A METHOD TO STUDY AFFECTIVE DYNAMICS AND PERFORMANCE IN ENGINEERING DESIGN TEAMS

Malte F. Jung, and Larry J. Leifer
Stanford University, USA

ABSTRACT
The practice of designing in teams is a socially mediated activity. Team members interact with one another to generate and develop concepts and physical artifacts over time. Researchers explored many different aspects of design interactions in order to generate new insights and theory about how interaction characteristics and performance relate. Despite many years of research not much has been done to explore the role of emotions in the context of the relationship of interaction dynamics and performance. The authors attribute this lack of research mainly to a lack of appropriate research methods to study emotions. In this paper we show that methods developed to study emotions in marital interactions can be used to study emotions in design team interactions. We describe the key components of such a method and share insights gained from its application three cases.

Keywords: Emotions, affective dynamics, thin slicing, systematic observation of behavior, design team interactions, performance

1 INTRODUCTION
The practice of designing in teams is a socially mediated activity [1, 2]. Team members interact with one another to generate and develop concepts and physical artifacts over time. The nature of this interaction has an impact on performance relevant outcomes such as the quality of final deliverables, the fulfillment of personal needs, or the willingness of a team to work together in the future [3]. Researchers interested in uncovering the relationship between social aspects of designing in teams and performance have looked at a large variety of phenomena such as question asking [4], gesturing activity [5], process changes [6], and many other aspects [7, 8]. Even specific laboratory environments were designed to study design interactions [9]. Despite these broad investigations of different interaction dynamics, not much has been done to look at the role of emotions in designing, and how the emotions designers express, or feel shape subsequent performance relevant outcomes. In design, researchers have primarily looked at how products elicit certain emotions [10, 11]. Emotions, however, are recognized to play an important role regarding the effectiveness of teams in general [12-16]. Especially the way a team handles disagreement and conflict has frequently been shown to relate to team outcomes [12, 17-19].

The reasons for a lack of research into the emotional dynamics of engineering design teams might be partially found in an absence of appropriate research methods. Currently, most research exploring the relationship between affect and performance relies on self-report measures. Tran, for example, explored how specific, self reported emotions predict performance in decision making teams [16]. Self-report measures of affect are easy to obtain and broad set of metrics have been developed and frequently validated [20-22]. However, studies that specifically explore performance relevant aspects using questionnaires are at danger of being biased by participants’ implicit theories about performance [23]. An exception to the self report studies, are Dong’s investigations of the role of affective appraisals in design interactions [24, 25]. These studies however only rely on the analysis of verbal protocols and neglect other behavioral channels such as facial muscle movements or changes in voice tone as indicators of affect. Overall the current methods have not led to the development of a coherent understanding about the mechanisms underlying the relationship between the affective characteristics of design team interactions and their performance relevant outcomes.

Researchers of married couples, on the other hand, have developed powerful methods to study affective dynamics in marital interactions. The methods developed by Gottman and his colleagues led
to new theory about the relationship between affective dynamics and subsequent outcomes [26], changed therapy practices [27], and direct insights that couples could use on their own to improve the quality of their relationships[28]. Especially illustrative is a study in which Gottman and Levinson [29] showed that it is possible to accurately (93%) predict the fate of a marriage based on the affective interaction quality determined from just a 15 minute video sample of a couple’s interaction. In particular, a classification of couples based on the ratio of positive to negative emotions into regulated (more positive than negative) and non-regulated (more negative than positive) couples was shown to be predictive of marital satisfaction and divorce. Using the same methods, it was even possible to predict divorce based on the affective interaction quality during the first 3 minutes of a conflict episode with 80% accuracy [30].

Given these powerful methods we see an opportunity to apply a set of methods that were developed to predict satisfaction and divorce in marriages to solve the problem of understanding the relationships between affective interaction characteristics and performance relevant outcomes in design teams. In our research we have applied these methods to study design team interactions. We are outlining the key components of that method, describe three cases of its application, and share the insights we gained from adapting and developing them.

2 KEY COMPONENTS OF A METHOD TO STUDY AFFECTIVE DYNAMICS AND PERFORMANCE IN ENGINEERING DESIGN TEAMS

We describe three core components of a method to study affective dynamics in engineering design teams that were derived from methods to study marital interactions and appropriated for the study of engineering design interactions.

Sampling the Affective Behavior of Real Teams in the Lab

Most researchers studying interactions in design teams make a trade-off between studying “real” teams in the field, or “artificial” teams in the lab. The study of real teams offers advantages regarding external validity of results, but due to varying contexts the teams are observed in, it is often difficult to separate the phenomena of interest from the context they were observed in and to generalize the results obtained. The study of laboratory teams, on the other hand makes it easier to isolate phenomena of interest and produce generalizable results, but it is questionable how externally valid insights are that were gained with teams that only existed for the duration of a study. The approach used by Gottman and his colleagues in the studies described above, combines advantages of both approaches. As it has been done with couples [31], we can bring “real” teams into a laboratory environment and observe their interaction behavior under controlled conditions. Even though teams are expected to behave differently in a lab context, we expect them to have different affective styles of interaction and that those styles will be observable in the lab. As with couples, we assume those styles to be indicative of a team’s behavior in the field. To generate a sample of behavior that is reflective of a team’s interaction style we modified a emotion elicitation task that was developed to elicit high arousal emotions in dyads [32] for the use with teams. The specific tasks sets up a conflict discussion that “amplifies” the emotional characteristics of a team interaction so that differences between teams can be observed more clearly.

Thin slicing of Behavior:

Thin-slicing refers to the process of making accurate classifications based on small samples, or “thin slices” of expressive behaviors [33]. The thin-slicing research showed powerfully that certain behavioral characteristics are stable over time and that only a small interaction sample is necessary to make meaningful judgments about behavior occurring over longer durations such as hours, or even months. For example, in a frequently cited study, Ambady and Rosenthal [34] showed that end-of-semester teacher evaluations could be accurately classified based on judgments of 30 second silent video clips of the respective teachers. In addition to its more prominent applications in predicting teacher evaluations [34] and marital outcomes [29, 30], the method of thin slicing has been applied successfully across a wide range of other contexts such as doctor-patient interactions, family interactions, interviews, or work related interactions [33]. In a meta-analysis across 38 different studies, Ambady and Rosenthal [33] were able to show that short behavioral samples ranging between 20 seconds and 5 minutes, are highly indicative of long-term characteristics, irrespective of the
specific context they were taken in. The method of thin slicing is particularly relevant when gathering
data about the interactions of real teams that exist not only for a few minutes or ours in the lab but
over weeks, months or even years.

**Systematic Observation of Behavior using SPAFF and RCISS:**
Systematic observation refers to a particular approach of quantifying behavior according to which
trained observers record occurrences of specific behaviors in correspondence with a precisely defined
coding scheme [35]. The goal of this approach is “for properly trained observers to produce identical
protocols, given that they observed the same stream of behavior” – [35]. Often audio or video
recordings are used as a basis for coding.

One of two central components of the method of Systematic Observation is the catalog of behavior
codes or “Coding Scheme”. Two approaches can be distinguished in developing a coding scheme: To
(1) derive it from existing theory “top down”, or (2) to develop it using a more grounded approach
“bottom up”, deriving it from the close observation of behavior itself [36]. The line between these two
approaches is necessarily blurred in practice and most approaches are somewhat hybrid, as their
development often iterates between the formation new codes from close observation and their
comparison to existing theory. A final coding scheme has to both define how a behavioral is unitized,
and then how the isolated units are classified. Some coding approaches start by unitizing the data and
then categorize the units afterwards. This can even be done be separate observers. Other approaches
use more simultaneous procedures for unitizing and categorizing behavioral streams. Finally coding
schemes can be placed on a continuum between physically based and socially based coding schemes.
According to Bakeman and Gottman [35], physically based schemes are schemes with clear and well-
understood roots in the organism’s physiology, and socially based schemes are schemes that deal with
behavior whose very classification depends far more on ideas in the mind of the investigator (and
others) than on mechanisms in the body.

The other central component of Systematic Observation is the demonstration of reliability of coders.
Coder reliability refers to the level of agreement between two independent observers of the same
stream of behavior. A common measure for the assessment of coder reliability is Cohen’s Kappa [37],
which determines the level of agreement corrected for chance agreements, in comparison to the often
used percent-agreement. Weingart and colleagues have distinguished between two types of reliability:
Unitizing reliability and interpretive reliability [36]. Unitizing reliability refers to the degree of
agreement regarding identifications of units to be categorized and interpretive reliability refers to the
degree of agreement in assigning the labels to the units.

Three prominent examples of coding systems for affect are the Facial Action Coding System (FACS)
[38], the Specific Affect Coding System (SPAFF) [39] and the Rapid Couples Interaction Scoring
System (RCISS) [40]. Of these three FACS could be characterized as a physically based scheme as it
classifies behavior on the level of movement. FACS assigns numbered Action Units (AUs) to specific
muscle movements in the face and a specific subset of AUs have been associated with emotional
expression. SPAFF and RCISS, on the other hand, classify behavior on the level of the affective
meaning that behavior has in a particular cultural context. They both make affect distinctions not only
based on facial muscle movements, but also on changes in vocal tone, content, and body movement.
Additionally, coding in RCISS and SPAFF relies on a cultural informant approach [39], by which
coders make categorizations on a gestalt level. These unique characteristics make RCISS and SPAFF
very powerful coding schemes, but also very difficult to train and use.

---

1 Both FACS and SPAFF have been referred to as “emotion coding schemes” although none of them actually codes emotions. They both
code behavior indicative of emotions but in two different ways. There are strong claims about a direct linkage between facial movement and
affective experience however facial movement is neither sufficient nor necessary for an emotion to occur.
3 AN APPLICATION OF THE METHOD IN THREE CASES

The three examples below describe how we applied the aforementioned method in three research cases. Each case is taken from an ongoing research project. In each case we took a small slice of an interaction, applied systematic observation of behavior, and measured performance at the end. The cases are only meant to illustrate the application of the method and to provide a context for a discussion of the insights we gained in its application.

Affective dynamics and performance in software design teams

Our first study focused on exploring the relationship between affective dynamics and performance in software design teams. We wanted to find out whether we can predict the performance of software designers by looking at the emotional expressions occurring in a small sample taken from the beginning of their interaction. Figure one gives an overview about the layout of the study. For the purpose of our exploration, we re-investigated a data-set with 17 teams, that was collected in a previous laboratory study [41, 42]. All teams consisted of programmers recruited from industry. Each of the teams had to design and implement a calendaring solution for scheduling meetings with multiple participants. Before starting with the main task each team had to complete a short warm-up task to familiarize themselves with the laboratory environment and the specific tools they were given. A detailed description of the study and its setup can be found in [43].

Thin-Slicing of Behavior: For our detailed analysis we selected a five-minute thin slice at 5 minutes into the main programming task. We chose a slice a few minutes in, to avoid the initial, non-task-related chatter that most teams exhibited when starting on the new task. We were pleased to find out that a five-minute slice was long enough to discover very distinct affective interaction styles amongst the different teams.

Systematic Observation of Behavior: For systematic observation of affective behavior we used an adapted version of the RCISS [40] coding scheme. Initially we tried to apply the coding scheme in its original form, but we had to modify the coding scheme through several interactions in order to adapt it to the much lower emotional intensity of the programming teams in comparison to couples. We discovered that many of the behaviors that distinguished the teams from one another were visible through a participant’s listening behavior. Table 1 shows the list of codes for speaker and listener that can be chosen from for each conversational turn. Speaker and listener always switch dependent on whose turn of speech it is.
<table>
<thead>
<tr>
<th>TABLE 1. Modified RCISS coding scheme</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>SPEAKER</strong></th>
<th>Interest</th>
<th>Validation</th>
<th>Excitement</th>
<th>Humor/Laugh</th>
<th>Empathy</th>
<th>Other Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>🎉</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SPEAKER</strong></th>
<th>Complain</th>
<th>Criticize</th>
<th>Constrained Anger</th>
<th>Defensive</th>
<th>Yes-but</th>
<th>Put down</th>
<th>Tension/Tense humor</th>
<th>Other negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>🙁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>LISTENER</strong></th>
<th>Backchannels present</th>
<th>Facial movement present</th>
<th>Looking at each other</th>
<th>Positive facial expression</th>
<th>Responsive facial movement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>🎉</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>LISTENER</strong></th>
<th>No backchannels</th>
<th>No facial movement</th>
<th>No looking at each other</th>
<th>Negative facial expression</th>
<th>Stonewalling</th>
</tr>
</thead>
</table>
Affective dynamics and performance in dyadic negotiations

With our second study we are exploring the relationship between affective dynamics and performance in dyadic negotiations. We chose the context of negotiations (1) because they have been referred to as a central component to the practice of designing [44] and (2) because negotiation setting provide a context for heightened emotional arousal and a context for which clear performance metrics can be established. As in the first case we were also interested in exploring whether the affective dynamics occurring during a small interaction slice would be indicative of overall performance relevant outcomes. 52 subjects divided into 26 pairs participated in the study. All participants were recruited from an ongoing negotiation class. Participants were invited into a laboratory environment and given instructions to prepare and negotiate one party in a dispute over the cost of a car repair. Participants were free to design the terms of their agreement and come up with creative solutions to solve the dispute. The entire negotiation was videotaped. Figure 2 provides an overview about the layout of the study over time.

Thin Slicing of Behavior: We selected a five-minute slice at 5 minutes into the discussion. As in the previous study, the five-minute slice proved long enough to distinguish the teams based on their affective interaction characteristics.

Systematic Observation of Behavior: For the systematic observation of behavior we used the same, modified RCISS coding scheme as in the previous study. Due to the different task, on initial examination, we were able to observe higher emotional arousal than in the first study, however most difference between teams still emerged out of the back channeling behavior.

Affective dynamics and performance in small design teams

Our third study combined all the three method components introduced above: Thin slicing of behavior, systematic observation, and sampling of affective behavior in the lab. With this study we are exploring the relationship between affective characteristics of interactions in small leader-less design teams and long-term performance relevant outcomes. Subjects in this study are 35 students divided into nine teams participating in a nine month long graduate course in mechanical engineering at Stanford University. A detailed description of the layout of the course can be found in [45]. For our investigation we invited each team into a laboratory environment where they were videotaped during a 15-minute conflict discussion task. The aim of our research is to explore the relationship between the emotions occurring during that 15-minute interaction and performance relevant outcomes as measured at the end of the class. Figure 3 provides a visualization of the overall layout of the study.
Thin Slicing of Behavior and Behavior Sampling in the Lab: To generate an appropriate slice of a team’s behavior we modified the sequence of tasks that were used in the studies with married couples. Figure 4 provides an outline of the task we employed. Each team started with a discussion of project requirements (1) in order to accommodate them to the lab environment and situate them in the context of their class project. Tasks (2) and (3) were designed to set up a conflict discussion topic. During task (4) each team discussed the selected topic for 15 minutes. This interaction delivered the 15-minute slice that was the bases for a detailed analysis of the affective characteristics. Tasks (5) and (6) consisted of a series of self-report measures of affect. We were surprised to see the level of emotional arousal exhibited by the teams when engaged in the conflict discussion task. The procedure employed was able to elicit a range and intensity of emotions that we were not able to see in any of the previous studies. The material obtained through this procedure was suitable for an analysis using the methods developed to study marital interactions.

Systematic Observation of Behavior: For a systematic observation of behavior we used a modified version of the SPAFF [39]. We found that the RCISS coding scheme used in the previous studies could not be successfully adapted for teams larger than two. SPAFF, on the other hand has been used to study in emotions in the interactions of cancer support groups with larger numbers of participants [46]. Due to the highly engaging conflict discussion task, we were also able to observe almost and emotion code listed. Several teams even exhibited short moments of sadness when engaged in interactions.
4 DISCUSSION AND CONCLUSION

The methods and cases introduced above, give an overview about a set of methods to study the relationship between interaction and performance in engineering design teams from an affective behavior perspective. Concluding we want to discuss one of the major of the limitations in using the approach outlined above its high costs in generating data. Coding a 15-minute segment of video, as recorded from the interactions of design teams, took about six to eight hours per team member. Additionally, extensive training (two to three weeks) is needed to prepare coders to reliably code behavior according to the coding schemes employed.

Other researchers have shown that it is possible to automatically generate data about interactions in teams using sociometric badges [47]. In one study of negotiations, Curhan and Pentlant were able to show that the performance of a negotiation can be predicted from a thin since of an interaction early on using automatically generated interaction metrics [48]. However the results of this approach are difficult to translate into recommendations for practitioners. The data collected through the badges are mere proxies for social interaction dynamics and are not meaningful on a human level. This aspect is what makes SPAFF particularly powerful, even though difficult to apply, its distinctions are meaningful from an intervention perspective. Each distinction gives a direct handle though with the quality of an interaction can be shaped.

With our paper we demonstrate, how methods to study affective dynamics of marital interactions can be applied to investigate the interaction dynamics of design teams. It is our aim to give other researchers a set of starting points for their own investigations into the affective dynamics of design teams. We think that by developing and applying these methods further we can gain new and important insights into the relationship between interaction dynamics and performance of engineering design teams.

REFERENCES


[38] Ekman, P. and Friesen, W.V., Facial Action Coding System: A Technique for the Measurement


Contact: Malte F. Jung
Stanford University
Center for Design Research
Stanford, CA, 94305
USA
Email: mjung@stanford.edu
URL: http://me.stanford.edu/research/centers/cdr/index.html

Malte is a Ph.D Candidate in Mechanical Engineering at Stanford University. His research is at the intersection of mechanical engineering, psychology, and organizational behavior. He is interested in the role of emotions in engineering design team interactions.