Indication of the appraisal in the moment on the appraisal in the memory

Taichi Uno

December 2023

Abstract

1 Introduction

One of the challenges in the field of intelligent systems is to understand how individuals subjectively appraise situations in interactions. Appraisal is defined as the process of evaluating situations based on personal belief, relevance, and significance [19]. This type of information is particularly beneficial for adaptive intelligence systems where they try to change their behaviours and responses such that they fit a user at a specific moment in time. A better understanding of users' perception of a situation is necessary to make the learning process more effective by identifying the specific pattern, phenomenon or social signals that can be utilized for inference of human perception. Traditionally, psychologists try to structurally uncover the process of how one perceives and evaluates a situation from different perspectives to understand different aspects of a situation, such as emotions [19, 20], situation perception [17], or situational interdependence [6].

As stated earlier, appraisal refers to the assessment process of situations based on personal beliefs, relevance, and significance [19]. The cognitive process of evaluation can be described by a set of appraisal variables/dimensions. The theories of appraisal are currently vague in how time plays a role and the dynamics of appraisal due to changes in situations, instead, they focus on classifying emotional reactions or cognitive evaluation of situations based on different dimensions. Therefore, the present research clearly distinguishes two appraisals based on the moment of appraisal and the stimulus events. Appraisal in the

Dimension	Description
Mutual Dependence	
Conflict of Interest	
Future Independence	
Information Certainty	
Power	

Table 1: Dimensions of situational interdependence [6]

moment is defined as appraisal that is carried out in the moment of events. It is closely related to the concept of appraisal dynamics, which captures evolving situations and evaluates appraisal variables dynamically [13]. On the other hand, appraisal as remembered is defined as appraisal of past events. This is attained from reappraisal of the past event by evaluating the mental recollection of happened events or recalling appraisal variables. The cognitive process of these two are different and it is expected to be dependent on how the recall was initiated (e.g. how the triggering prompt is formulated [**?**].

As an example of different aspects of appraisal, situation perception is a type of cognitive process that interprets and understands situations subjectively based on an individual's experiences, beliefs, and expectations [22]. The expressed emotions and behaviour can be indicative of how one perceives a situation [8, 9]. When it comes to different aspects of evaluation, as one of the facets of situation perception, there is situational interdependence, which particularly highlights the perceived interdependence in a certain situation. The situational Interdependence Scale (SIS) proposed by Gerpott et al. [6] provides a way to quantify subjective (or perceived) interdependence based on the five dimensions (mutual dependence, conflict of interest, future independence, information certainty, and power) that articulate individuals' perceptions of their interdependence.

Another specific example of different aspects of cognitive appraisal, the component process model (CPM) of appraisal is inherited from the appraisal theory of emotion, in which specific configurations of those dimensions in combination define emotional labels, such as sadness, happiness and more [18, 20]. The components consist of relevance, implications/consequences, coping potential, and norm compatibility. Individual differences in perception make its process subjective, which leads to diverse outcomes when evaluating the same stimulus event. This variation occurs across different individuals and within the same individual over time, influenced by changes in perception, and the situation [3]. The temporal distance between the moment of appraisal and the moment of the stimulus event also has an impact on its evaluation [23]. The outcome of the appraisal can be inferred from observable cues, thus one can estimate the internal cognitive states of a person by exploiting facial expressions [11], verbal contents [], verbal cues [12], or physiological signals [].

2

Dimension	Description
Relevance	
Implications/Consequences	
Coping potential	
Norm compatibility	

Table 2: Components of component process model from appraisal theory of emotion [20]

Having explained what appraisal is and specific examples of different constructs, modelling appraisal in a situation can be inherently challenging. One major obstacle lies in the difficulty of precisely determining the extent to which each component of the appraisal process is triggered from observable signals, regardless of the specific constructs of appraisal being considered. This is due to the complexity of the cognitive system, which complicates both the data collection and annotation processes [18].

Another problem is that the appraisal process is known to be temporally dynamic and sequential, presenting challenges in determining an optimal frequency for data collection [22]. The temporal distances between the appraising stimulus event and the reporting moment can significantly influence participants' appraisal processing [3]. When observing conversations to extract indications of appraisal from observable cues for modelling, it is important to appropriately slice interactions [15]. Therefore, ensuring conceptual validity and computational feasibility, as well as understanding the assumptions and limitations, are imperative.

Lastly, while it may seem intuitive to assume a direct correlation between the appraisal in the moment and the appraisal as remembered, this connection remains ambiguous. One primary reason for this ambiguity is the conceptual disparity between these two forms of appraisal. The appraisal in the moment deals with the ongoing situation and stimuli, whereas the latter evaluates the remembered interactions. It is necessary to distinguish between these two because the temporal gap between the stimulus and the moment of appraisal influences the assessment[23]. However, current research has not extensively delved into this differentiation; instead, it has primarily focused on the appraisal as remembered. For instance, researchers often utilize post-interaction questionnaires, treating them as an appraisal of the situation.

Acknowledging these gaps in research and limitations, this research is going to investigate the following research questions to uncover the relationship between the appraisal in the moment and the appraisal of remembered interactions.

1. How can appraisal in the situation extracted from verbal contents be uti-

lized for inference of appraisal as remembered?

2. How does the context of appraisal as remembered impact its inference by appraisal in the situation extracted from verbal contents?

To address these questions, two datasets that contain video recordings of dyadic conversations and a post-conversation questionnaire for evaluating the interaction are used, essentially representing the appraisal as remembered. With these datasets, two modelling steps are carried out. Firstly, appraisal in the moment during a conversation is modelled as a time-series signal. The conversation is segmented appropriately to maintain conceptual coherence. At each segmented time window, the appraisal in that specific moment is estimated by leveraging observable signals (video recordings) known to indicate appraisal from literature. Subsequently, this time-series data, representing the evolving appraisal in the moment throughout the conversation, is fed into a machine learning model. This model aims to predict and estimate how one appraises the remembered conversation. The resulting predictions are then analyzed to understand the correlation between the appraisal in the moment and the appraisal as remembered. The details of the datasets and methodologies are discussed later in this paper.

One of the hypotheses of the first research question is that there exist certain dimensions of an individual's appraisal that are closely linked to indicating their appraisal as remembered. Gerpott et al. [6] investigated how a participant felt during an interaction, using four emotion labels (anger, disgust, happiness and sadness) in a situation and tried to find a correlation with the situational interdependence, which represent appraisal as remembered. The report of the emotion labels from participants can be seen as the appraisal in the moment in terms of emotion labels, as they are asked to recall how they felt during interactions instead of reevaluating the interaction. The results support that each dimension correlates with specific emotions. For example, more participants felt sad in a situation where they perceived the situation with a higher conflict. Thus, we can hypothesize the claim that certain dimensions of an individual's appraisal in the moment can be indicative of certain dimensions of appraisal as remembered.

In terms of the time variance, the peak-end rule states that people tend to evaluate the overall experience of an event based on its peak and the end rather than based on the sum or weighted average of overall experience [10]. This allows us to hypothesise that the peak and the end of one's appraisal in the moment have more impacts on one's appraisal as remembered compared to other parts of the conversation.

2 Related Work

2.1 Appraisal Modelling

Appraisal in the moment can be seen as the dynamics of appraisal during an interaction as defined earlier. In terms of capturing the dynamics of the internal states of a person, previous attempts are limited to capturing the dynamics of emotion during interactions []. It is also questionable if emotional expressions, which sometimes are used interchangeably in the context of emotion inference, are truly reflecting the internal states of one. To our best knowledge, the present study is the first attempt to model the dynamics of appraisal with regard to the situational interdependence scale.

There are several attempts to quantify the components of the appraisal using large language models (LLM) [1, 5, 21, 25, 26]. LLM has a huge potential in analysing textual content. This is realized by the emergence of Transformer [24]. This architecture has achieved great success in natural language processing, relying on self-attention mechanisms to process input data in parallel, which makes it possible to capture long-range dependencies in a sequence of data efficiently.

When it comes to applying LLMs to appraisal modelling, one of the examples is research by Zhan et al. [26]. They investigated the usage of LLMs to predict dimensions of emotional appraisal (CPM) with zero-shot learning using the CovidET dataset [27], which consists of covid-related posts on Reddit. These posts are likely more closely related to the appraisal of remembered events, as it's improbable that individuals are posting on Reddit while the event is ongoing. Similar approaches have been adopted, employing text descriptions of specific scenarios [1, 21, 25]. This research demonstrates the potential of LLMs in extracting dimensions of appraisal from textual data. However, to the best of our knowledge, there have been no attempts so far to explore other aspects of appraisal, such as situation perception or situational interdependence.

While much of the research on employing LLMs for appraisal modelling focuses on short texts or scripted dialogue scenarios in controlled environments or social media posts, there has been limited exploration specifically targeting appraisal within conversational settings. One notable study conducted by Feng et al. [5] addresses this gap by analyzing each utterance to identify emotions within a conversational context using a set of emotional labels. The study found that, compared to state-of-the-art supervised approaches, LLMs exhibited lower performance in zero-shot learning scenarios. However, performance notably improved with instruction-following demonstrations, indicating the effective utilization of prompts. While the research highlights that LLMs still have some way to go to match the performance of state-of-the-art supervised models, it

underscores their greater generalizability to other natural language processing tasks and robustness to errors in automatic speech recognition. Overall, this study unveils the potential of employing LLMs within conversation settings to infer abstract phenomena, such as emotion, from conversational text data.

2.2 Predicting Appraisal

Research by Dudzik et al. [4] investigated recognizing perceived situational interdependence in face-to-face negotiations by exploiting facial expressions, upper body behaviour and non-verbal vocal behaviour. They used the aforementioned SIS to measure situational interdependence and built a model based on the Ridge Classifier to analyse multivariate time series of those behavioural features. Their main discovery is that people's behaviour seems to be predictable of the perceptions of conflict of interest and power, while the conversation partner's behaviour is for conflict of interest, future independence and information certainty. Their research focuses on predicting appraisal as remembered, which is recorded via post-interaction reports, from behavioural features directly instead of interring appraisal. Also, this research did not take the temporal dynamics into account, instead, they constructed "aggregated" features from the behavioural features by using ROCKET (Random Convolutional Kernel Transformation) [2], which lacked the implications of the temporal dynamics of how each behavioural feature in the moment has significance to the situational interdependence.

Since the appraisal in the moment is time-variant, preserving the temporal dynamics when predicting appraisal as remembered becomes crucial. This is why the Long Short-Term Memory network (LSTM) is commonly employed in the field of affective computing [16]. The LSTM architecture has hidden internal states that function as "memory," retaining information from the past to be utilized in the future. This inherent capability of LSTMs to capture long dependencies over time aligns well with predicting appraisal as remembered.

In the field of psychology, there is a widely known phenomenon called the "Peak-end rule", which states that people tend to evaluate the overall experience of an event based on its peak and the end rather than based on the sum or weighted average of overall experience [10]. (definition of peak) This theory implies that the appraisal as remembered of interaction is hugely or solely based on the peak and the end of the appraisal in the moment. (ADD AN EXAMPLE APPLICATION OF PEAKEND RULE)
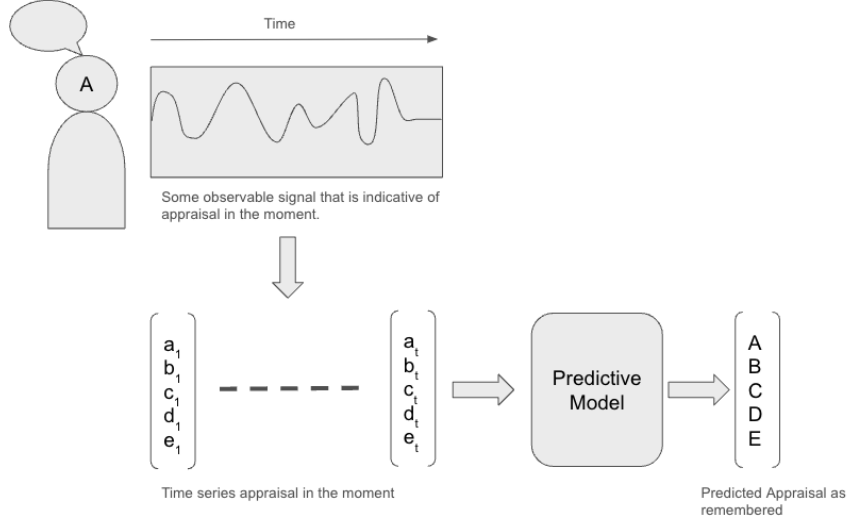
Figure 1: Schematic Representation of Situational Interdependence Prediction Model

3   Methodology

This research involves creating a model of the appraisal process to predict the appraisal as remembered in terms of situational interdependence. The schematic representation of the prediction model is described in Figure 1. The diagram illustrates Person A in a conversation, showing that the appraisal in the moment is estimated from some observable signals that indicate the appraisal. For example, Person A can be smiling (facial cue) while talking (verbal cue) and making hand gestures (behavioural cue), where each of these cues could potentially indicate a different aspect of appraisal. The modelled Person A's appraisal in the moment is the input for the predictive model, generating predicted values for appraisal as remembered. Subsequently, the outcomes are compared with actual data to test the hypotheses.

3.1   Dataset

This research employs two distinct datasets: the PACO dataset [14] and the Negotiation dataset [4]. Although they were collected for different purposes, both datasets provide comparable information relevant to our research aims. This includes audio-visual recordings of participants engaging in conversations from similar angles, along with post-conversation SIS measurements. In short, con-

versation records under the listed five contexts were present in the two datasets.

1. Joint trust task setting (TRUST)
2. Joint competence task setting (COMP)
3. Feud Task setting (FEUD)
4. Low conflict Negotiation setting (LConf)
5. High conflict Negotiation setting (HConf)

The PACO dataset [14] was initially developed to model partner selection and explore its relationship with human impressions during social interactions [14]. The dataset includes recordings of 3-minute online conversations between two individuals over three different contextual settings. The settings include performing 1) a joint trust task (TRUST), which is a cooperative decision-making task with mixed motives, 2) a joint competence task (COMP), where the outcome depends on the competence of two individuals, and 3) a Feud Task (FEUD), where facilitation of collaboration and active communication are important for the best outcome. Half of the participants were selected for performing the joint trust task and the other for the joint competence task. The feud tasks were performed by all participants. Each participant was asked to have one-on-one online video conversations for the two tasks with 3-5 other participants and to fill out a questionnaire including a 10-item version of questions from SIS after each conversation.

The other dataset is the Negotiation dataset [4]. It contains recordings of negotiation conversations in face-to-face settings between two individuals. One of the participants was assigned to the "applicant" and the other to the "HR manager", where the latter has more power in the setting. The conversations are assigned to one of two conditions in terms of the intensity of conflict. The low conflict situation (LConf) has more options, five out of five options, which leads to a preferred outcome of both of them while the high conflict situation (HConf) has only one option and the rest leads to a consequence where one's best outcome is the other's worst outcome. Participants needed to carry out the 5-minute negotiation three times and fill out a survey that included a 10-item version of SIS, the same set of questions with the PACO, after each session. After the third session, they had to come up with an agreement and failure to do so had a consequence of either 60%, 40%, or 20% of an even split.

3.2   Modelling appraisal in the moment

As mentioned earlier in Section 2.1, the usage of LLM has been attempted previously in modelling emotional appraisal from textual content. This research

Prompt type	Prompt template
Task Definition	Consider the following list of concepts, called "Situational Interdependence" : [Mutual Dependence : (definition), Conflict of Interest : (definition), ...]
Query	Given the dialogue history between PersonA and PersonB : [PersonA: ..., PersonB: ..., ...], what is the "Situational Interdependence" in the next utterance of PersonA "..."?

Table 3: Templates of prompts for modelling appraisal in the moment.

is also going to follow this trend of using LLM to model the appraisal in the moment by looking at the textual conversation contents. Since the dimensions of the appraisal as remembered are in terms of the situational interdependence scale (SIS), the modelling of appraisal in the moment will also be in terms of the SIS.

In formulating the prompt to model appraisal in the moment, we adopt a methodology similar to that of Feng et al.[5]. Their study is particularly relevant as it specifically addresses conversation settings. We incorporate two types of prompts: task definitions and queries. It's worth noting that Feng et al. also experimented with in-context learning; however, due to the absence of ground truth labels for appraisal dimensions in the moment in the datasets, we are unable to utilize this aspect in our setup. The task definition involves outlining each dimension of situational interdependence and describing the contextual setting.

Meanwhile, the structure of the query prompt draws inspiration from Feng et al.[5], with a focus on the process of appraisal estimation for each new utterance while maintaining the history of the conversation as context for future queries. The utterance of the conversation partner is also added to the conversation history. The example prompt is shown in Table 3.

3.3   Predictive model

The predictive model takes the time-variant appraisal in the moment in terms of situational interdependence as its input and the predicted values of appraisal as remembered in terms of situational interdependence as its output. To maintain the time-variant nature of appraisal for prediction, we utilize the Long Short-Term Memory network (LSTM). The LSTM's ability to capture long dependencies over time is well-suited for our task. To address RQ1, all instances of data are utilized, whereas for RQ2, a new network is trained for each contextual

setting.

4 Discussion

4.1 Limitations

A combination of different modalities has more information regarding modelling appraisal in the moment [7].

References

[1] Joost Broekens, Bernhard Hilpert, Suzan Verberne, Kim Baraka, Patrick Gebhard, and Aske Plaat. 2023. Fine-grained Affective Processing Capabilities Emerging from Large Language Models. arXiv:2309.01664 [cs.CL]

[2] Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. 2021. MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & amp; Data Mining (KDD '21)*. ACM. https://doi.org/10.1145/3447548.3467231

[3] Bernd Dudzik and Joost Broekens. 2023. A Valid Self-Report is Never Late, Nor is it Early: On Considering the "Right" Temporal Distance for Assessing Emotional Experience. arXiv:2302.02821 [cs.HC]

[4] Bernd Dudzik, Simon Columbus, Tiffany Matej Hrkalovic, Daniel Balliet, and Hayley Hung. 2021. Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 121–130. https://doi.org/10.1145/3462244.3479935

[5] Shutong Feng, Guangzhi Sun, Nurul Lubis, Chao Zhang, and Milica Gašić. 2023. Affect Recognition in Conversations Using Large Language Models. arXiv:2309.12881 [cs.CL]

[6] Fabiola Gerpott, Daniel Balliet, Simon Columbus, Catherine Molho, and Reinout de Vries. 2017. How Do People Think About Interdependence? A Multidimensional Model of Subjective Outcome Interdependence. *Journal of Personality and Social Psychology* In press (09 2017). https://doi.org/10.1037/pspp0000166

[7] Judith A Hall and Marianne Schmid Mast. 2007. Sources of accuracy in the empathic accuracy paradigm. *Emotion* 7, 2 (2007), 438.

[8] Ursula Hess, Jonas Dietrich, Konstantinos Kafetsios, Shimon Elkabetz, and Shlomo Hareli. 2020. The bidirectional influence of emotion expressions and context: emotion expressions, situational information and real-world knowledge combine to inform observers' judgments of both the emotion expressions and the situation. *Cognition and Emotion* 34, 3 (2020), 539–552.

[9] Kai T Horstmann and Matthias Ziegler. 2019. Situational perception and affect: Barking up the wrong tree? *Personality and Individual Differences* 136 (2019), 132–139.

[10] Daniel Kahneman. 2000. Evaluation by moments: Past and future. *Choices, values, and frames* (2000), 693–708.

[11] Susanne Kaiser and Thomas Wehrle. 2001. Facial expressions as indicators of appraisal processes. *Appraisal processes in emotion: Theory, methods, research* (2001), 285–300.

[12] Reza Lotfian and Carlos Busso. 2019. Building Naturalistic Emotionally Balanced Speech Corpus by Retrieving Emotional Speech from Existing Podcast Recordings. *IEEE Transactions on Affective Computing* 10, 4 (2019), 471–483. https://doi.org/10.1109/TAFFC.2017.2736999

[13] Stacy C Marsella and Jonathan Gratch. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10, 1 (2009), 70–90.

[14] Tiffany Matej Hrkalovic. 2022. Designing Hybrid Intelligence Techniques for Facilitating Collaboration Informed by Social Science. In *Proceedings of the 2022 International Conference on Multimodal Interaction* (Bengaluru, India) *(ICMI '22)*. Association for Computing Machinery, New York, NY, USA, 679–684. https://doi.org/10.1145/3536221.3557032

[15] Nora A Murphy and Judith A Hall. 2021. Capturing behavior in small doses: A review of comparative research in evaluating thin slices for behavioral measurement. *Frontiers in psychology* 12 (2021), 667326.

[16] Desmond C Ong, Zhengxuan Wu, Zhi-Xuan Tan, Marianne Reddan, Isabella Kahhale, Alison Mattek, and Jamil Zaki. 2019. Modeling emotion in complex stories: the Stanford Emotional Narratives Dataset. *IEEE Transactions on Affective Computing* 12, 3 (2019), 579–594.

[17] John F Rauthmann, David Gallardo-Pujol, Esther M Guillaume, Elysia Todd, Christopher S Nave, Ryne A Sherman, Matthias Ziegler, Ashley Bell Jones, and David C Funder. 2014. The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology* 107, 4 (2014), 677.

[18] David Sander, Didier Grandjean, and Klaus R Scherer. 2005. A systems approach to appraisal mechanisms in emotion. *Neural networks* 18, 4 (2005), 317–352.

[19] Klaus R Scherer. 2005. What are emotions? And how can they be measured? *Social science information* 44, 4 (2005), 695–729.

[20] Klaus R Scherer. 2013. The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion review* 5, 2 (2013), 150–162.

[21] Ala N Tak and Jonathan Gratch. 2023. Is GPT a Computational Model of Emotion?. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.

[22] Jurriaan L Tekoppele, Ilona E De Hooge, and Hans CM van Trijp. 2023. We've got a situation here!–How situation-perception dimensions and appraisal dimensions of emotion overlap. *Personality and Individual Differences* 200 (2023), 111878.

[23] Yaacov Trope and Nira Liberman. 2003. Temporal construal. *Psychological review* 110, 3 (2003), 403.

[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[25] Nutchanon Yongsatianchot, Parisa Ghanad Torshizi, and Stacy Marsella. 2023. Investigating Large Language Models' Perception of Emotion Using Appraisal Theory. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. 1–8. https://doi.org/10.1109/ACIIW59127.2023.10388194

[26] Hongli Zhan, Desmond Ong, and Junyi Jessy Li. 2023. Evaluating Subjective Cognitive Appraisals of Emotions from Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 14418–14446. https://doi.org/10.18653/v1/2023.findings-emnlp.962

[27] Hongli Zhan, Tiberiu Sosea, Cornelia Caragea, and Junyi Jessy Li. 2022. Why Do You Feel This Way? Summarizing Triggers of Emotions in Social Media Posts. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 9436–9453. https://aclanthology.org/2022.emnlp-main.642