Using the time-varying vector autoregressive model to study dynamic changes in situation perceptions and emotional reactions

Author's Response

Dear Editor,

On behalf of all the authors, I am pleased to submit a revised version of the manuscript JOPY-18-0302, entitled "Using the time-varying vector autoregressive model to study dynamic changes in situation perceptions and emotional reactions".

We appreciated the opportunity of submitting a revision you offered us. Based on the very insightful suggestions and detailed comments provided by you and the three reviewers, we have revised the manuscript. You will find below our responses to each issue. We hope you will evaluate this new version of the manuscript as suitable for publication for the Journal of Personality.

Since the revised manuscript includes many changes in the main text, we have decided not to highlight the modified parts using a different color. However, we used the relevant page numbers to refer to any change mentioned in our responses to reviewers' comments.

Editor

As I said, I have very little to add in terms of specific comments. Further, most of the reviewers comments are reasonably aligned, such that I don't see the need to adjudicate any conflicting requests. The largest one may be that R2 thinks your framing of the introduction is largely well-founded, whereas R3 points to various sources in the personality literature that should be incorporated. On this point, I would follow R3's advice, it will make your manuscript better received by the personality community.

Reply. We have added several references throughout the text to better root our contribution to the personality literature.

Finally, all reviewers commented that this manuscript could benefit from some native English speaker editing. It is impressive that the English language is so good already given it is not your native tongue, but given the highly technical nature of these materials I would request that this be given particular attention in a revision.

Reply. The manuscript has been proofread.

Reviewer: 1

I appreciate the overarching aim of this work, as it links the application of novel intra-individual time-varying analysis methods and research questions applicable to those methods, while at the same time, providing insight into substantive personality-context interactions. This aim is facilitated by a valuable and unique empirical data set.

Reply. We thank the Reviewer for the encouraging comments.

The description and implementation of the TV-VAR approach is quite vague and, at times, difficult to follow. I realize that details are provided in other manuscripts, but as written, this manuscript does not stand independently from those works.

1. The exact model for the TV-VAR is not clear. For equation 1, it is stated that all terms are time-varying parameters, but this does not align with other statements in the text. How can the error term (i.e., residuals) be "mutually independent" if they are allowed to vary across time (p. 11)? Also, the text refers to "time-invariant" parameters already being included in the model (p. 12) and the significance of time-varying parameters (p. 13), which suggests that time-invariant and time-varying terms are estimated separately; in that case, equation 1 seems to be missing key terms. It would be helpful to include equations for both the AR and TV-VAR models (e.g., as referenced in Table 3) in the text.

Reply. We agree this is a point we have to describe more thoroughly. In the revised version of the manuscript, we provided a more comprehensive explanation of the TV-AR model (we reasoned that this new slightly modified acronym is more appropriate and adapted the manuscript accordingly) that addresses all the points raised above. Specifically, in the new explanation of the model, we clarify that only the intercept and the autoregressive parameter are allowed to vary over time, while the variance of the error term (innovations) is not. Instead, the innovations should follow a white noise process over time. We also specified that only a time-invariant intercept is automatically (but implicitly) included in the TV-AR. Furthermore, we followed the suggestion to include the equation for both the TV-AR and the standard AR models.

2. What software or programs were used for data analysis? There seems to be one since reference is made to plots being generated for parameters (p. 13).

Reply. All the analyses were conducted using R and, more specifically, the mcgv (Wood, 2006) and tvvarGAM (Bringmann & Haslbeck, 2017) packages. We have included this precision (p.14).

3. Do I understand correctly that a separate model was run for AR and TV-VAR for each situational dimension/emotion of each participant (so over 50 models)? How does this influence the interpretation of the significance of time-varying parameters?
Relatedly, do I understand correctly that no cross-dimension/emotion relations are computed (e.g., there is no estimation of Intellect on Duty ratings across time)? If so, this seems important to note as a limitation of the model, as these processes are not unfolding independently.

Reply. We ran separate AR and TV-AR models for each situational dimension and emotion of both participants. We did not correct for multiple tests, but we decided to follow a specific process in our way of considering the significance of each result. As a preliminary note, each individual is considered as a single study. Moreover, we first examined the information given by the BIC to indicate whether TV-AR was better or not than the AR. If the AR was better, we did not investigate further and concluded for a time-invariant process. If the TV-AR appeared to perform better based on the BIC, then we examined the significance of the intercept and the autoregressive parameters. This parsimonious and conservative approach allowed us to consider only non conflicting and thus reliable results. We have detailed the process in the revised manuscript and presented results fitting these criteria.

Although the TV-AR would allow testing for time-varying processes in situation perception and emotion relationships, we did not investigate them. First, it was not the focus of our contribution. Moreover, it would increase the number of statistical tests performed on the data for each participant. We acknowledge this possibility in the discussion as a potential future direction in the use of the TV-AR (p. 20).

4. There seems to be a good deal of subjectivity in the model selection process (e.g., as research team

consensus discussion was required in several cases, p. 13). This may be inevitable, but these procedures need to systematized, so that they can be replicated.

Reply. In the model selection process, there are some specific criteria to consider, as indicated by Bringmann et al. (2017, 2018). However, as we acknowledged in the previous point, we have decided to adopt a more conservative approach for didactic purposes and to follow a specific process to consider results as indicative of a time-varying process. We thus believe that the part of subjectivity is now largely reduced. Furthermore, data and R code for the analyses are available at the link provided in the manuscript (p. 15).

5. The authors only briefly mention the existence of other time-varying analytic approaches, but there are several (e.g., see those implemented in R's dynr package). Considering the pros and cons of the TV-VAR approach in light of alternatives is vital for contextualizing findings.

Reply. We thank the reviewer for the useful suggestion. In the revised version of the paper, we included a paragraph about other possible time-varying analytic approaches in which we explained the advantages of the TV-AR model compared to the others (p. 6-8).

6. Why were data from only 2 participants analyzed for this manuscript, and how were these 2 participants chosen (from the sample of 198)?

Reply. We chose to use only 2 participants for the sake of brevity. With this contribution, we aim at illustrating the usefulness of the TV-VAR and using 2 participants who differed in terms of some personality dimensions is sufficient to reach this aim. Regarding the particular 2 participants, our choice was based on two criteria. The first criterion was statistical. Ideally, to detect a time-varying process, one needs around 100 data points (Bringmann et al., 2017). Our design led to a maximum of 95 data points, and only five individuals showed more than 80. The second criterion was theoretical. We were interested in studying adolescents with different personality traits to detect whether such a difference might be mirrored in differences in their interaction with situations. We have included these choices in the text (p. 11-12).

7. How were missing data handled? The two participants had different numbers of observations, and it is unclear how this may affect results. The authors also mention that "additional missing data points were imputed at the end of the day" (p. 12). So, how many observations were in the time series in total, and what imputation approaches were used? There is great debate surrounding time series imputation.

Reply. The note specifying "additional missing data points were imputed at the end of the day" was a mistake because no data were imputed. Participant 1 completed the 5 data entries per day except for the last day for which he did not complete the last data entry leading to a total of 94 data points. Participant 2 ended the data collection two days before Participant 1, leading to 84 data points. Therefore, there were no missing values within each day. The analyses were, therefore conducted on the 94 data points for Participant 1 and 84 for Participant 2. We have specified this issue in the text (p. 12).

8. It would be helpful to provide example items for each of the measures.

Reply. We have included example items for each measure (p. 10-11).

9. It is not clear how personality differences in the HEXACO dimensions were determined between the two adolescents. Based on the in-text description, a z-score (with standard errors informed by a larger group)

appears to have been used, but the data in Table 1 do not align with this (e.g., participant differences scores – not standard errors – are not shown).

Reply. Differences between the two individuals were tested using standard errors. We apologize for the mistake in the title of Table 1 that might have been perceived as contradictory to what was written in the text. Table 1 reported mean scores and standard errors of difference.

10. The authors claim that the "TV-VAR model was superior to the standard time-invariant model" (p. 17). On what information is this conclusion based? One participant had no significant time-varying components in his models, so how could the time-varying model be superior for him? Relatedly, the authors claim that this same "Participant 2 showed an adaptive psychological functioning, especially characterized by a flexible life approach" (p. 18). How is this interpretation linked to the absence of time-varying effects?

Reply. We have toned down our conclusions regarding the TV-AR and rewritten the discussion to avoid speculative interpretation.

11. There are typos throughout the manuscript. For instance, only 7 dimensions are listed for the Situational Eight DIAMONDS model (p. 4). Also, the grammar and writing is unclear in many places, making it difficult to follow the conceptual argumentation. Examples include "only a few research as dealt statistically with" (p. 6), "It allows shedding light on" (p. 8), "heterogeneity within each people and time courses underlying their functioning" (p. 20).

Reply. We have corrected the typos, and the entire revised manuscript has been proofread.

Reviewer 2

The current paper provided a method for testing person-level time-varying autoregressive models to control for non-stationarity using situation and affect data. Overall, I think there is a critical need for the implementation of idiographic time series models within personality. There are a number of theoretically grounded and important questions that are yet unanswered. However, I have a number of concerns about the structure of the paper, its goals, and its conclusions. I think that this paper could be a valuable contribution to the personality literature, but my read is that this paper requires very substantial revision.

Reply. We thank the Reviewer for her/his encouraging remarks.

1. The introduction proceeds from the observation that (1) person-situation interactions are important in personality and (2) personality has essentially exclusively studied these questions from a nomothetic perspective. There are issues with both of these discussions.

a. The paper does not actually test person-situation interactions, unless the argument is that by looking at individual differences in the intercept and autoregressive functions across different psychological characteristics of situations, this is the interplay of personality and situations. This argument is not made in the paper. In addition, the literature on person-situation interactions (e.g. Sherman et al., 2015) study how personality predicts situation characteristics, not just individual differences in situation perception, concluding these characteristics are independent predicts of behavior.

Reply. Although the interplay of personality and situations is embedded within the person-situation perspective, we agree that the former is more adapted to describe our work. We have investigated potential

changes in intra-individual dynamics in the perception of situations and the emotional reactions of individuals varying in personality traits. We have modified the Introduction and the Aim of the contribution to avoid potential confusion.

b. Interest in idiographic personality (and idiographic personality systems) stems back to the earliest days of personality psychology (c.f. Allport, 1937 and Cattell, 1957). Moreover, it misses a number of idiographic studies specifically examining person-situation interactions (see work by Albert Bandura, Dan Cervone, Jack Wright, Phil Peake, Yuichi Shoda, and others for pre-ESM approaches to these studies).
Reply. We have included the earlier work on idiographic personality.

2. Stationarity is an oft ignored but very important consideration in time series modeling. However, the authors need to make a compelling case for why TV-VAR is a better alternative to other methods like detrending or using residuals (cite) from a time series (cite).
Reply. We thank the reviewer for raising this point. In the revised version of the paper, we included a paragraph about other possible time-varying analytic approaches in which we explained the advantages of the TV-AR model compared to the others (p. 6-8).

3. Inertia is time series measure not often examined in personality (although Cattell actually discussed it in his 1957 book), but the reasoning for looking at inertia in personality in particular was not immediately clear here in a manuscript on "person-situation interactions." The introduction should include a theoretically meaningful explanation of why this matters for personality.
Reply. We have included a section in the Introduction dedicated to inertia and why it is important to study it (pp. 4-5).

4. Based on the paper, it is not clear why two individuals were chosen for examination – that is, no compelling argument was made to why the rest of the sample was not also examined and variability in the parameters compared.
Reply. We chose to use only 2 participants for the sake of brevity. With this contribution, we aim at illustrating the usefulness of the TV-AR and using 2 participants who differed in terms of some personality dimensions is sufficient to reach this aim. Regarding the particular 2 participants, our choice was based on two criteria. The first criterion was statistical. Ideally, to detect time-varying processes in terms of small changes, such as a small linear increase over time, one needs approximately 100 data points. However, 60 data points are enough for detecting larger changes over time (Bringmann et al., 2017). Our design let to a maximum of 95 data points, and only five individuals showed more than 80. The second criterion was theoretical. We were interested in studying adolescents differing in personality traits to detect whether such a difference might reflect also differences in their situation perceptions and emotions. We have included this precision in the text (p. 11-12).

a. Relatedly, adolescent personality has a limited but important literature that is not discussed. The authors might want to use some of that literature in order to motivate why they studied individuals at this age and how that might influence the results that they found.
Reply. We agree with the reviewer that adolescent personality has not been studied as much as for the adult

population. We have included a few more references for justifying our choice in the Aim section (p. 8-9) and address this point in the Discussion (p. 18-21). However, we have decided not to extend too much on this point, considering that it was not the main focus of the paper.

5. In the discussion section, after a manuscript dedicated to idiographic trends, the authors note "Finally, another limitation of the TV-VAR model is its idiographic approach, that is, it considers one individual at a time." In my read, this comment comes off as dismissive of the model as a whole. The authors should consider reframing this in terms of what the idiographic model offers us. Do you think that group-level results would be meaningful in a case like this where you are testing for a (1) changing mean, (2) changing relationship among the predictor, and (3) need for a time-varying model at all? I would caution careful wording here, particularly since justification for using this method was the lack of attention to idiographic questions. Reply. We agree with the Reviewer that our phrasing might have been perceived as dismissive. We corrected the manuscript to emphasize the importance of considering the idiographic and the nomothetic approach for a better understanding of personality and temporal dynamics in everyday life functioning jointly.

6. If published, most readers of this paper will have limited background on TV-VAR models and are unlikely to return to source articles to understand them more fully. I found the description of the autoregressive function in particular to be lacking. A number of readings might not have strong backgrounds in autoregression in general, so the idea of a changing autoregressive relationship might be missed completely despite being foundational to the paper. In particular, the description of the time-varying autoregressive function on p. 12 appears to be the same as the intercept.
Reply. We have followed the suggestion and explained the concept of autoregression or inertia in the Introduction when introducing the concept of temporal dynamics (pp. 4-5).

On the whole, I would like to suggest a number of papers across several fields and labs whose inclusion could significantly improve this paper.
Beck, E. D., & Jackson, J. J. (2019). Consistency and Change in Idiographic Personality: A Longitudinal ESM Network Study. https://doi.org/10.31234/osf.io/pb92q
Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. Assessment, 23(4), 447-458.
Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time?. Journal of Research in Personality, 32(2), 202-221.
Cervone, D. (2005). Personality architecture: Within-person structures and processes. Annual Review Psychology, 56, 423-452.
Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2008). Integrating dispositions, signatures, and the interpersonal domain. Journal of Personality and Social Psychology, 94(3), 531.
Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. Psychological Review, 102(2), 246.
Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement, 2(4), 201-218.
Wright, A. G., Gates, K., Arizmendi, C., Lane, S., Woods, W., & Edershile, E. A. (2018). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and

psychopathology.

Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: the local predictability of social behavior. Journal of Personality and Social Psychology, 53(6), 1159.

Reply. We thank the reviewer for the useful references. We have included many of them in the text.

Reviewer 3

This manuscript makes a very important methodological, statistical, and theoretical contribution to the study of personality and person functioning within different situations. The Introduction makes a succinct and well organized, if not compelling argument for the importance of studying individuals intra-individually using an idiographic approach with time-series EMA data, often referred to as intensive longitudinal data. The authors explain that virtually all past research in this vein has used autoregressive (AR) or vector AR (VAR) approaches with models that assume invariance or stationarity of the AR process over time. That is, the direction and magnitude of yesterday's level of anxiety on today's is assumed to be constant throughout the course of the time series. The authors argue how a newly articulated Time Variant VAR (TV-VAR) model can estimate changes in the intercept, AR, and cross-lagged regression coefficients using Generalized Additive Models. They present the results of these analyses for two older adolescents who complete EMA assessments five times a day for 19 days with ratings along 8 dimensions of situational characteristics or appraisals using the DIAMONDS measure and of five emotions (e.g., anxiety, anger). As predicted based on the differing scores of the two participants on the HEXACO Personality Inventory, the intra-individual processes for the two adolescents were quite different, highlighting the capability and value of the TV-VAR approach. As mentioned, in my opinion this article makes several very important methodological, statistical, and substantive contributions. However, I would like to see addressed a number of concerns, methodological, statistical and substantive. The latter involves, in particular, interpretations of processes presented in the Discussion, Introduction

Reply. We thank the reviewer for her/his positive remarks.

1. p. 5: "Note that the more the data points and the more extended the assessment period, the more the intra-individual processes will unfold themselves …" Whereas I would generally agree with this statement, there are several issues or limitations to this statement that should be acknowledged: a) This statement is only accurate to the extent that the sampling method and interval – whether time or event based, and how often the participants are assessed with the EMA – matches that optimal approach and intervals to detect the intra-individual process. Sampling intervals may be too fine grained (often), resulting for instance, in an artefactual blowing up of the inertia parameter estimates because the ratings on that variable are relatively stable over shorter periods. Alternatively, important processes may occur within a time frame that is substantially shorter than the 2 hour sampling interval, for instance, changes may occur in situation perception or in emotional states in a 10 to 15 minute time frame. Such changes would simply be missed with a 2 hour sampling time frame. See Haynes, O'Brien, and Kaholokula (2011) for a good summary discussion of these issues. These concerns need to be recognized and mentioned.

Reply. We agree with the reviewer, and we have followed the suggestion by including some conditionality to the initial statement (p. 20).

b) Additionally, the EMA assumes that the rating variables are important and relevant for that individual(s). For

instance, there is evidence that the "Deception" rating on the DIAMONDS is probably not. Alternatively, have variables that are important for the intra-individual process for that individual been included for the ratings made by that person? This is an issue that is discussed in, for instance, Haynes, Mumma, and Pinson (2009) and should be mentioned in the discussion (see below). In short, the authors' statement here is probably accurate provided these two conditions are met. As mentioned, aside from this issue I found the Introduction to be well-reasoned, clearly presented, and making a rather compelling argument for the need for the intra-individual approach to examining dynamic processes.

Reply. We have included a section in the Discussion about the relevance and importance of some EMA items in our research (p. 20).

Method

1. p. 10: HEXACO: Authors report Cronbach's alphas – are these from the Ashton & Lee, 2009 study?
Reply. The alphas initially reported referred to the whole sample from where the data were extracted. We have removed these indexes from the text and included the ones from Baiocco et al.'s data that were used to determine the differences between the two participants on the HEXACO dimensions (p.12).

2. p. 11, last line; Consider inserting (italics): "level of situation characteristics at time t is thus dependent …". This should help the reader understand that this "mean" is time-varying.
Reply. We have followed the Reviewer's suggestion.

3. p. 12: "Since a time-invariant parameter is already automatically included in the model, a significant time-varying intercept could be interpreted as a significant change in the intercept over time;…". Having already discussed in their (1) in that paragraph issues related to a timevarying intercept, I believe the authors meant to say in their (2) "…a significant time-varying autoregressive parameter could be interpreted as a significant change in the direction or magnitude of autoregression over time:…" Correct?
Reply. We thank the Reviewer for raising this point. It made us understand that our explanation was not clear enough. In contrast to the intercept, a significant autoregressive parameter merely means that it deviates from zero. To see if it is time-varying, one can look at the edf and visually inspect the plot of the autoregressive function. We have now clarified this in the text.

4. p. 12: a) The authors indicate that "missing data points were imputed at the end of the day …" which is a good practice given the interval between the 5th rating on one day and the first rating of the next day is substantially different from the 2 hour interval between the other ratings. Pl. indicate how these "missings" were imputed. Also, for each variable please indicate how many missings were imputed (e.g., a single one or multiple missings) and what the likely effect on the results for would be.
b) Similarly, please describe here or elsewhere how the approximately 11 embedded missings for pt. 2 were handled in the data set and analyses. Describe any pattern here: e.g., did they occur in a couple of clusters, at particular times of the day, early vs. late in the time series, etc. Were these imputed also? How?
Reply. The note specifying "additional missing data points were imputed at the end of the day" was a mistake because no data were imputed. Participant 1 completed the 5 data entries per day except for the last day for which he did not complete the last data entry leading to a total of 94 data points. Participant 2 ended the data collection two days before Participant 1, leading to 84 data points. Therefore, there were no missing values

within each day. The analyses were, therefore, conducted on the 94 data points for Participant 1 and 84 for Participant 2. We have included this information in the text (p. 12).

5. p. 12: It might be useful to remind the reader that the BIC is a parsimony-adjusted fit index so that comparisons across models that differ in parsimony is meaningful. Perhaps in the sentence beginning with "The BIC function indicates …"?
Reply. We have included this useful precision in the text (p. 14).

6. p. 13, end first paragraph: "…the issues that may have contributed to these conflicting results were individuated for each adolescent." I'm unclear what "individuated" means here and what the results of this are.
Reply. We have changed the process of examining the results following a sequential process. We first examined the information given by the BIC to indicate whether TV-AR was better or not than the AR. If the AR was better, we did not investigate further and conclude for a time-invariant process. If the T-VAR appeared to perform better based on the BIC, then we examined the significance of the intercept and the autoregressive parameters. This parsimonious and conservative approach allowed us to consider only nonconflicting results as reliable. We have detailed the process in the revised manuscript and presented results fitting these criteria.

7. p. 13 last paragraph: "Intra-individual changes in mean scores of less than 5 points (< 5% of the 0-100 scale) …:" I'm confused. Earlier the authors indicate that ratings for the DIAMONDS and emotional states were made on a 1 to 7 scale (not 0 – 100 scale)!
Reply. We apologize for this confusing sentence. The scale was referring to the number of data points. We have corrected it and it now reads as follow:"Intra-individual changes in mean scores of less than 5 data points (< 5% of the total 95)....".

Results
1. Table 1: a) The two columns labelled "SE" appear to the scores for Pt 1 and 2, correct? Pl. change column label.
b) For Emotionality, the SE diff is reported as 0.49 yet the difference of 0.6 between the two scores has a "*". Please clarify.
c) 'n' column. Given that these are the same for each variable for each of the two participants, remove the "n" column and replace with a Note to the table indicating number of occasions for each participant.
d) It would increase readability of the table if there was an extra row between the 8 DIAMONDS ratings and the 5 Emotions ratings.
3. Table 3: a) Please clarify that the entries in the two columns labelled "Intercept" and "Autoregressive coefficient" are the F tests for the significance of the smooth functions for the TV parameter tests. Perhaps indicate this in the table's title?
b) Indicate either in the column header or via Note that the entries in the "Model fit" columns are the BIC.
c) For pt. 1, Duty, the AR model BIC of 281.85 should be bolded.
d) Readers familiar with the tests for the smooth functions may find inclusion of the edf useful to evaluate whether the test for the AR smooth functions suggests a time varying model. Thus, I'd suggest including the edf for these tests for both the Intercept and the AR coefficient.

e) For models that are time invariant, please report the parameter estimates for the Intercept and AR coefficients in Table 3.

f) Given d) and e) above and, thus, the suggestions to add four columns of information, I'd suggest breaking this table into two separate tables – one for Participant 1 and another for Participant 2.

Reply. We have made all the changes. In particular, we have followed the Reviewer's suggestion of including edf and intercept and autoregressive t tests values. For Table 1, we have included more details in the text as well as in the Table to avoid confusion about the meaning of the results.

4. P. 14, Time-varying analyses: a) "Results concerning emotions showed the intercepts of all the emotions to be time-varying for Participant 1…" but Table 3 indicates this is not the case for Anger (with F = 1.36).

b) cont.: "…while for Participant 2 only those [intercepts] of Anxiety and Hurt were timevarying." However, Table 3 indicates that the Happiness intercept was also time varying.

Reply. We have corrected the errors and modified the text accordingly.

c) For the results for Deception in this paragraph, please refer to Comment #6 later.

Reply. Please see our response to #6.

5. What is the likely effect of embedded missing on results, especially for pt. 2? For example, this brings the n down to well below the recommended 100 observations, how is this likely to influence the results of the TV tests for him?

Reply. Ideally, to detect time-varying processes in terms of small changes, such as a small linear increase over time, one needs approximately 100 data points. However, 60 data points are enough for detecting larger changes over time (Bringmann et al., 2017). Moreover, as noted earlier, there was no missing data within each day for Participant 2, only two days worth of data points (= 10) less than Participant 1. It is why we do not think that the differences in data points between the two participants affected the results. We have included these details in the text (p. 12).

6. a) Note that the ratings by both participants on Deception are highly restricted in range, and have very high skewness and kurtosis. A rule of thumb in many statistical analyses is to not analyze the data from such highly skewed and kurtotic distributions (generally accepted cut points are |3.0| and 10.0 for skewness and kurtosis, respectively). This problem may explain the very large results of the smooth functions tests for Pt. 2 on this variable, and may explain the inconsistency between these tests and the BIC fit results. Note that a similar problem/issues are present for Pt. 2's Hurt ratings and Angry ratings also.

b) For Pt. 1 on Deception: The very low F test results (close to 0) may suggest a violation of assumptions or otherwise invalid or problematic results for the smooth functions tests. These results, as indicated above, may be due to the very small variance (SD = 0.10) and very large skew and kurtosis for the Deception ratings by Pt. 1.

c) Thus, I'd recommend that the authors consider reporting the descriptive rating results for the above dimensions in Table 2 but removing these ratings from the analyses for the stationary and TV models. This could be explained in a note to Table 3 or in the text.

Reply. We agree with the Reviewer and decided to 1) include a short paragraph on the normality issues for these two variables at the beginning of the Analyses section (p. 12), and 2) remove from Table 3 the results

regarding Deception for both Participant 1 and Hurt and Angry for Participant 2.

Discussion:

A. Issues related to interpretation of the results:

1. General issue: It is difficult to follow the interpretations for the rating dimensions that do not have the plots of the TV Intercept, AR estimate, and inferred TV mean. Please include these, perhaps as an Appendix for those ratings discussed in this section. These, of course, only need be done for the ratings that are TV. As suggested above, estimates for the time-invariant intercept and AR parameter can be added to Table 3. Reply. We have included all the plots in the Appendix as well as the AR intercept and AR parameter in Tables 3a and 3b.

2. p. 17, middle of main paragraph: "This means that Participant 1 constantly attributes a negative valence [Negativity rating] to the situations he encounters." First, please include the plots for Negativity. Second, assuming that the TV AR parameter estimates are generally high (as suggested by the interpretation), the indicated start and end point of the TV Mean (4.8 -> 1.1) indicates that his Negativity ratings are, at least some of the time, low. The authors' interpretation confounds apparently high AR TV parameter estimates with high ratings on this dimension. However, these ratings are not high on average (M = 2.37, SD = 1.23 with minimal skewness) and appear to cover a substantial part of the rating scale (as indicated also by the start/end point of the TV Mean estimates. A better description would be something like: "Although Participant 1's Negativity ratings vary from low to (moderately) high, they seem to have high inertia overall."

3. p. 17 "Although he showed variation in his perception of situations as positive, this did not have any positive impact on the perception of subsequent situations." I'd suggest rephrase the latter part as "…, this did not have any impact on the perception of positivity on the subsequent rating ([indicate the estimated AR time-invariant parameter = 0]).

4. p. 18, "Participant 2 seemed to be prone to perceive situations as deceptive and not to trust others." But the average Deception rating = 1.04 with a SD = 0.33! This is a clearly inaccurate interpretation. See also above recommendations regarding the Deception ratings and tests.

5. p. 18, same paragraph: a) "Participant 2 "Indeed, he was alerted to perceive different psychological characteristics of situations…" I'm not sure how "was alerted" is relevant to the interpretation of his DIAMONDS results.

b) "…and motivated to respond to all the potential challenges he could encounter during everyday life. This finding …" This description is not a finding and seems, to me at least, as quite speculative and not necessarily supported by the intra-individual results.

c) ",,,on Anger, Anxiety, and Hurt. Participant 2 showed inertia on these negative emotions."

i) As mentioned above, I'd consider not including the TV tests of ratings for Hurt for Pt. 2 given their limited range, and very high skewness and kurtosis.

ii) For Participant 2, the distribution of Angry ratings is also highly skewed and kurtotic.

iii) Again for Pt. 2, given the significant results for the time varying AR coefficients on Anxiety and the plots in Figure 2 which support that the autoregression is TV, seems rather antithetical to the goal of this article to interpret the results simply as he "showed inertia" Rather, the TV AR results support the importance of conducting the tests and inspecting the figures for the ratings (which indicate TV autoregression).

iv) end of paragraph: "while his high levels of extraversion might be responsible for his being more prone to

feel angry." But his median Anger rating is a 1.00 and the average is rather low also.

Reply. We have rewritten the discussion of the results in line with the change in the approach of considering results as indicators of time-varying or time-invariant processes. Moreover, we have limited the interpretation of some results to avoid speculation.

B. General Discussion Issues

1. Please discuss or at least comment on the effects of the embedded missings on the results for Participant 2. For example, each embedded missing disrupts the estimation of lagged effects because, assuming the missing is at time "t," lagged effects from t – 1 to t cannot be estimated and neither can lagged effects from t to t + 1. Could this reduction in the n to estimate lagged effects partly explain the inconsistent results between the BIC and smooth function tests for TV Intercept and AR for him for Anxiety and Happy?

Reply. The note specifying "additional missing data points were imputed at the end of the day" was a mistake because no data were imputed. Participant 1 completed the 5 data entries per day except for the last day for which he did not complete the last data entry leading to a total of 94 data points. Participant 2 ended the data collection two days before Participant 1, leading to 84 data points. Therefore, there were no missing values within each day. The analyses were, therefore, conducted on the 94 data points for Participant 1 and 84 for Participant 2. We have included this information in the text (p. 12).

2. Idiographic and Nomothetic: The authors nicely summarize some of the issues relevant to this in the Introduction. In the Discussion please note that a) the measures and items used for the EMA ratings were from standardized measures and were thus nomothetic in their development. This may explain why some of the ratings had very little variability and may not have been seen as relevant by one or both participants. It is possible that a more idiographic approach to measurement might address this issue.

b) More generally, the authors discuss the idiographic approach solely in terms of the pattern of intra-individual results and make no mention of this other major domain (idiographic measurement and assessment) relevant to an idiographic approach. Issues related to idiographic measurement and assessment pertinent to clinical psychology have also been referred to as "patient-centered assessment" and been discussed by several research groups (e.g., Haynes and colleagues; Mumma and colleagues; Weisz, Chorpita and colleagues).

Reply. We thank the Reviewer for raising this issue. We have included a section in the Discussion about it.

3. To emphasize, although I am suggesting a number of changes, most of these are readily doable. Overall, as indicated above, I think this article makes very important contributions to psychological science in the methodological, statistical, and substantive domains.

Reply. Thank you again for the insightful comments and the encouragements. We hope the revised version will meet the high standards of the journal.