Understanding the Affect of Developers: Theoretical Background and Guidelines for Psychoempirical Software Engineering

Daniel Graziotin
Faculty of Computer Science,
Free University of
Bozen-Bolzano
Piazza Domenicani 3
Bolzano/Bozen, Italy
daniel.graziotin@unibz.it

Xiaofeng Wang
Faculty of Computer Science,
Free University of
Bozen-Bolzano
Piazza Domenicani 3
Bolzano/Bozen, Italy
xiaofeng.wang@unibz.it

Pekka Abrahamsson
Department of computer and
information science
Norwegian University of
Science and Technology
NO-7491 Trondheim, Norway
pekkaa@ntnu.no

ABSTRACT

Affects—emotions and moods—have an impact on cognitive processing activities and the working performance of individuals. It has been established that software development tasks are undertaken through cognitive processing activities. Therefore, we have proposed to employ psychology theory and measurements in software engineering (SE) research. We have called it "psychoempirical software engineering". However, we found out that existing SE research has often fallen into misconceptions about the affect of developers, lacking in background theory and how to successfully employ psychological measurements in studies. The contribution of this paper is threefold. (1) It highlights the challenges to conduct proper affect-related studies with psychology; (2) it provides a comprehensive literature review in affect theory; and (3) it proposes guidelines for conducting psychoempirical software engineering.

Categories and Subject Descriptors

D.2.9 [**Software Engineering**]: Management—*Productivity, Programming Teams*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors, Software psychology*; J.4 [**Social and behavioral Science**]: [Psychology]

Keywords

Affects, emotions, moods, human aspects in software development, psychology of programming, psychoempirical software engineering

1. INTRODUCTION

The Nobel prize winner Daniel Kahneman has pointed out that it is unrealistic to limit our understanding of human behaviors solely through rational models [41]. Yet, software engineering (SE) research has been known to be too much confined in the fallacy of rationality-above-everything paradigm [63], to miss out the possibility to be a social discipline [82], and to focus too much on domains of technical nature while neglecting the so-called *soft aspects* or *human-related* topics [52]. But software development *is* a very human activity. Software development happens in our minds first, then on artifacts [26]. It has been established that development is intellectual, and it is carried out through cognitive processing activities [25, 26, 42]. Indeed, we are human beings, and, as such, we behave based on affect as we encounter the world through our emotions and moods [10]. The affects pervade organizations by coloring the workers' thoughts, and they influence their behavior [8]. Affects have a role in the relationships between workers, deadlines, work motivation, sense-making, and human-resource processes [3]. Although affects have been historically neglected in the studies of industrial and organizational psychology [61], an interest in the role of affects on job outcomes has accelerated over the past fifteen years in psychology research [27]. While research is still needed on the impact of affects on cognitive activities and work-related achievements in general, this link undeniably exists according to psychology research.

We have shown elsewhere [33] that practitioners are deeply interested in their affects while developing software, which causes them to engage in long and interesting discussions when reading related articles. Thus, it is important to understand the role of affects in software development processes. Even more, we share the view of Lenberg et al. [52] that SE should also be studied from a behavioral perspective. We have, in fact, focused on these issues for some time now, by producing several articles on this avenue, i.e., [30, 16, 31, 32, 33, 34]. We have also proposed the term *psychoempirical software engineering* [35] to denote research in SE with proper theory and measurement from psychology. Our message was well-received by the community with some degree of agreements regarding terminology, e.g., [51, 52]. However, we show below that long is the road to properly address the human aspects of SE with psychology.

Problem: SE Lacks in Theoretical Background of Affects and Guidelines for Using Psychology.

SSE'15, September 1, 2015, Bergamo, Italy
ACM. 978-1-4503-3818-9/15/09...$15.00
http://dx.doi.org/10.1145/2804381.2804386

Given the rising number of recent SE articles that deal with the affects of developers, e.g., [62, 28, 37, 20], we believe that it should be important for researchers to adopt a critical view of the phenomenon under study, and that they do not fall into the several misconceptions when dealing with the affect of developers [34].

Yet, we understand that we have placed ourselves in a "very confused and confusing field of study" ([64], p. 2). We experienced this confusion especially during our talks at ISERN 2014, where we chaired a workshop called psychoempirical SE [35], and during the CHASE 2015 workshop [6], where we presented some common misconceptions and measurements of the affect of software developers [34]. Such misconceptions include confusing affect and the related constructs of emotions and moods with motivation or job satisfaction, which has happened even in articles already dealing with misconceptions of motivation with respect to job satisfaction, e.g., [29], although affects were not the focus of the study in this case.

Other issues lie in missing out the opportunity of using validated measurement instruments for affect. An example is the use of the niko-niko calendar for assessing the mood of a software development team, e.g. [77], or the so-called happiness index, e.g., [56]. Another example of the missed opportunity is when a single truth is assumed in the writing of articles, like in a CACM positional article claiming that "psychologist recognize eight basic emotions, with each positive balanced by a negative", e.g. "love-hate" ([19], p. 34), or in a proper empirical study where it has been claimed that "there are six basic emotions or universal emotions: anger, happiness, fear, [..]" ([13], p. 1079). We will show below that it is not true that a unique dominant, accepted theory exists for affect, emotions, and moods. Researchers should recognize this issue when employing such delicate concepts for conducting research.

Proposal: Theoretical Background of Affects and Guidelines for Psychoempirical SE.

While it would be preposterously arrogant on our side to claim the all-encompassing knowledge of the topic, we would like to share what we have learned so far in our journey to understanding software developers through their affect. This article builds upon our experience, the feedback collected at our talks and peer review processes, and the previously conducted research, to build some theoretical background for understanding the affect of software developers. We draw from research in psychology in the last decades, and offer a comprehensive review of the theory of affect (section 2) and, as a follow-up of our ISERN 2014 workshop [35], we propose our guidelines for psychoempirical SE (Section 3) for conducting studies in SE with psychological theory and measurement.

2. AFFECT, EMOTIONS, MOODS: THEORETICAL BACKGROUND

The fields of psychology have failed to agree on the definitions of affects and the related terms such as emotions, moods, and feelings [64, 72]. Yet, it is desirable that we provide a starting set of definitions, which we will however criticize.

Let us start by stating that the term *affect* (or affective state) has been defined as "any type of emotional state [...] often used in situations where emotions dominate the person's awareness" [85]. This definition is problematic as it contains the term *emotion*, which has not yet been defined, and it does not help in defining the (now apparently) superconstruct *affects*. Indeed, the term *affects* is often associated in the literature with *emotions* and *moods*. We now are left with three terms, which look remarkably similar to each other.

Plutchik [66] has defined *emotions* as the states of mind that are raised by external stimuli and are directed toward the stimulus in the environment by which they are raised. However, Kleinginna et al. [45] reported one year later that more than 90 definitions have been produced for this term, and no consensus in the literature has been reached. The term has been taken for granted and often defined with references to a list, e.g. anger, fear, joy, surprise [9]. To worsen this, *emotion* as a term is not universally employed, as it is a word that does not exist in all languages and cultures [71].

Moods have been defined as emotional states in which the individual feels good or bad, and either likes or dislikes what is happening around him/her [65]. Yet again, a definition of one construct contains another construct of our interest.

How Do Emotions and Moods Differentiate, Then?

While for some researchers certain moods are emotions and vice versa [18], it has been suggested that a distinction is not necessary for studying cognitive responses that are not strictly connected to the origin of the mood or emotion [89]. Distinctions between emotions and moods are clouded, because both may feel very much the same from the perspective of an individual experiencing either [5] and are now a part of common sense [72]. They are embedded in psychologists' questions and, as a consequence, answers. Reisenzein [68] argued that "*the consensual definition of emotion is not a precondition but the result of scientific research; and even then, it remains a revisable empirical hypothesis*" (p. 2). So, affects, emotions, and moods are an emergent construction rather than a latent entity [11, 58].

We have adopted the same stance of several researchers in the various fields [79, 78, 88, 17] and employed the noun *affects* (affective states) as an umbrella term for emotions and moods. We will show that, according to a recent unifying theory, this strategy does make sense.

2.1 The Major Frameworks for Affect Theories

According to Huang [39], four major theories exist for emotions (moods, affects) in psychology. However, we see that these four theories and all the other we could review fall into two competing frameworks.

The Discrete Framework.

One framework, namely the discrete approach, collects a set of basic affective states that can be distinguished uniquely [66], and that possess high cross-cultural agreement when evaluated by people in literate and preliterate cultures [23].

The Differential Emotions Theory [40] states that the human motivation system is based on ten fundamental emotions (interest, joy, surprise, distress, anger, disgust, contempt, fear, shame, and guilt). These fundamental emotions function for the survival of human beings, possess an own neural network

in the brain, and an own behavioral response. Finally, these emotions interact with each other simultaneously.

Ekman [23] proposed a set of basic affects, which include anger, happiness, surprise, disgust, sadness, and fear. However, the list has received critique, leading to an extended version of eleven elements [24]. They include amusement, embarrassment, relief, and shame.

In the Circular Model of Emotion [66], a structure describing the interrelations among emotions has been proposed. Eight primary, bipolar affective states were presented as coupled pairs: joy versus sadness, anger versus fear, trust versus disgust, and surprise versus anticipation. These eight basic emotions vary in intensity and can be combined with each other, to form secondary emotions. For example, joy has been set as the midpoint between serenity and ecstasy, whereas sadness has been shown to be the midpoint between pensiveness and grief. Emotions can vary in intensity and persistence (to form moods, for example). Emotions, under this theory, serve an adaptive role in dealing with survival issues.

Developing a minimal list of basic affective states appears to be difficult with the discrete approach. Subsequent studies have come to the point where more than 100 basic emotions have been proposed [80].

The Dimensional Framework.

The dimensional framework groups affects in major dimensions that allow a clear distinction among them [70, 47]. In the PAD models, three dimensions of Pleasure-displeasure, Arousal-nonarousal, and Dominance-submissiveness [75, 70, 57] characterize the emotional states of humans. *Valence* (or pleasure) is the attractiveness (or adverseness) of an event, object, or situation [53] [49]. The term refers to the "direction of a behavioral activation associated toward (appetitive motivation) or away (aversive motivation) from a stimulus" [47]. *Arousal* represents the intensity of emotional activation [47]. It is the sensation of being mentally awake and reactive to stimuli, i.e. vigor and energy or fatigue and tiredness [90]. *Dominance* (or control, over-learning) represents a change in the sensation of the control of a situation [7]. It is the sensation by which an individual's skills are perceived to be higher than the challenge level for a task [15]. Figure 1 provides a representation of a PAD model of valence and arousal, and examples of related discrete affects with an indication to where they might correspond on the axes.

Emotional states under the PAD models include moods, feelings, and any other feeling-related concepts. The dimensions are usually bipolar, indicating that the presence of pleasure excludes the possibility of displeasure. Some variations of these models have been proposed using different notations but without changing the core meaning [72], some of which omit the dominance dimension [47].

In the Positive and Negative Affect Schedule (PANAS) [86, 87], the positive and negative affects are considered as the two primary emotional dimensions. However, these two dimensions are the result of the self-evaluation of a number of words and phrases that describe different feelings and emotions. That is: discrete emotions are rated but two dimensions are evaluated. This theory is designed to present a mood scale. Finally, positive and negative affects are mutually independent. In Figure 1, the positive (negative) dimension would comprise of positive (negative) valence, or positive arousal, or both according to the different theories.

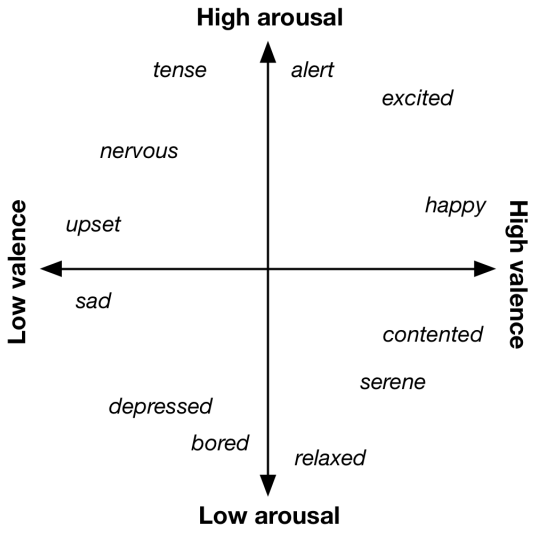

Figure 1: A PAD model of valence and arousal, and examples of related discrete affects.

We note here that several other theories exist, although they are less prominent. One example is the cube of emotion [55], which is a dimensional theory of affect that expresses affect in terms of combinations of dopamine, adrenaline, and serotonin, which intersect in eight basic (but extreme) affects, e.g. distress, interest, joy.

The Unifying Theory.

A prominent unifying theory exists as well. Russell and Barrett [74, 72, 73] have proposed the concept of *core affect* to unify the theories of emotions and moods in psychology. Core affect is "a pre-conceptual primitive process, a neurophysiological state, accessible to consciousness as a simple non-reflective feeling that is an integral blend of hedonic (valence) and arousal values: feeling good or bad, feeling lethargic or energized" [72] (p. 147). The state is accessible at a consciousness level as the simplest raw feelings, which is distinct in moods and emotions. A feeling is an assessment of one's current condition. Therefore, an affect is a very raw concept, upon which the more complex of mood and emotion is built upon. *Pride* can be thought of as feeling good about oneself. The "feeling good" is core affect and the "about oneself" is an additional (cognitive) component.

Changes in affects result from a combination of happenings, such as stressful events on the job. Sometimes the cause of the change is obvious. However, sometimes one can undergo a change in core affect without understanding the reasons. The individuals possess a limited ability to track this complex causality connection. Instead, a person makes attributions and interpretations of core affect.

Affect can be felt in relation to no obvious stimulus—in a free-floating form—as *moods* are perceived. Indeed, mood is defined as a prolonged core affect without an object, i.e. an unattributed affect.

In the core affect theory, *emotions* are episodes instead of simple objects. An emotion is a complex set of interrelated sub-events about a specific object.

The core affect theory is interesting because it unifies the previous theories, and it maintains compatibility with the

majority of the existing measurement instruments, regardless of them being about moods or emotions. Although we do not neglect moods and emotions *per se*, when adopting the core affect theory we chose to understand the states of minds of software developers at the *affective* level only, which is the foundation of moods and emotions.

Core Affect Is our Current Suggestion to Frame SE Research on Affect. However, a researcher should select the affective framework and theory that better suits the research objective and the level of details that are desirable. We provide more details in the next section. What is important is that researchers are aware of an absence of an absolute truth and of the many existing alternatives, and that they justify their choice.

3. GUIDELINES FOR PSYCHOEMPIRICAL SOFTWARE ENGINEERING

A much requested feature in our previous discussions at recent academic venues such as ISERN 2014, CHASE 2015 [6], and ICSE 2015 had been *How should one conduct research with psychological measurements?* By making sense of the hundreds of articles we reviewed on psychology and organizational behavior, we came up with a simple series of steps, listed below.

Defining a Research Objective.

As with any research activity, it is important to understand what we want to do in a study. Suppose two different, yet common scenarios with the affects of developers. They have been adapted from two of our previous studies [32, 31].

Scenario A Assessing how happy developers are generally.

Scenario B Assessing over a time frame the emotional reaction of a stimulus (e.g., employing a software tool) on developers.

Both of them require a deep understanding of the topic under study.

Theoretically Framing the Research.

Scenario A—From a comprehensive literature review, we would understand that we can call happy those developers who are in a strongly positive mood, or those who frequently have positive and meaningful experiences (see [34] for more), thus having a positive affect balance. We decide to focus on dimensions of affects, e.g. with the Positive and Negative Affect Schedule (PANAS) [86, 87], which still lets us evaluate discrete affects before the aggregated scores.

Scenario B—Suppose that, instead of asking a developer what emotions she is feeling when using a tool, we are interested in knowing how she feels in terms of more aggregated dimensions like pleasure, energy, and dominance. We focus then on the dimensional theory of affects like the one in the PAD models [75, 70, 57].

Selecting a Validated Measurement Instrument.

Scenario A—The PANAS dimensional model recommend employing the PANAS [86, 87] measurement instrument which is one of the most notable measurement instruments for affective states. However, a deeper look at the literature shows that there are several shortcomings that have been criticized for this instrument. The PANAS reportedly omits core emotions such as *bad* and *joy* while including items that are not considered emotions, like *strong*, *alert*, and *determined* [21, 54]. Another limitation has been reported in its non-consideration of the differences in desirability of emotions and feelings in various cultures [84, 54]. Furthermore, a considerable redundancy has been found in PANAS items [14, 83, 54]. PANAS has also been reported to capture only high-arousal feelings in general [21].

Recent, modern scales have been proposed to reduce the number of the PANAS scale items and to overcome some of its shortcomings. Diener [21] developed the Scale of Positive and Negative Experience (SPANE). SPANE assesses a broad range of pleasant and unpleasant emotions by asking the participants to report them in terms of their frequency during the last four weeks. It is a 12-items scale, divided into two sub-scales. Six items assess positive affective states and form the SPANE-P scale. The other six assess negative affective states and form the SPANE-N scale. The answers to the items are given on a five-point scale ranging from 1 (*very rarely or never*) to 5 (*very often or always*). For example, a score of five for the *joyful* item means that the respondent experienced this affective state *very often* or *always* during the last four weeks. The SPANE-P and SPANE-N scores are the sum of the scores given to their respective six items. Therefore, they range from 6 to 30. The two scores can be further combined by subtracting SPANE-N from SPANE-P, resulting in the Affect Balance Score (SPANE-B). SPANE-B is an indicator of the pleasant and unpleasant affective states caused by how often positive and negative affective states have been felt by the participant. SPANE-B ranges from -24 (*completely negative*) to +24 (*completely positive*). The SPANE measurement instrument has been reported to be capable of measuring positive and negative affective states regardless of their sources, arousal level or cultural context, and it captures feelings from the emotion circumplex [21, 54]. The timespan of four weeks was chosen in SPANE in order to provide a balance between the sampling adequacy of feelings and the accuracy of memory [54], and to decrease the ambiguity of people's understanding of the scale itself [21].

Scenario B—The PAD dimensional models have been implemented in several measurement instruments. One of the most notable instruments is the Affect Grid [76], which is a grid generated by intersecting the axes of valence and arousal accompanied by four discrete affects, i.e. depression-relaxation and stress-excitement, to guide the participant in pointing where the emotional reaction is located. The affect grid has been employed in SE research, e.g. in [12]. Yet, the grid was shown to have only moderate validity [43], thus other measurement instruments would be more desirable. Thus comes the Self-Assessment Manikin (SAM, [7, 48]). SAM is a pictorial, i.e. non-verbal, assessment method. SAM measures valence, arousal, and dominance associated with a person's affective reaction to an object (or a stimulus) [7]. As a picture is worth a thousand words, we reproduce SAM in figure 2. The figures of the first row range from a frown to a smile, representing the valence dimension. The second row depicts a figure showing a serene, peaceful, or passionless face to an explosive, anxious, or excited face. It represents the arousal dimension. The third row ranges from a very little, insignificant figure to a ubiquitous, pervasive figure. It represents the dominance affective dimension. As reported

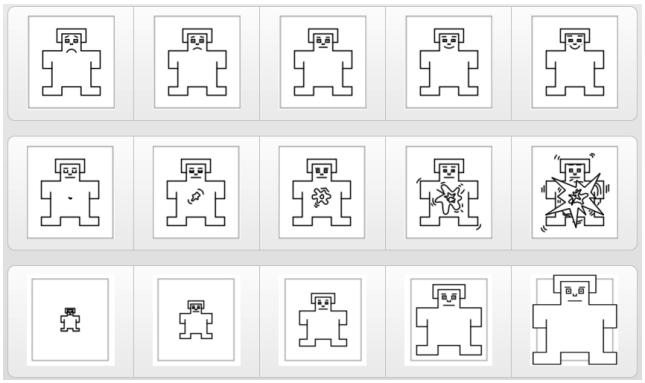

Figure 2: The Self-Assessment Manikin.

in [44], SAM has the advantage of eliminating the cognitive processes associated with verbal measures but it is still very quick and simple to use.

Considering Psychometric Properties.

As we noted in a previous paper [34], a selected measurement instrument has to possess acceptable validity and reliability properties, which are provided in psychometric studies of the measurement instrument. Psychometrics is a term, which has been misused in SE including ourselves. It is a subfield of psychology that focuses on the theory and techniques of psychological measurements. Psychometric studies deal with the design, development and especially the validation of psychological measures.

A modification to an existing measurement instrument (e.g., adding, deleting, or rewording items) often requires a new psychometric study because the reliability of a measurement instrument can be compromised. Therefore, it is not advisable to modify validated psychological measurements or models as it happened in [13].

Scenario A—The SPANE has been validated to converge with other affective states measurement instruments, including PANAS [21]. The scale provided good psychometric properties in the introductory research [21] and in numerous follow-ups, with up to twenty-one thousand participants in a single study [81, 22, 54]. Additionally, the scale proved consistency across full-time workers and students [81].

Scenario B—The SAM has been under scholarly scrutiny, as well. The original article describing SAM already reports good psychometric properties [7]. A very high correlation was found between the SAM items and those of other verbal-based measurement instruments [59, 60], including high reliability across age [2]. Therefore, SAM is one of the most reliable measurement instruments for affective reactions [44].

Administering the Measurement Instrument Correctly.

The psychometric properties of a measurement instrument in psychology are also calculated by administering the instrument in the same way in each study. This is because the instructions might influence the participants' responses. For this reason, any good measurement instrument is always accompanied with the instructions for the participants. We encourage administering a measurement instrument as it is reported in the accompanying instructions, and to further share the instructions with participants. Furthermore, the gained transparency ensures a higher reproducibility of the studies.

We strongly encourage the authors of SE studies to report the participants' instructions when publishing an article, preferably in an archived format. [1]

Scenario A—The SPANE instructions for participants are clearly stated in the original paper [21] and in the instrument itself, which is freely available. [2]

Scenario B—The SAM instructions for participants are exhaustively reported in the accompanying technical report [48].

Performing Strong Analyses.

We encourage the authors in SE to spend some time to understand whether such complex and delicate constructs require accurate analyses.

Scenario A—The SPANE scores can be considered as ordinal values or as discrete pinpoints of a continuous scale. Regression analyses on the aggregated SPANE-P, SPANE-N, and SPANE-B scores are possible given that the assumptions for linear regression are met. Otherwise, especially when groups have to be compared, the usual assumptions for employing the *t-test* or non-parametric tests should be taken into account. It is also important to report an effect size measure such as the Cohen's *d*.

Scenario B—Repeated measures within-subject that need a between subject comparison pose several issues. First, there is not a stable and shared metric for assessing the affects across persons. For example, a score of one in valence for a person may be equal to a score of three for another person. However, a participant scoring two for valence at time t and five at time $t+x$ unquestionably indicates that the participant's valence increased. As stated by Hektner [38], "*it is sensible to assume that there is a reasonable stable metric within persons*" (p. 10). In order to have comparable measurements, the raw scores of each participant are typically transformed into *z-scores* (also known as standard scores). A z-score transformation is such that a participant's mean score for a variable is zero, and scores for the same variable that lie one standard deviation above or below the mean have the value equivalent to their deviation. One observation is translated to how many standard deviations the observation itself is above or below the mean of the individual's observations. Therefore, the participants' measurements become dimensionless and comparable with each other, because the z-scores indicate how much the values are spread [50, 38].

Second, the repeated measurements often present dependencies of data at the participants' level and the time level grouped by the participant. The analysis of variance (ANOVA) family provides rANOVA as a variant for repeated measurements. However, rANOVA and general ANOVA procedures are discouraged [36] in favor of mixed-effects models, which are robust and specifically designed for repeated, within-participant longitudinal data [46, 36, 1]. A linear mixed-effects model is a linear model that contains both fixed effects and random effects [69]. The estimation of the significance

[1] For the participants' instructions in [32], see `https://dx.doi.org/10.7717/peerj.289/supp-1`. For the participants' instructions in [30, 31], see `http://dx.doi.org/10.6084/m9.figshare.796393`

[2] `http://internal.psychology.illinois.edu/~ediener/SPANE.html`

of the effects for mixed models is an open debate [4, 67]. We encourage the reader to follow our reasoning in [31] for a deeper discussion.

4. CONCLUSION

Affects—emotions and moods—are beginning to be comprehensively studied in SE, and other psychological constructs are being incorporated in related research. However, there is a risk of underusing and misusing the theory and the measurement instruments from psychology, and falling into the many misconceptions tied to such intriguing and complex research topics.

For this reason, we have proposed the term *psychoempirical software engineering* to refer to the research in SE with psychology theory and measurement. This paper described the challenge to conduct proper affect-related studies with psychology, provided a comprehensive literature review in affect theory, and proposed guidelines for conducting psychoempirical software engineering.

With this article, we hope to raise much needed awareness for better use of psychology in SE studies and to begin a sane discussion with our peers towards a more standard and sound way of conducting studies on the human and social aspects of SE.

5. ACKNOWLEDGMENTS

We are grateful for the comments received by all at the ISERN 2014 and CHASE 2015 workshops. We are also grateful to every member of the scientific community who spent time in reviewing our articles on the affect of developers, including the present one, which brought us to research in deep about this fascinating topic that deserves to be properly researched in SE.

6. REFERENCES

[1] R. Baayen, D. Davidson, and D. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412, 2008.

[2] R. W. Backs, S. P. da Silva, and K. Han. A comparison of younger and older adults' self-assessment manikin ratings of affective pictures. Technical Report 4, 2007.

[3] S. G. Barsade and D. E. Gibson. Why Does Affect Matter in Organizations? *Academy of Management Perspectives*, 21(1):36–59, 2007.

[4] D. Bates. lmer, p-values and all that, 2006.

[5] C. Beedie, P. Terry, and A. Lane. Distinctions between emotion and mood. *Cognition & Emotion*, 19(6):847–878, 2005.

[6] A. Begel, R. Prikladnicki, Y. Dittrich, C. de Souza, A. Sarma, and S. Athavale. Proceedings of the 8th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE 2015). In *CHASE 2015*, 2015.

[7] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[8] A. P. Brief and H. M. Weiss. Organizational behavior: affect in the workplace. *Annual review of psychology*, 53:279–307, 2002.

[9] M. Cabanac. What is emotion? *Behavioural Processes*, 60(2):69–83, 2002.

[10] C. Ciborra. *The Labyrinths of Information: Challenging the Wisdom of Systems*. Oxford University Press, USA, New York, USA, 1 edition, 2002.

[11] G. L. Clore and A. Ortony. Appraisal Theories: How Cognition Shapes Affect into Emotion. In M. Lewis, J. M. Haviland-Jones, and L. Feldman-Barrett, editors, *Handbook of Emotions*, chapter 39, pages 628–642. The Guilford Press, New York, USA, 3 edition, 2008.

[12] R. Colomo-Palacios, C. Casado-Lumbreras, P. Soto-Acosta, and A. García-Crespo. Using the affect grid to measure emotions in software requirements engineering. *Journal of Universal Computer Science*, 17(9):1281–1298, 2011.

[13] R. Colomo-Palacios, C. Casado-Lumbreras, P. Soto-Acosta, and A. García-Crespo. Decisions in software development projects management. An exploratory study. *Behaviour and Information Technology*, 32(11):1077–1085, 2013.

[14] J. R. Crawford and J. D. Henry. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *The British journal of clinical psychology / the British Psychological Society*, 43(Pt 3):245–65, 2004.

[15] M. Csikszentmihalyi. *Finding flow: The psychology of engagement with everyday life*, volume 30. Basic Books, New York, USA, 1 edition, 1997.

[16] Daniel Graziotin. The Dynamics of Creativity in Software Development. In *Doctoral Symposium of the 14th International Conference on Product-Focused Software Process Improvement (PROFES 2013)*, pages 1–6, Paphos, Cyprus, 2013. `http://dx.doi.org/10.6084/m9.figshare.703568`.

[17] C. K. W. De Dreu, B. a. Nijstad, M. N. Bechtoldt, and M. Baas. Group creativity and innovation: A motivated information processing perspective. *Psychology of Aesthetics, Creativity, and the Arts*, 5(1):81–89, 2011.

[18] C. DeLancey. Basic Moods. *Philosophical Psychology*, 19(4):527–538, 2006.

[19] P. J. Denning. Moods. *Communications of the ACM*, 55(12):33, 2012.

[20] P. Dewan. Towards Emotion-Based Collaborative Software Engineering. In *ICSE*, 2015.

[21] E. Diener, D. Wirtz, W. Tov, C. Kim-Prieto, D. Choi, S. Oishi, and R. Biswas-Diener. New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Social Indicators Research*, 97(2):143–156, 2009.

[22] T. Dogan, T. Totan, and F. Sapmaz. The Role Of Self-esteem, Psychological Well-being, Emotional Self-efficacy, And Affect Balance On Happiness: A Path Model. *European Scientific Journal*, 9(20):31–42, 2013.

[23] P. Ekman. Universals and cultural differences in facial expressions of emotion. In *Nebraska Symposium on Motivation*, pages 207–283, 1971.

[24] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3):169–200, 1992.

[25] R. Feldt, L. Angelis, R. Torkar, and M. Samuelsson. Links between the personalities, views and attitudes of

software engineers. *Information and Software Technology*, 52(6):611–624, 2010.

[26] G. Fischer. Cognitive View of Reuse and Redesign. *IEEE Software*, 4(4):60–72, 1987.

[27] C. D. Fisher and N. M. Ashkanasy. The emerging role of emotions in work life: an introduction. *Journal of Organizational Behavior*, 21(2):123–129, 2000.

[28] D. Ford and C. Parnin. Exploring Causes of Frustration for Software Developers. In *2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 115–116, 2015.

[29] C. França, H. Sharp, and F. Q. B. da Silva. Motivated software engineers are engaged and focused, while satisfied ones are happy. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM '14*, pages 1–8, New York, USA, 2014. ACM Press.

[30] D. Graziotin, X. Wang, and P. Abrahamsson. Are happy developers more productive? The correlation of affective states of software developers and their self-assessed productivity. In *14th International Conference of Product-Focused Software Process Improvement. LNCS*, volume 7983, pages 50–64, 2013.

[31] D. Graziotin, X. Wang, and P. Abrahamsson. Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering. *Journal of Software: Evolution and Process*, Early View:1–21, 2014. `http://doi.wiley.com/10.1002/smr.1673`.

[32] D. Graziotin, X. Wang, and P. Abrahamsson. Happy software developers solve problems better: Psychological measurements in empirical software engineering. *PeerJ*, 2:e289, 2014.

[33] D. Graziotin, X. Wang, and P. Abrahamsson. Software developers, moods, emotions, and performance. *IEEE Software*, 31(4):24–27, 2014.

[34] D. Graziotin, X. Wang, and P. Abrahamsson. The Affect of Software Developers: Common Misconceptions and Measurements. In *2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 123–124, Firenze, Italy, 2015. IEEE Computer Society.

[35] D. Graziotin, X. Wang, P. Abrahamsson, and A. Jedlitschka. Psychoempirical Software Engineering. In *2014 Annual Meeting of the International Software Engineering Research Network (ISERN 2014)*, Turin, Italy, 2014. `http://dx.doi.org/10.6084/m9.figshare.1169833`.

[36] R. Gueorguieva and J. H. Krystal. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Archives of General Psychiatry*, 61(3):310–7, 2004.

[37] L. Haaranen, P. Ihantola, J. Sorva, and A. Vihavainen. In Search of the Emotional Design Effect in Programming. In *37th International Conference on Software Engineering (ICSE 2015)*, 2015.

[38] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Sage Publications Inc, 2007.

[39] M. Huang. The theory of emotions in marketing. *Journal of Business and Psychology*, 16(2):239–247, 2001.

[40] C. E. Izard. *Human Emotions*. Plenum Press, New York, USA, 1 edition, 1977.

[41] D. Kahneman. Maps of bounded rationality: Psychology for behavioral economics, 2003.

[42] I. A. Khan, W. Brinkman, and R. M. Hierons. Do moods affect programmers' debug performance? *Cognition, Technology & Work*, 13(4):245–258, 2010.

[43] W. D. Killgore. The Affect Grid: a moderately valid, nonspecific measure of pleasure and arousal. *Psychological reports*, 83(2):639–642, 1998.

[44] J. Kim, J. A. Geason, C. Woo, and J. D. Morris. The Power of Affect: Predicting Intention. *Journal of Advertising Research*, 42(3):7–17, 2002.

[45] P. R. Kleinginna and A. M. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5:345–379, 1981.

[46] N. M. Laird and J. H. Ware. Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4):963–974, 1982.

[47] R. D. Lane, P. M.-L. Chua, and R. J. Dolan. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*, 37(9):989–997, 1999.

[48] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (IAPS): Technical manual and affective ratings. *Gainesville FL NIMH Center for the study of emotion and attention University of Florida*, pages Technical Report A–6, 1999.

[49] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[50] R. Larson and M. Csikszentmihalyi. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15(15):41–56, 1983.

[51] P. Lenberg, R. Feldt, and L.-G. Wallgren. Towards a behavioral software engineering. In *Proceedings of the 7th International Workshop on Cooperative and Human Aspects of Software Engineering - CHASE 2014*, pages 48–55, New York, USA, 2014. ACM Press.

[52] P. Lenberg, R. Feldt, and L. G. Wallgren. Behavioral software engineering: A definition and systematic literature review. *Journal of Systems and Software*, page In Press, 2015.

[53] K. Lewin. *A dynamic theory of personality*, volume 2008. McGraw-Hill, New York, USA, 1935.

[54] F. Li, X. Bai, and Y. Wang. The Scale of Positive and Negative Experience (SPANE): psychometric properties and normative data in a large Chinese sample. *PloS one*, 8(4):e61137, 2013.

[55] H. Lövheim. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses*, 78(2):341–348, 2012.

[56] A. Medinilla. Process Kaizen Agile Kaizen. *Process Kaizen: Managing Continuous Improvement Far Beyond Retrospectives (Part 1)*, pages 85–119, 2014.

[57] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14(4):261–292, 1996.

[58] M. Minsky. A Framework for Representing Emotional States. In M. Lewis, J. M. Haviland-Jones, and L. Feldman-Barrett, editors, *The Handbook of Emotions*, chapter 38, pages 618–627. The Guilford Press, New York, USA, 3 edition, 2008.

[59] J. Morris and C. Waine. Managing the creative effort: Pre-production and post-production measures of emotional response. In *Proceedings of the 1993 Conference of the American Academy of Advertising*, pages 158–156, 1993.

[60] J. D. Morris. SAM: The Self-Assessment Maniking - An Efficient Cross-Cultural Measurement of Emotional Response. *Journal of Adverstising Research*, 35(6):63–68, 1995.

[61] P. M. Muchinsky. Emotions in the workplace: the neglect of organizational behavior. *Journal of Organizational Behavior*, 21(7):801–805, 2000.

[62] S. C. Müller and T. Fritz. Stuck and Frustrated or In Flow and Happy : Sensing Developers ' Emotions and Progress. In *37th International Conference on Software Engineering (ICSE 2015)*, 2015.

[63] A. Murgia, P. Tourani, B. Adams, and M. Ortu. Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In *11th Working Conference on Mining Software Repositories*, pages 262–271, New York, USA, 2014. ACM Press.

[64] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, New York, USA, 1 edition, 1990.

[65] B. Parkinson, R. Briner, Reynolds S., and P. Totterdell. *Changing moods: The psychology of mood and mood regulation*. Addison-Wesley Longman, London, 1 edition, 1996.

[66] R. Plutchik and H. Kellerman. *Emotion, theory, research, and experience*, volume 1. Academic Press, London, 1980.

[67] R Community. Conservative ANOVA tables in lmer, 2006.

[68] R. Reisenzein. What is a definition of emotion? And are emotions mental-behavioral processes? *Social Science Information*, 46(3):424–428, 2007.

[69] G. K. Robinson. That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6(1):15–32, 1991.

[70] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[71] J. A. Russell. Culture and the categorization of emotions. *Psychological Bulletin*, 110(3):426–450, 1991.

[72] J. A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172, 2003.

[73] J. a. Russell. Emotion, core affect, and psychological construction. *Cognition & Emotion*, 23(7):1259–1283, 2009.

[74] J. A. Russell and L. F. Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805–819, 1999.

[75] J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977.

[76] J. a. Russell, A. Weiss, and G. a. Mendelsohn. Affect Grid: A single-item scale of pleasure and arousal., 1989.

[77] D. Sato, D. Bassi, M. Bravo, A. Goldman, and F. Kon. Experiences tracking agile projects: an empirical study. *Journal of the Brazilian Computer Society*, 12(3):45–64, 2006.

[78] N. Schwarz. Feelings as information: informational and motivational functions of affective states. In E. T. Higgins and R. M. Sorrentino, editors, *Handbook of Motivation and Cognition: Foundations of Social Behavior*, number 89, chapter 15, pages 527–561. Guilford Press, New York, NY, US, 1990.

[79] N. Schwarz and G. L. Clore. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3):513–523, 1983.

[80] P. Shaver, J. Schwartz, D. Kirson, and C. O'Connor. Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52(6):1061–1086, 1987.

[81] A. J. Silva and A. Caetano. Validation of the Flourishing Scale and Scale of Positive and Negative Experience in Portugal. *Social Indicators Research*, 110(2):469–478, 2011.

[82] D. Sjoberg, T. Dybå, B. Anda, and J. Hannay. Building theories in software engineering. *Guide to Advanced Empirical Software Engineering*, 1(1):312–336, 2008.

[83] E. R. Thompson. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38(2):227–242, 2007.

[84] J. L. Tsai, B. Knutson, and H. H. Fung. Cultural variation in affect valuation. *Journal of personality and social psychology*, 90(2):288–307, 2006.

[85] G. R. VandenBos. *APA dictionary of clinical psychology*. American Psychological Association, 2013.

[86] D. Watson, L. A. Clark, and A. Tellegan. The Positive and Negative Affect Schedule. *Journal of Personality*, 8(6):1988, 1988.

[87] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988.

[88] J. Wegge, R. V. Dick, G. K. Fisher, M. a. West, and J. F. Dawson. A Test of Basic Assumptions of Affective Events Theory (AET) in Call Centre Work. *British Journal of Management*, 17(3):237–254, 2006.

[89] H. Weiss and R. Cropanzano. Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior*, 18(1):1–74, 1996.

[90] M. Zajenkowski, E. Goryńska, and M. Winiewski. Variability of the relationship between personality and mood. *Personality and Individual Differences*, 52(7):858–861, 2012.