1. INTRODUCTION

1.1 Motivation

The major challenge in Affective Education is to accurately recognize and respond based on learner’s affective state [2] and to unify experiments and findings from various disciplines and create a cohesive picture [3].

Affect is a construct used in psychology to describe unspecified feelings; the superordinate umbrella of constructs involving emotion, mood, and emotion-related traits [4]. Though there are finer distinctions between affect and emotions, research articles refer to these terms interchangeably.

Affect impacts learning. Underlying mechanisms of how affect impacts learning is still an open problem [5]. To understand this relationship, accurate detection of affect is essential. Although the role of affect on learning is established through various studies, most of the current research treat cognition and affect on two different dimensions, and at best only loosely coupled with cognition and indirectly mediated via Personality Traits, and Motivation.

Relationship between learning and affect is deeper and more intertwined than what the current learning theories advocate. When learning proceeds through different phases, it evokes associated affect (for example when diagnosing a problem for a sustained period, the student may have to endure frustration) [5]. Figure 1 depicts Affect cycle during learning.

As an illustration, the student may be in a “state of stuck” or “state of flow” (Though the behavioural correlates could be similar when the student is solving a difficult problem – the student should be encouraged to persevere through when in state of flow but intervene when in state of stuck [6]. Each student is unique, to cater to the unique needs of the students the learning needs to be tailored, so deep personalization is essential - machines need to understand affect to respond effectively.

1.2 Limitations of previous works

1.2.1 Challenges in emotion detection

Currently, the accuracy of emotion detection is low [7]. Emotions are not universal, and there is inter and intrapersonal variability. Distinctions reported in one study are rarely precisely replicated in another [8].

Figure 1: Affect cycle during learning and playing chess [5]
Detection of affect is a challenge as it involves subjective aspect [9] and self-reporting is not very accurate [9]. Besides, the task-complexity also confounds the results.

The accuracy of Automatic emotion detection systems is low due to lack of unique-fingerprints & visible co-relates of emotion using facial expressions [8]. Valuable Gaze Metrics, related to engagement, from chess studies, can be used to complement and improve emotion detection [10].

1.2.2 Lack of Integrated framework – affect, cognition and performance

Findings from multiple disciplines need to be Unified and create a cohesive picture [3]. (Judgement and Decision making – “Incidental-influences (Incidental affect - feelings at the time of decision making not normatively relevant for deciding) impact decision making [4], Education (Students have to wade through varying emotions that are integral with learning successfully and not give-up up mid-way) [5], learning agents (agents that don’t just respond to valence and arousal but factor in the student engagement – whether in-flow or stuck and then respond accordingly) [11], Affective research (Recent findings have shown that Emotions are constructed and not universal as believed earlier [8] – This brings to suspicion emotion detection systems that rely on unique fingerprints for detecting emotions), Chess studies (chess players in order to deal with complexity use emotions “as information” to narrow down their options for further exploration) [12, 13].

1.2.3 Scalability and repeatability challenges

Lack of framework for Scalability, repeatability and reproducibility in chess studies.

Prior empirical studies [12, 13] were conducted in the lab using specialized equipment, and experts select the tasks to be used in the subsequent experiments – this model though very useful for understanding the underlying mechanisms cannot be scaled.

1.2.4 Lack of quantifiable affect mitigation strategies

Lack of Quantifiable results related to short term strategies, in mitigating the adverse impact of incidental affect.

Though, Numerous studies have established the positive impact of mindfulness practice (requiring long and deliberate practice) on Test performance [14]. None have so far, explored the seemingly simple strategies like “time-out” and awareness of one’s state on task performance to reduce the impact of transient emotion dynamics and its influence on performance. Even, Short-term mindfulness intervention reduces the adverse attentional effects [15].

Chess gives us a unique opportunity to test this hypothesis [Refer to Figure 6].

A comprehensive meta-analysis has shown that Chess skill correlated positively and significantly with fluid reasoning, comprehension-knowledge, short-term memory, and processing speed [16].

1.3 The Structure of this Paper is as follows:

- Section 1 presents the background and goal.
- Section 2 introduces the proposed framework and details of the architecture.
- Section 3 and 4 describes the method and results of the pilot experiment (chess puzzles) to validate the framework.
- Section 5 provides a discussion.
- Finally, section 6 explains the conclusion and future works.

1.4 Objectives of our study [Figure 2]

Create a scalable framework so that experiments based on field trials can be conducted (beyond Lab). To validate the framework, we study the impact of regulation strategies to mitigate the influence of incidental affect on learning outcomes (Using chess Puzzles). Based on the results, derive a set of policies for improving the effectiveness of learning outcomes (Both short term and long-term affect regulation strategies).

2. PROPOSED FRAMEWORK

2.1 Framework

This section presents the framework for adapting the tasks based on the current strength and affect.

Figure 3 shows High-level architecture and Figure 4 elaborates on the affect management module. Though this framework is currently tailored for chess, we believe this architecture can be adapted for other domains easily.

As an illustration of adaptability of chess frameworks and their relevance in education: CHREST [17], a chess framework, was developed initially to explain expert behaviour in chess, a domain that draws upon a variety of cognitive abilities, such as perception, memory, decision making, and problem-solving. However, this framework is now able to explain phenomena in verbal learning.
the acquisition of language, developmental psychology, implicit learning, and concept formation [18].

Also, our primary objective is first to understand the underlying mechanisms during game play and select effective response strategies for the user – so the user can learn effectively. The other objective is to test out the recent findings in chess studies that posit – players use “emotions as information” to deal with the complexity of games like chess [12]. Towards this end, we need a framework that will help test out hypothesis and useful for further research.

Some of the significant additions to the earlier framework – NITM [19] are the affect management module & adaptive task management. The earlier framework was designed keeping regular game-play in mind and adapting only the behaviour of the computer (by ways of speech) to impair the gameplay of the opponent, whereas our focus is on the task selection and appropriate response strategies based on affect & skill level to help the user learn optimally.

Figure 3 describes the main components of the affect management system.

2.1.1 Adaptable Framework – Architecture

Affect management system is tightly integrated with the system and coupled with Chess Gaming Engine. For details on other components refer - [19]. Marked blue is the key interface changes.

2.1.2 Affect management Module – Architecture (Figure 4)

Key Inputs and outputs to this system are described below:

2.1.2.1 Affect detection and response

Self-report or feedback by a human tutor using VAS scale – Visual Analog scale [9,11]

2.1.2.2 Emotion Detection

Automatic emotion detection systems, used in research, associate Action Units (i.e. movement involving the minimal number of facial muscles) with six (or discrete number) prototypical facial expressions of emotion.

The systems are trained on video databases made of dynamic emotional facial expressions as tagged by annotators [7]. However, Emotions are not universal, and there is inter and intrapersonal variability. Distinctions reported in one study are rarely replicated in another [7].

Using Action Units directly (Intensity and change) in addition to the existing categories of emotions will minimize the inter-person and intra-person variability.

2.1.2.3 Emotional Dynamics

“Balance between two opposing forces, the tendency to resist change and the tendency to regulate to achieve optimal fit continuously, determines to a large extent how an individual’s emotions unfold across time” [20]

Using experience sampling the following can be calculated [21].

- Emotional inertia – Carryover from the previous moment
- Emotion augmentation and blunting – predicts another emotion
- Emotion (component) – co-variation (differentiation/granularity)
- Duration – Feature of an emotional episode
- Intensity profile shape – fluctuations in emotions
- Emotional variability – Feature of emotional trajectory

2.1.2.4 Eye Gaze Metrics [22] – Possible trajectory

- The total number of fixations – Efficiency of searching or engagement.
- The total number of gazes – Complexity of the inferential process.
- Mean fixation duration – Complexity or engagement.
- Fixations on AOIs – Element importance or noticeability.
- Gazes on AOIs – Element importance or noticeability.
- Viewing Time on AOIs – Information content, complexity, or engagement.
- Time to First Fixation on AOI – Attention-getting properties.
- Scan Path and Fixation time and duration (Chess) – differences between Expert vs. Novice can be understood.

2.1.2.5 Posture Detection – Body Agitation, Volume, and self-touching [13].

2.1.2.6 Pupil Detection(size) – Task complexity [13].

2.1.2.7 Current Skill of user – ELO Rating of User.

2.1.2.8 Task complexity – ELO score of the Task.

2.1.2.9 Feedback Intervention – Refer to Figure 8.
2.1.3 “Affect & Emotion” Effective Response Strategies

2.1.3.1 Refer to Figure 5 and 6.

We outline the Affect Integrated Model of Decision-making (AIMD) and highlight the influence of incidental emotions on decision making and then contrast it to the corresponding flow diagram for chess play—showcasing the importance of emotions in reducing the complexity.

According to AIMD model and “Atlas of emotions”, understanding the triggers of emotions and taking “time out” to be an effective way in responding to emotions [4, 23].

“The simplest strategy for minimizing emotional magnitude is to let time pass before making a decision. Emotions are short-lived (Levenson 1994). Facial expressions are fleeting (Keltner et al. 2003), and physiological responses quickly fade (e.g., Mauss et al. 2005)” [4].

2.1.3.2 Refer to Figure 7 and 8.

We highlight why “one size fits all” strategy will not work for emotion regulation and detection—emphasizing the need for deep personalization and holistic approach.

In contrast to the classical view of emotions [24], which states that each emotion has a unique fingerprint and can be identified uniquely—which form the basis of automatic emotion detection systems. Theory of constructed emotions [8] has demonstrated that emotions are constructed depending on the context and they are not unique. The expression of each emotion also varies from time to time. Consequently, emotions elicit action tendencies.

Hence the emotion identification system must be built factoring in the unique expression for each person. “People attend to their feelings as a source of information, with different feelings providing different types of information. Whereas feelings elicited by the target of judgment provide valid information, feelings that are due to an unrelated influence can lead us astray” [25].

For the insights, from prior research on mindfulness, to be accessible for general public Dr Paul Ekman and his team had compiled the Atlas [23]—“The interactive map” for people to be aware of the triggers of emotions and effective ways to mitigate the adverse effect of these emotions. He suggests understanding the triggers of emotions and taking “time out” to be an effective way of responding to emotions. This is a valuable guide for people to understand their emotions and consequences. Once the participants are informed about the easy to understand website, subsequent experiments can be conducted relating to the influence of emotions - This will reduce the training time.

2.1.3.3 Refer to Figure 9.

Rationale behind the experimental design for studying the influence of incidental affect.

Chess is a very complex game. (Shannon showed a calculation for the lower bound of the game-tree complexity of chess, resulting in about $10^{120}$ possible games, to demonstrate the impracticality of solving chess by brute force, in his 1950 paper “Programming a Computer for Playing Chess”. As a comparison, the number of atoms in the observable universe, to which it is often compared, is roughly estimated to be $10^{80}$ - Which is orders of magnitude lower). The challenges faced during Learning are similar – Students must successfully navigate through the information overload.
Despite the enormous complexity, Chess players perform very well, in addition to making accurate moves that too under time constraints [26].

Recent experiments [12] have shown that one possible hypothesis that enables chess players to deal with complexity is by relating emotions to game situations and simplifying the position so that the ensuing positions can be evaluated in detail. Hence, we hypothesize:

H1: Incidental affect is negatively related to task performance [Refer to Figure 6].

H2: Time-delay & being aware of one’s state between tasks will decrease the influence of Incidental affect (leading to an increase in Task performance).

3. METHOD

3.1 Participants

The participants are hobbyist chess players and University students from Hyderabad, India. They volunteered to participate in the experiment. Familiarity with chess rules and Chess notation was a mandatory pre-requisite to Participate.

In addition to the above criteria, to ensure the reliability of rating estimate, the following criteria were enforced:

Active lichess.org [27] account and rating (ELO [28] rating evaluated by GLICKO-2 [29,30]) with minimum of 100 games in Classical/Blitz/Rapid format and “1200 ELO” in Puzzles-score (having solved minimum of 100 Puzzles) (This criterion was enforced to ensure reliable rating estimate during the experiment).

Total of 92 Participants responded. 12 Students were eliminated (after applying the reliability criteria). Participants were male (age 19.1 (1.5)). The reason for choosing male participants and restricting the age group was to limit the confounding variables influencing the results. Besides, we did not receive a significant response from Female Participants.

The Participant rating evaluation was the following: Puzzle Rating (ELO) 1520 (100); Min ELO = 1276 and Max ELO = 1679. No emotional stimuli were used. Informed written consent was obtained from all participants, and prior approval was taken to conduct the experiment [31].

3.2 Platform

Lichess Platform was used to conduct the experiment. The puzzles were automatically selected based on the current strength of the participant (Puzzle ELO). The puzzles were randomly drawn from a large pool of games played on this platform by member participants (each puzzle is randomly drawn from a large pool (670 Million Games and close to 250000 games/Rating) [lichess.org]. The Participants Task is to find the Best move. The puzzles are not restricted to mate in N but also include finding the best moves – replicating the Game scenario (Refer to Figure 10 as an illustration).

The Rational for choosing to draw games from such a sizeable Random pool is to include only games played by Human Participants (instead of randomly generated puzzles by machine) [32], mitigating the selection and familiarity bias (and enabling this design of experiment reproducible/scalable for future research). In contrast, Prior research on chess studies [12,28], though having its advantages, have used carefully selected chess problems by an expert. Hence this approach has a limitation concerning scalability and generalizability.

3.3 Rating System

Glicko-2 rating system (in contrast to the FIDE rating system used in Tournament play) is used for assessing a player’s strength in games such as chess, and Go.

Ratings are calculated using the Glicko-2 rating method [30]. This is a very popular rating method used by chess platforms. This rating system uses confidence intervals when representing the rating. When a player is starting the rating will change dramatically, but after several games, the confidence interval will narrow, and the number of points gained/lost will lessen. Another advantage of this system compared to the FIDE system is that the confidence interval will adjust depending on not just on the current score but related to past performance.

![Figure 9: Experimental Design for effective intervention strategies. Also used for Validating the hypothesis that incidental affect/emotions are tightly coupled to cognition and quantify the influence of intervention strategies on performance.](image)

![Figure 10: Illustration of Puzzle with performance score](image)
3.4 Setup and Task Description

During the experiment, data were recorded from Screen capture, built-in webcam, user clicks, Tobii-Bar (built-in).

Our Test setup though similar in some respects described in [12] differed, in the following ways:

To avoid calibration and for future scalability, off-the-shelf commercially available equipment with an integrated infrared camera was used.

Hardware: Alienware 17 R4 with built-in Tobii Eye Tracker [33].

Software: Opensource OpenFace [34] Software was used for post-processing.

First, 80 Players were randomly divided into two groups – Control and Test (Forty Each).

Forty students worked on Test setup first while another set of forty students worked on Control setup.

Next, the roles were reversed in the next round.

Eventually, at the end of the tests, the set of Eighty people participated in both the setups (Test and Control). This was done to reduce familiarity bias and compare results between intervention and no-intervention.

Problem selection: Randomly selected from 640 Million puzzles (250000 games/Rating) depending on the participants’ current strength.

Refer to Figure 9 for the architecture of the experimental design.

3.5 Control Group

- Chess Puzzles [27] were given one by one to the participants to be solved in 30 minutes.
- When the participant completes one puzzle, the next puzzle is automatically selected (Depending on the result, the difficulty of the puzzle is automatically adjusted to match the rating of the player)
- An overall score (Starting and Final ELO), Time spent on each problem is recorded.

3.6 Test Group

- Chess Puzzles [21] were given one by one to the participants to be solved in 30 minutes.
- When the participant completes one puzzle, unlike Control Condition a mandatory break is given between puzzles.
- Break duration is variable depending on the affective state of the participant (But for this experiment we have fixed the duration to 30 seconds). The participants are told to relax (and if possible, be aware of one’s state during this time)
- Next puzzle is automatically selected (depending on the result the difficulty of the puzzle is automatically adjusted to match the rating of the player)
- An overall score (Starting ELO and Final ELO), Time spent on each problem is recorded.

3.7 Explanatory and Dependent variables

The independent variable is the intervention strategy (Time duration between puzzles & Instruction to Relax), and Dependent variable is the Performance (ELO).

The Mean of the ELO before and after the Puzzles, for both Control and Test Conditions, was calculated.

Paired T-Test was performed to determine the effectiveness of the strategy.

4. RESULTS

4.1 Hypothesis Evaluation

H1<sub>a</sub>: Incidental affect is negatively related to task performance [Refer to Figure 6].

H2<sub>a</sub>: Time-delay & being aware of one’s state between tasks will decrease the influence of Incidental affect (leading to an increase in Task performance).

A Two-sample Paired t-test was conducted to compare the performance (ELO change) of chess participants in “time-delay between tasks” (Test) and “no time-delay between tasks” (Control) conditions.” (Refer to Figure 12)

“There was a significant difference in the scores (Mean ELO) for Control (M=1535.4, SD = 78.01) and Test (Mean ELO) (M = 1591.1, SD=49.7) conditions; t(79) = -5.28 p<0.001,95% CI for mean difference 34.78 to 76.80, r = .05, Cohen’s d=0.591.” on an average The Test group was 55.79 ELO points higher due to intervention.

Also, 57(71.5%) participants improved their performance 21(26.25%) decline in performance and 2 participants did not have any change in rating (Refer to Figure 11).

4.2 Selection of Puzzles (Refer to Figure 13)

The number of puzzles solved with respect to the Relative strength of puzzles compared to participants strength (ELO).

Much weaker: -200(0.9%), Weaker: -100(5.3%), Similar (85.9%), Stronger: +100(6.4%), Much stronger: +200(1.3%).

Figure 11: Rating change between Control and Test conditions (Y-axis “mean ELO” of participant)
4.3 Fluctuation in Action Units while solving Task

There was rapid fluctuation in emotions (AUs) during the experiment - Facial Action Unit Intensity and Number of Action Units detected while solving the puzzle.

The OpenFace system can detect the intensity (from 0 to 5) of 17 AUs & presence/absence (0 = absent, 1 = present) of 18 AUs. (Refer to Figure 13)

We did not observe any statistically significant differences in Fixation time, Fixation duration, number of saccades between Control and Test conditions.

5. DISCUSSION

Affect impacts learning. In order to respond effectively based on affect, we need a holistic framework as the well-known quote states - "The whole is other than the sum of the parts". Towards this goal, we have described an architecture for an integrated framework that simultaneously factors in affect, cognition, regulation strategies and performance (we have built on earlier framework NITM [19]); highlighted some of the limitations in current tools being used for affect/emotion detection and suggested practical ways to overcome these limitations. Then to validate our framework and to control for task difficulty, we have used chess as a tool to study effective ways to mitigate the influence of incidental affect. Our results show that participants associate emotions with game situations ("emotion as information") to deal with complexity. This is in line with earlier findings in chess studies [12]. However, unlike the earlier studies where the experiment was conducted in controlled conditions, we have performed the experiment utilizing open chess Platform [27] where the task is drawn from large population and task is controlled for difficulty level based on the current strength of the participant and task performance. This approach helps us systematically eliminate task complexity from the list of confounding variables and makes the approach scalable. Another critical difference compared to earlier experiments: we did not include chess experts for this task rather chess amateurs.

The results also show that incidental affect can be mitigated by introducing time between tasks and being aware of one’s state [4,14]. Even though, Earlier work based on AIMD [Refer to Figure 5] highlighted the fact that incidental influences impact decision making, our way of eliciting the incidental influence is unique (integral to the previous task) and our focus is on the influence of transient nature of affect. Only 72% of the participants had performance gain, and 28% had poorly performed, these findings need more in-depth evaluation. We believe conducting these experiments over longer duration & multiple times may even out some un-factored variables like mood, stress, variability in task selection.

We have only considered the intensity of action units, and the number of Action Units detected. We can find correlation among individual action units and affect dynamics (emotion dynamics) to explain the variance in performance – but we have not yet done this exercise.

We expected differences in Fixation times, Fixation duration, saccade time and time spent on Area of Interest, but surprisingly, we did not find any statistically significant differences.

We speculate that the reason for this because the participants were all amateurs and earlier findings from empirical studies [10] show that there is a significant difference between novices and experts. Experiment on experts may reveal differences.

We have combined findings from various studies and created a holistic framework which we believe will uncover some surprising results.

Participants can perform this experiment at the convenience of their home and need not come to the lab.
Though there are bound to be anomalies, the sheer scale of data will dwarf the anomalies in results introduced due to setup or those who may game the system.

The emotion dynamics, self-reports, and human-feedback could be factored in, and its relation to the increase and decrease in performance could be evaluated. Individual gain and losses could be evaluated in detail for all the 29 AUs [34].

6. CONCLUSION

We validated the proposed framework using a pilot study.

The findings support our initial hypothesis that the adverse effect of incidental influences can be mitigated by introducing delay & being aware of the one’s state between tasks adapted to the current skill.

Both the intensity and number of Facial action units (AU) vary when solving a task. Complementing the emotion data, with Action units can be used to overcome the findings in the research that emotions vary across cultures and people.

We have outlined the architecture by addressing some of the limitations observed in the NITM framework [19].

Though AIMD framework (The Affect Integrated Model of Decision-making (AIMD) has proposed that incidental influences impact decision making & suggested ways to effectively respond to them, prior research so far has only focused on the long-term states like mood and not transient emotions, like those encountered during chess.

This study also validates the previous research findings that emotions are crucial and integral to cognitive tasks [12] and gives a new perspective for further research.

Further research should improve the model to account for all the observed effects like Gaze, Fixation time and duration [22]. If this model receives empirical support in further research, it will provide a precious relationship between affect and performance.

Most of the research on Incidental emotions has focused on mood, and not on transient affect/emotions that last for short duration, like those that come into play in games like chess, and relatively little is known about the ways to mitigate the influence of these emotions on performance in a quantifiable manner – both short term and long-term strategies.

Our goal is to integrate these findings into Learning companion that will detect and respond based on the student affect, this is in line with the vision outlined for affective cognitive agents [2].

This holistic architecture we believe will make the learning experience more effective and sustainable.

Finally, we should like to remark again that the characterization of affect regulation strategies on performance represented by this pilot study and framework is not unique to chess. As the results and architecture demonstrate, is generalizable to other domains like education, where strategies appropriate to a domain can account for the abilities of learners to acquire information based on affect with great rapidity & joy.

REFERENCES

12. Guntz, T., Crowley, J., Vaufreydaz, D., Balzarini, R., and
A Framework for Affective & Sustainable Learning


Amarnath DASAKA (Non-Member)
Amarnath Dasaka is a research scholar at the Department of Cognitive Science Lab, Kohli Centre for Intelligent Systems (KCIS), IIIT-Hyderabad. His ongoing research work focuses on exploring the interaction between Affect (emotions, feelings, mood) and learning. He is interested in understanding the human mind in a holistic way by synthesizing the research findings from various cognitive disciplines and insights from Indian philosophy. His focus for the future is to apply his research findings in developing commercially viable and socially sustainable solutions for affective learning, and virtual agents that can seamlessly interact with humans. He has previously designed solutions for Power, Performance and Stability for Mobile Chips and worked in multiple domains and technologies - AI, Wireless communication (3G/4G/5G), GPU, AR/VR, Autonomous Vehicles, Drones and IoT.

Bapi Raju SURAMPUDI (Non-Member)
Bapi S. Bapi received the BE (Electrical Engineering) degree from Osmania University, the MS (Biomedical Engineering) and PhD (Mathematical Sciences Computer Science) degrees from the University of Texas, Arlington, USA. He worked at BHEL, India; the University of Plymouth, UK; and ATR Research Labs, Kyoto, Japan before joining the University of Hyderabad. After having served University of Hyderabad till 2019, he joined as Professor at IIIT Hyderabad, India and is associated with the Cognitive Science Lab, Kohli Centre for Intelligent Systems (KCIS), IIIT-H. His research interests include the practical applications of various neural network and machine learning techniques, investigation of biological neural architectures, empirical experiments using behavioral and neuroimaging assays and cognitive modelling. He is a Senior Member of IEEE, member of the ACM, Society for Neuroscience and Cognitive Science Society.