

Daniel Graziotin
Faculty of Computer Science
Free University of Bozen-Bolzano
Piazza Domenicani 3
39100 Bolzano-Bozen, Italy.

July 23, 2015

Dear Philipp Leitner,

We are grateful for the editorial process so far, and for the generous and respectful reviews that we received.

The feedback has been rich and insightful to the point that—despite of the clear editorial line on which you expected us to focus on—we decided to respond every single comment raised by the reviewers, and we addressed the majority of their concerns in the revised text.

Best regards,

Daniel Graziotin
on behalf of all authors.

encl: Response to Reviewers.

Editor's comments

Thank you for this excellent contribution to PeerJ Computer Science. As you can see below, both anonymous reviewers have enjoyed the paper a lot. However, both have requested minor updates prior to publication.

Specifically, I ask you to address the comment by Reviewer 1 relating to the transparency of some of the interpretation steps. I agree with Reviewer 1 that, in general, the “minimum standards” of the field are easily fulfilled, but I think we could still do better.
Philipp Leitner Academic Editor for PeerJ Computer Science

This revision addresses all the concerns raised by the reviewers, which we comment and summarize in this letter. In particular, we expanded and added several explanations towards a higher transparency. In this revision, much work has been done to (1) address the request for chain of evidence of the data analysis process, (2) further develop the description of the reliability and quality of the data procedures, and (3) clarify about the few discrepancies that we encountered.

Overall, we feel that the manuscript was significantly improved by this revision.

Reviewer Comments

Reviewer 1 (Anonymous)

Basic reporting

1) I could not find where the authors have made available the intermediate results, such as the codes and other data. As it is presented, the “chain of evidence” for interpretive purposes is not very explicit. I understand that is the case for many IS research papers, but I would like to rise the bar and have a bit more concrete description of the intermediate results during the analysis/interpretation stage.

We thank the reviewer for pointing out this concern, which we feel was born due to a lack in explanations and examples when listing the four phases of coding. The revised version provides a significantly expanded Data analysis section, which elaborates on Charmaz (2006) chapter on coding in grounded theory. In particular, we provide in Figure 1 a chain of evidence of our coding activities that explains Charmaz (2006) steps. The text also explains that providing examples when going from focused coding to axial coding is impractical as it would require including a substantial part of the transcripts and the initial coding.

We also regret to being unable to provide access to the raw data for this article. The participants were not comfortable with sharing the recordings and the transcripts, because of the many personal issues and details disclosed during the interview process. As we are very careful with ethical and privacy issues, we will comply with their requests.

2) Some terms should be clarified, and some expressions need to be re-formulated: Examples: Page 2, line 64: what you mean by “removed treatment”?

The reviewer is right that we have here the opportunity to show a rarely employed design to software engineering researchers. The revised document provides a clarification of removed treatment designs

“Removed treatment designs are part of single-group quasi-experiment designs. A removed treatment design allows one to test hypotheses about an outcome in the presence of the intervention and in the absence of the intervention (Harris et al., 2006). A pre-treatment measurement is taken on a desired outcome; a treatment is provided; a post-treatment measurement is conducted; a second post-treatment measurement is conducted; the treatment is removed; a final measurement is performed (Harris et al., 2006).”

Page 3, line 123: please explain what you mean by “variance-based”

The revised manuscript references the seminal work in organizational behavior by Langley (1999) and by Mohr (1982) on theorizing from process data, which explains how process studies differ from variance studies.

“Variance theories, as opposed to process theories, provide explanations for phenomena in terms of relationships among dependent and independent variables (Langley, 1999; Mohr, 1982). In variance theory, the precursor is both a necessary and sufficient condition to explain an outcome, and the time ordering among the independent variables is immaterial (Pfeffer, 1983; Mohr, 1982). Strictly speaking, variance theory studies are hypothesis-driven studies, which aim to quantify the relationship between two variables in their base case.

Process research is concerned with understanding how things evolve over time and why they evolve in the way we observe (Langley, 1999). According to Langley (1999), process data consist mainly of ‘stories’—which are implemented using several different strategies—about what happened during observation of events, activities, choice, and people performing them, over time. Mohr (1982) has contrasted process theory from variance theory by stating that the basis of explanation of things is a probabilistic rearrangement instead of clear causality, and the precursor in process theory is only a necessary condition for the outcome.”

Page 3, line 125: “Our theoretical framework was primarily *based* on..” Page 11, line 503: “*A* more tangible focus..”

We thank the reviewer for these corrections. We included them in the revised text. We also ensured a stronger proof read before resubmitting the manuscript.

Experimental design

I am very curious about the four-staged coding process. In particular, I am missing some middle-level results when going from focused coding to axial coding. In particular, why are no examples of categories extracted via axial coding? you jump directly from 308 codes to 6 themes, which is a quite big jump.

As with our first comment to the reviewer, we feel that this concern was born due to a lack in explanations and examples when listing the four phases of coding. The revised version provides a further elaboration on Charmaz (2006) chapter on coding in grounded theory. Furthermore, we provide in Figure 1 a chain of evidence for our coding activities, which complements the elaboration on Charmaz (2006) steps. The text also explains that providing examples when going from focused coding to axial coding is impractical as it would require including a substantial part of the transcripts and the initial coding. This would make the text unreadable and Charmaz (2006) is of the same opinion when she cuts the example right after the focused coding.

In addition, I am interested on if there was any cross-validation of the coding (in particular during axial coding) across multiple researchers, or if it was only one researcher who did the entire interpretation.

Also, can you elaborate further on the difference between initial and focused coding? - maybe with some examples?

As with the cross-validation of the coding across the authors, one author performed the entire coding process. Towards the later stages of coding and model building, there were intensive discussion and collaboration among all three authors. We updated the main text to reflect on this.

The revised Data analysis section further elaborates on the difference between all four stages of coding, and it includes a running example from our data analysis.

When it comes to conflicting results between the manikin questionnaire and the qualitative data (here I am assuming the open ended interviews, the regular meetings and the observational sessions), how and when did you performed the clarifications when conflicting answers were identified? - an example could be very illustrative.

The reviewer is right that an example would clarify the origin of the discrepancies between the questionnaire and the qualitative data. The revised text clarifies that all three discrepancies were minor, and it includes the following case:

“For example, in one case the participant P1 reported low values of valence and arousal, and a neutral value for dominance. During the interview, P1 often stated that he had a frustrating day, but there were no mentions of low-arousal negative affects. When asked to explain how the Self-Assessment Manikin values were representative of the work day, the participant added that he felt low esteem, which was caused by episodes of frustration. Overall, P1 was unexcited and lost over the day; thus the reported low value for arousal.”

Validity of the findings

A major aspect that I am not extremely convinced in the proposed theory is the distinction made between Affects and Attractors. I believe the authors need to present a better case to justify the need of having two separate constructs, which appear to be difficult to clearly separate.

We agree with the reviewer that the manuscript could do a better job in describing the most novel part of the theory, which is attractors, and how attractors differentiate with affects. Therefore, we have expanded the subsection extensively, and we have offered a clearer view and added explanations. Furthermore, we did a better job in linking the interview quotes of that section to the sections of Focus and Performance. We thank the reviewer for pointing out this issue. We now feel that the section has greatly improved.

Also, I would suggest the authors to touch more actively upon how the specifics particular to the software engineering context affect the applicability of more general theories explaining the relationship of affect and performance. Basically, a more active comparison between the existing, more general theories with respect to the theory presented is advisable.

We positioned our work with the related theory of AET and existing SE work in section “Comparison of the theory with related work”. As for generalization, we see the reviewer’s concern as an immediate continuation. The Future works section quotes the reviewer’s comments in the revised paragraph

“As our study was explanatory, we suggest future research to test the proposed theory and to quantify the relationships in quantitative studies, in software engineering field but also in other domains to understand if and how the specifics particular to the software engineering context affect the applicability of our theory.”

In the limitations of the work, the authors need to discuss their choice of having only males in an inductive study. It is well known that gender plays a major role on outcomes and behavior related to affect, and they should at least discuss how their sample may have affect the dynamics of the project, and thus the narrative from which the theory is derived.

The reviewer raises here a very fascinating point that we definitely have to address in the manuscript. To our knowledge, some evidence has been found for a gender difference in emotion regulation but it is soon to consider this evidence as established fact. Indeed, we reflected on this in the Limitation section, which now contains the following added paragraph:

“The non-inclusion of female participants might be considered a further limitation of this study. There is a widespread popular conception that there are gender differences in emotionality (McRae et al., 2008). Evidence has been found for gender differences at the neural level associated to reappraisal, emotional responding and reward processing (McRae et al., 2008), and for a female having greater reactivity to negative stimuli (Gardener et al., 2013) and adoption of different emotion regulation strategies (Nolen-Hoeksema and Aldao, 2011). While more studies on gender differences are needed as the produced evidence is not enough yet (Nolen-Hoeksema, 2012), it might be the case that the inclusion of a female developer would have made the dataset richer, and perhaps would have led to a more gender-balanced theory.”

Comments for the author

This paper is well written, provides a good line of argumentation in relation to the relevance of the topic, it presents a sound methodology, and the results are presented in a clear, concise manner. The implications of affect on programming performance is a topic that it has caught up attention recently in SE, and the authors make a clear case for the “knowledge gap” they are trying to address. It was a pleasant experience to review this paper.

We further thank the reviewer for writing a respectful and insightful report, which brought us to improve the quality of the submission.

Reviewer 2 (Anonymous)

Basic reporting

This paper presents a theory that explains the affects that software developers have and how these affects impact their performance while programming. To come up with the theory, the authors conducted a qualitative study and collected data through observations and various kinds of interviews. Using grounded theory techniques, the authors then established an explanatory theory.

The authors do a very good job in motivating why their research and in particular a theory of the impact of affects on software developers is necessary. In the paper, it is clearly explained what the implications of such a theory are, why a better understanding is necessary, and how the theory can help to get a better understanding. The authors also provide reasons why such a theory was not yet established, and why in general researchers have not yet focused much on software developers affects. Furthermore, the authors also put a lot of effort into explaining the theoretical background of their theory construction approach.

The paper is also well structured and written in a way that makes it easy to follow.

However, in some cases, important details and explanations are missing.

For example, the authors state that there were three discrepancies between the qualitative data and the SAM data, but all of these three discrepancies could be resolved. It would be interesting to know how these discrepancies were resolved and why there were discrepancies in the first place.

We thank the Reviewer for the nice words expressed towards our work. This first issue was raised by Reviewer 1 as well. The revised text clarifies that all three discrepancies were minor. We also explained one of three cases in the text, as follows:

“For example, in one case the participant P1 reported low values of valence and arousal, and a neutral value for dominance. During the interview, P1 often stated that he had a frustrating day, but there were no mentions of low-arousal negative affects. When asked to explain how the Self-Assessment Manikin values were representative of the work day, the participant added that he felt low esteem, which was caused by episodes of frustration. Overall, P1 was unexcited and lost over the day; thus the reported low value for arousal.”

Regarding why there were discrepancies in the first place, this is an interesting question. It can happen in every study and they are not predictable and deterministic. This is part of the human nature. We can only do our best to limit the discrepancies.

Another aspect that needs clarification is the questionnaires that were used directly after a developer made a commit. What was the reason to let the participants fill out a questionnaire exactly at this point in time and what happened if a developer could not make any commit for a whole week (as mentioned in quote 16)? Didnt that lead to a loss of data for this participant? How did the authors handle this situation?

The only reason for employing the questionnaire was for enhancing the quality of the qualitative data, as the data from the questionnaire was not employed otherwise. We asked the developers to fill the questionnaire at code commit time in order to trigger an inner reflection about how they felt during the development task. We also aimed at being as less intrusive as possible, and to avoid burden them with research-related tasks while they were working.

As for the case of the developers who could not make any commit for a whole week, there was not really a loss of data for the participant. The participant there experienced work withdrawal (which is a well-defined construct in I-O psychology indicating poor job performance) and was not available for interviews. This apparent “loss of data” was instead a unique—alas, unfortunate for the developer—opportunity for us to analyze the loss of performance when he came back to work. There, he was fully committed to offer us explanations on why this happened, and began again working, committing code, and filling up the questionnaire.

Furthermore, for me it is unclear how the paragraph from line 379 to 391 fits into the section about developers affects.

We agree with the reviewer that this paragraph somehow “appears out of nowhere” if left at that place. We broke the paragraph into two new paragraphs. Lines 379-385 were moved at the end of the Affects section, because it provides a theoretical link between events and affects. Lines 385-391 were moved at the end of the Attractors section, because the text further helps us to clarify between affects and attractors.

Experimental design

The experimental design that the authors have chosen is a good fit for their research. It allows them to observe developers during their work, and using grounded theory techniques, come up with a theory that explains the observations. However, there are two issues with the experimental design.

The authors should clarify what exactly they consider to be an emotion. There is already a lot of research in this direction, and different authors have a different opinion on how to distinguish between emotions, feelings, moods, etc. The authors state that in their work, they use the term affect to include emotions and moods and that's fine, but I doubt that, for example, “feeling motivated”, or “feeling alone” would be considered neither an emotion nor mood by a lot of researchers in this area.

We agree with the reviewer that the manuscript as it was submitted failed to provide at least some background information on affects, emotions, and moods.

We added a new subsection under Background, called “Affect, emotion, and mood”. The section introduces the concept of affects and why we do not distinguish emotions with moods. The section is brief because we do not want to burden the reader with too technical details. However, the text points to two of our recent articles that are aimed toward clarifying the constructs of affects, emotions, moods, and the several misconceptions that arise when studying them.

As for the specific cases of “feeling motivated”, but also “satisfied” and so on, the reviewer is correct that stating that complex constructs like motivation are emotions would be considered as superficial by many readers. In addition to the added section about what we consider as affect, we added the following footnote at the occurrence of the first term identified by the reviewer, namely “motivation”.

“The careful readers might turn up their nose here. As we wrote in (Graziotin et al., 2015b), affects are not motivation, and they are not job satisfaction, etc. Yet, affects are important components of these psychological constructs, and studying complex multifaceted constructs like motivation would require different approaches and different measurement instruments. For this reason, if the participants only stated that they felt motivated or satisfied, we considered them as affects, as it might well be the case that they were expressing emotional judgments about such constructs. In any case, the inclusion or exclusion of such terms as affects would not change the results of this study.”

The other issue is about the sample size of the study and the environment in which the study took place. The authors state that theorists argue that theories should be simple in the beginning and by studying only two developers, they could keep the complex situation simple. Still, I'm a little bit concerned about the very small sample size and that the study took place in a separated work space. The two study participants did not interact with other developers, neither formally nor informally.

The Limitations section further highlights the limitations provided by the sample size (see also the comments of Reviewer 1 regarding the non inclusion of female developers).

In general, the authors describe their experimental design in a very detailed way that would make it possible for other researchers to reproduce it. However, there are also some minor open questions that would benefit from clarification. First, it would be interesting to know whether the interviews were conducted in English or if the authors translated the participants' answers.

We thank the reviewer for giving us the opportunity to further enhance the transparency of the article. The revised manuscript provides a new paragraph (in the Design section) that deals with the language of the study, as follows:

“The participants' native language is Italian, but they have been certified as proficient English speakers. The first author of the present article employs Italian as first language, as well, and he was the reference person for the participants for the duration of the entire study. The other two authors of the present article have been certified as proficient and upper intermediate Italian. The choice for the design of the study was therefore to conduct the interviews in Italian, as the native language let the participants express their opinion and feelings in the richest, unfiltered way (van Nes et al., 2010). The interviews were subsequently transcribed in English as suggested by the common research practices van Nes et al. (2010); Squires (2009), but the present case had the added value that the authors could validate the transcripts with the participants over the course of the study, given their advanced proficiency with English.”

Second, the authors state that they conducted 5 impromptu meetings. Could the authors clarify which events or circumstances lead to calling an impromptu meeting?

The revised manuscript now clarifies on the nature of the 5 impromptu meetings, with the following added text:

“The impromptu meetings were held mostly because of trivial issues, like casual chatting which turned into a proper interview. Only in one case an impromptu meeting was called by P2 when he finally came back to work after the week of absence. We also did not distinguish between the data coming from the scheduled meetings and the impromptu meetings.”

Third, for the affects, the authors performed a quantitative triangulation. Was there also such a triangulation done for the performance? If not, why?

There was no triangulation for the performance. One reason is that the nature of performance was not the primary focus of this study, but the relationships of affects and performance. Besides the fact that there is not a commonly accepted quantitative way to measure the performance of developers, the architecture, design, and implementation details of the system developed by the participants were not known during the design of the study as the developers were free in their implementation choices. For these reasons, it was unrealistic to build up a measurement environment suitable for data triangulation on performance.

Fourth, in the section “Theory construction and representation”, the authors discuss the issues that come with the development of a theory and state that their study design has taken these issues into account. Could the authors elaborate how exactly their study design has tackled these issues?

We were not clear enough in the original text. How we tackled these issues is actually fully explained in the Methodology section, which follows the claim “These issues were taken into account when we designed our study”. We extended the sentence to make the connection more clear.

Validity of the findings

The findings that the authors present are interesting and appropriately stated. The authors also discuss some implications of their findings for research but also for practitioners. However, there are also some open questions that the authors should address. First of all, the authors state that affects and focus are in general positively correlated. Did the authors consider that there might also be negative affects that have a positive influence on the focus? For example, a developer might get very frustrated because s/he cant fix a difficult bug, so that s/he puts a lot of effort in it, leading to a state of high focus that eventually helps to solve the difficult problem. Similarly, in Figure 1, it looks like events always trigger affects, but couldnt it also be the other way round?

The reviewer has raised interesting and fascinating issues here. However, as we explain in section “Happy, therefore productive or Productive, therefore happy?”, the theory has been constructed but only based on what we observed. As much as “we could not find support in the data for a double causation [happy → productive → happy and so on], but for a causality chain *Happy, therefore productive*”, any further discussion like the one of affects generating events, no matter how much sense it makes, is unfortunately out of the scope of this study, and would be considered a speculation without systematic evidence.

Furthermore, the category “Interventions” in Figure 1 seems to be very specific to this particular study setting, where a researcher could intervene when two bachelor students working on a software project couldnt cope with each other. What would be the real world analog?

The real world analog would be the one that we call as a mediator, as we explained at the end of the Interventions section. In the text, we specified that an agile coach or a team leader could be concrete examples of a mediator.

Comments for the author

I only have two minor comments: i) in the enumeration starting on line 158, number 3) is missing, and ii) the authors grouped affects in two different categories: positive and negative affects. Later they refer to these two categories as two different dimensions of affects. Is this really correct? For me it looks more like two different values or categories of affects. Wouldnt dimensions of affects be something like valence or arousal, i.e. different dimensions to measure an affect?

We thank the reviewer for spotting out the first issue. The revised text has a correct enumeration.

As for the second point, thanks to the reviewer for pointing out the inconsistency in the terminology we use to call positive and negative affects. They are dimensions rather than categories. We revised the text accordingly.

Regarding the dimensions suggested by the reviewer, the reviewer here is referring to the dimensional framework for expressing affects, but the case of valence and arousal (and sometimes also dominance) is part of the so called PAD models. The following three references are exemplar studies explaining the PAD models.

A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14(4):261292, 1996.

J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):11611178, 1980.

J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273294, 1977

The newly created section Affects, emotions, and moods opens up the discussion about the definition and categorization of the constructs. Moreover, it points to two recent studies that offer a review of the theoretical foundations of affects, emotions, and moods, namely Graziotin et al., (2015b,c)

We thank the reviewer for taking the time to evaluate the manuscript and for offering such a rich report, which let us improve the manuscript considerably.