All models consisting of text inputs are trained for 10 epochs. We found that the performance is stagnant and the fine-tuned BERT models overfit after 10 epochs. On the other hand, models that only use appraisal/emotion ratings are trained from scratch for 2000 epochs, where overfitting occurs after. We implemented separate models for the two PCB variables- a) intention to promote, and b) intention to repurchase. For evaluation, we used the accuracy and the weighted F1 scores.

Baseline models. For the text -> PCB model, we fine-tuned BERT on the dataset and added a

Appraisal	Measure
Accountability-circumstances	To what extent did you think that circumstances beyond anyone's control were responsible for what was happening in the situation?
Accountability-other	To what extent did you think that someone else other than you was responsible for what was happening in the situation?
Accountability-self	To what extent did you think that you were responsible for what was happening in the situation?
Attentional activity	To what extent did you think that you needed to attend to the situation further?
Certainty	To what extent did you understand what was happening in the situation?
Control-circumstances	To what extent did you think that circumstances beyond anyone's control were controlling what was happening in the situation?
Control-other	To what extent did you think that other people were controlling what was happening in the situation?
Control-self	To what extent did you think you had control over the situation?
Coping potential	To what extent were you able to cope with any negative consequences of the situation?
Difficulty	To what extent did you think that the situation was difficult?
Effort	To what extent did you think that you needed to exert effort to deal with the situation?
Expectedness	To what extent did you expect the situation to occur?
External normative significance	To what extent did you think that the situation was consistent with external and social norms?
Fairness	To what extent did you think the situation was fair?
Future expectancy	To what extent did you think that the situation would get worse/better?
Goal conduciveness	To what extent was the situation consistent with what you wanted?
Goal relevance	To what extent did you think that the situation was relevant to what you wanted?
Novelty	To what extent did you think that the situation was familiar?
Perceived obstacle	To what extent did you think that there were problems that had to be solved before you could get what you wanted?
Pleasantness	To what extent did you think that the situation was pleasant?

Table 2: The cognitive appraisal dimensions measured in the PEACE-Review Dataset.

PCB	Low (%)	Medium (%)	High (%)
Intent to repurchase	36.6	21.3	42.1
Intent to promote	32.3	24.1	43.6

Table 3: Distribution of post-consumption behavioral (PCB) intentions in terms of the low (1-2), moderate (3-5), and high (6-7) classes for the whole dataset used in the 3-way PCB classification task.

FFNN at the last layer to predict PCB. For the appraisal/emotion -> PCB models, we trained a neural network that has 3 layers of 1024, 512, and 3 nodes, respectively.

Constrained models. For the Text -> Appraisals/Emotions -> PCB models, the embeddings are obtained after passing to the BERT model. These embeddings are then fed to a FFNN to predict the appraisals/emotions. After which it goes through 3 layers of FFNN of 1024, 512, and 3 neurons, respectively. For the Text -> Appraisals -> Emotions -> PCB model, the appraisal dimensions obtained after passing through the BERT model are fed into a FFNN of 2 layers of 512, and 8, respectively. This 8-dimensional emotion vector is then fed into another FFNN which has 3 layers of 1024, 512, and 3, respectively.

Multi-modal models. The model of each modality was trained separately to predict PCB. After which, the second-to-last layers (excluding the final FFNN layer) of the models are concatenated and passed through a FFNN of 3 layers of 1024, 512, and 3 nodes, respectively.

Multi-task and theoretical models. For the two multi-task models that predict appraisals/emotions and PCB, the embeddings of the text reviews after passing through the BERT model are used to predict either the appraisal or emotions through a 1-layer FFNN. After which the result is concatenated with the BERT embeddings and feed through 2 FFNN of 512 and 3 neurons to predict PCB. For the theoretical model, the BERT embeddings are used to first predict the appraisals through a 1-layer FFNN. After which, the resulting embeddings go through 2 FFNN of 512 and 8 neurons to predict the emotions. The BERT, appraisal, and emotion embeddings are then concatenated and feed through 2 FFNN of 512 and 3 neurons to predict PCB.