**Editorial Requirements:**

1. Delete the sub-heading “Introduction”.

We have deleted the sub-heading “Introduction.”

1. Do not use the expression "and colleagues" as a substitute for "et al.". The former is used when referring to more than one study by an author - typically when that author has taken a lead role. If you are referring to a single study by that author, use the expression "et al.". You can also list more than one study by simply adding the date such as in "(Smith et al., 2002, 2004)".

We have substituted “and colleagues” with “et al.”

1. According to APA style, use past tense to refer to work that you have already conducted (final section of your introduction).
2. Just like you write “Table 1” and not “table 1”, write “Study 1” and not “study 1”.

We have corrected this typo by writing “Study 1” and “Study 2” with uppercase.

1. When describing the Likert-scale labels such as "not at all", use italics instead of quotation marks.

We have corrected this typo by writing the Likert scale labels in italics.

1. Present "1.", "2." etc as "(1)", "(2)" etc.
2. Throughout the entire manuscript, including tables and figures, italicize variable labels such as "p", "n", "d", "g", “d”, “β”, and also "M" and "SD".
3. Please add doi numbers for all articles, if available.

**Reviewers' Comments:**  
**Reviewer #1:**

Thank you for the opportunity to review this paper. The paper examines Neff's Bipolar Continuum Hypothesis with two longitudinal studies that use Ecological Momentary Assessment (EMA). The studies are novel and contribute significantly to our understanding of the dynamic nature of self-compassion. In particular, I was impressed with the design of study 2 which allowed the examination of how context impacted on state compassionate and uncompassionate self-responding in real-life contexts. I do have some comments that I hope the authors will consider.

**Point 1.** In the abstract, the number of participants was noted as 494. Please amend this to indicate clearly the number of participants for each of the two studies (i.e., Study 1; n = 326 and Study 2; n = 168).

**Point 2.** The introduction provides a reasonable rationale for testing the bipolar continuum hypothesis using EMA. However, I wonder if the authors could provide a brief review of recent self-compassion research using EMA (e.g., Mey et al., 2023; DOI: 10.1007/s12671-022-02050-y, which was in the reference list but was not cited) and explain how previous research might inform hypotheses.

Thank you for raising this point. We have added a comprehensive literature review on state self-compassion and ecological momentary assessment (EMA) to the introduction section.

Specifically, we have discussed the study of Mey et al. (2023) and cited studies by Biehler & Naragon-Gainey, 2022; Ewert et al., 2022; Sahdra et al., 2023, as follows:

*The conceptualization of self-compassion as a dynamic process mirrors broader developments in psychological science that emphasize state-dependent variability over trait stability. Ecological momentary assessment (EMA) methodology has been particularly instrumental in capturing these temporal dynamics. Mey et al. (2023) showed that fluctuations in momentary self-compassion predict concurrent changes in affect and stress reactivity, with uncompassionate self-responding emerging as a particularly robust predictor of negative emotional states. This finding aligns with a growing body of research examining the temporal covariation between self-compassion and related psychological processes. Multiple studies have identified robust associations between momentary self-compassion and various adaptive outcomes, such as enhanced mindfulness, reduced stress reactivity, and improved well-being (Biehler & Naragon-Gainey, 2022; Ewert et al., 2022; Sahdra et al., 2023). The consistency of these temporal relationships supports a reconceptualization of self-compassion as fundamentally dynamic rather than dispositional*.

**Point 3.** Please consider including hypotheses for both studies. I understand that traditional hypotheses are not necessary when using Bayesian analyses, but it might be helpful. Furthermore, hypotheses are mentioned in the results sections so it would be best to report them in the introduction.

We have revised the introduction section and, based on Ferrari et al. (2022), we have clarified our hypothesis in this section of the Introduction:

*Building on Ferrari et al.’s (2022) suggestions, we conceptualize self-compassion as a dynamic process responsive to contextual influences. We test three specific hypotheses:*

1. *Trait-level CS and UCS may demonstrate greater independence over extended periods, while state-level components may show stronger temporal coupling.*
2. *Stressful or negative contexts may enhance the bipolarity between components.*
3. *The relationship between components may vary across individuals, necessitating person-centered approaches to self-compassion research.*

**Point 4.** Method - please provide more information about how psychology students were recruited. How were participants informed of the study? Was participation for course credit?

In the Method section, we specify the recruitment strategies as follows:

*Recruitment was conducted through university advertisements, and no incentives or course credits were provided for participation.*

**Point 5.** The authors note one additional item each for the state CS and UCS scales. Please note what were these items?

In the Method section, we added this specification:

*For the CS component, we added the item 'In this moment, I am able to accept my flaws and weaknesses' (In questo momento riesco ad accettare i miei difetti e debolezze); for the UCS component, we added the item 'In this moment, I let myself be carried away by my emotions' (In questo momento mi lascio trasportare dalle mie emozioni).*

**Point 6.** Pg 11 Please clarify "The posterior distributions were derived from at least 2,000 samples across four chains". Do you mean 2000 samples in total or 2000 samples per chain = 8000 in total?

We have clarified that in total we have 8000 samples from the posterior distribution. The revised sentence is the following:

*The posterior distributions were estimated from a minimum of 2,000 samples per chain, across four chains, following 1,000 adaptation steps.*

**Point 7.** Pg 13 The authors state "Specifically, the CS and UCS components of the state self-compassion were modeled as functions of six predictors: negative affect and the level of unpleasantness of the event." Please clarify what are the six predictors.

We revised this sentence as follows:

*Specifically, the CS and UCS components of state self-compassion were modeled as functions of six predictors: three derived from negative affect and three from the level of unpleasantness of the event. To generate each set of predictors, we centered both negative affect and event unpleasantness to isolate three distinct sources of variance: inter-individual differences, between-day variations within the same individual, and within-day fluctuations within each individual (see SI for details).*

**Point 8.**Readers not familiar with Bayesian analyses might note the 89% credibility intervals for posterior distributions as unusual. Why not 95%? Please provide a more convincing rationale and perhaps also note that it is common practice in Bayesian statistics?

To clarify this point, we added the following specification:

*We generated 89% Credibility Intervals instead of the conventional 95% to avoid encouraging implicit hypothesis testing, as suggested by McElreath (2020). Bayesian inference prioritizes estimation over hypothesis testing, making alternative intervals, such as 89% or even 50%, equally valid and informative (e.g., Burger, Ralph-Nearman, and Levinson 2022).*

**Point 9.** Pg 20. I wonder if the authors could add the phrase "once a week" to reiterate how often participants were assessed over the 3 months?

We added this specification in the Method section of both Study 1 and Study 2, as follows:

*The EMA protocol extended over three months, encompassing 10 specific days within this timeframe, once a week.*

*The EMA protocol spanned three months, with data collection occurring over 16 selected days, once a week.*

**Point 10.** For study 2, the authors note on pg 18 ": Substantial inter-individual variability in the responses to stress, where some individuals show no change in CS while others show no change in UCS, would challenge the notion of a uniform bipolar relationship. This variability might suggest that the CS and UCS components are influenced by distinct psychological or contextual factors that are not accounted for by the BCH". Later in the results section, the authors reported that there was "robust inter-individual variability" for CS (pg 26) and "credible inter-individual differences" for UCS (pg 27). In spite of the inter-interindividual variability, the authors conclude that the "findings provide substantial support for the BCH". Furthermore, on page 30, the authors note that "There were individual differences in how CS influences UCS, with a median estimate of 0.29 (89% CI [0.26, 0.32]). There was moderate unexplained variability in UCS, with a median estimate of 0.39 (89% CI [0.38, 0.41])." These results seem to be inconsistent with the BCH at the individual level. Please clarify these inconsistencies.

Thank you for raising this critical point. In response to your suggestion, we have clarified the level of analysis in the revised manuscript. In the original submission, we used multilevel modeling to capture individual variation through random effects, assuming the same functional form for all subjects. However, while this approach accounts for individual differences, it retains nomothetic characteristics, such as the shrinkage phenomenon, which reduces individual variability to improve the estimation of fixed effects (as noted by Sahadra et al., 2024; Ciarrochi et al., 2024).

Recognizing this limitation, we have addressed it by incorporating a novel idionomic analysis, inspired by Ciarrochi et al. (2024). The results from this analysis show that a non-marginal subgroup of participants exhibits an association between CS and UCS that is inconsistent with the BCH—an important aspect of individual differences that was underrepresented in the original submission due to the focus on nomothetic methods.

In light of the Reviewer's comments and the new analysis, we have revised our interpretation. The revised manuscript now presents both the previous nomothetic findings and the new idionomic results, offering a more nuanced understanding. We have highlighted the inconsistencies between the nomothetic analysis and the BCH, and, together with the idionomic findings, we conclude that our study provides partial, rather than robust, support for the BCH.

In the revised manuscript, we have changed several sections of the manuscript. Specifically, we revised the Title, the Abstract, Discussion of the results of both Studies, and the General Discussion.

For example, the revised Discussion section of Study 1, now is formulated in these terms:

*These findings offer nuanced support for the Bipolar Continuum Hypothesis in the context of daily experiences. The robust opposing effects of negative affect on CS and UCS align with the Bipolar Continuum Hypothesis prediction of inverse relationships between these components. However, the minimal and sometimes parallel effects of event unpleasantness on both components suggest that the bipolar relationship may be more pronounced for internal emotional states than for external contextual factors. This pattern indicates that the dynamic interplay between compassionate and uncompassionate self-responding may be more complex than initially theorized, varying across different types of situational influences.*

The revised Discussion section of Study 2, now is formulated in these terms:

*While most findings support the Bipolar Continuum Hypothesis, some results suggest inconsistencies. Negative affect had a strong, symmetrical effect on both CS and UCS, but decentering had a stronger, asymmetrical impact on UCS than CS. This asymmetry indicates that the relationship between these components may be more complex than the strictly bipolar model of the BCH, highlighting the need for a more nuanced understanding of how contextual factors influence self-compassion.*

*In conclusion, Study 2 provides partial support for the inverse relationship between CS and UCS, particularly in high-stress environments. These results underscore the importance of considering both contextual and individual factors, such as stress and decentering, when examining the dynamics of state self-compassion.*

We also revised the General Discussion section. For example:

*Our findings provide partial support for the Bipolar Continuum Hypothesis, which posits an inverse relationship between CS and UCS.*

Later on:

*In addition to stress, negative affect and decentering emerged as robust influences on state self-compassion. Negative affect had symmetrical but opposite effects on CS and UCS, reinforcing the bipolarity of the construct as proposed by Neff. However, decentering, a measure of mindfulness and emotional regulation, had a stronger impact on reducing UCS than on increasing CS. This imbalance raises questions about whether the Bipolar Continuum Hypothesis fully captures the complexity of self-compassion dynamics. While decentering reduces self-criticism, it may not equally promote self-compassion, challenging the notion of a strictly bipolar model.*

Later on:

*In summary, while our findings support key aspects of the Bipolar Continuum Hypothesis, they also point to complexities that may not be fully explained by a purely bipolar model. Contextual factors like negative affect and mindfulness, particularly decentering, influence CS and UCS in asymmetrical ways, suggesting the need of more flexible models to capture the full range of self-compassion dynamics.*

**Point 11.** While there is indication that the findings are consistent with the BCH, it could be argued that the results presented do not equate to "strong support" as claimed by the authors. One important limitation of the study is that the authors have used a nomothetic approach to test what is essentially an idionomic research question. Using multilevel analyses, the authors showed that state CS and UCS negatively correlate. However, this method of analysis does not examine whether there may be some individuals where these associations (i.e., within-person associations) are not as predicted. As noted in Ferrari et al. (2022; DOI: 10.1007/s12671-022-01897-5) there can still be quite a lot of variability in within-person associations between CS and UCS (suggesting that the BCH is not supported for some individuals) and this needs to be tested using an idionomic approach (see Sahdra et al., 2023; DOI: 10.1007/s12671-023-02187-4). While I understand that reanalyses of this data using an idionomic approach may be beyond the scope of this paper, I hope that the authors can note this in the limitations, propose future research, and temper their conclusions with more caution.

**Reviewer #2:**

I mainly did a high level review of this manuscript for overall theoretical coherence. I'm not familiar with all the statistical methods used so am not qualified to comment on those. Overall, however, I thought it was a novel and well-conceptualized series of studies. There is a need for direct empirical tests of the important issue of whether or not CS and UCS operate independently or as a bipolar continuum, especially as this has important implications for intervention. The data clearly supported the study conclusions - that CS and UCS do operate as a bipolar continuum in the moment. I thought the authors methods were clever and well-executed.

**Point 1.** The only deficit I saw in the study was there was no mention of how study measures such as the State SCS were translated in Italian (assuming they were). This should be clarified.

**Reviewer #3:**

This paper aims to assess the bipolarity of state self-compassion in daily life in several samples. This is a good method to address this issue, which is likely of interest to many readers. However, there are some major concerns with the study conceptualization/design, analyses, and presentation which limit its contribution to the literature, as described below.

1. A primary concern is that I was not clear exactly what the author's a priori criteria were for determining whether results indicated bipolarity or not, and I do not believe the approaches used are comprehensive or optimal on their own. In particular, I was very surprised that the authors did not conduct a multilevel EFA or CFA to examine structurally whether one or two factor are optimal, which seems to be to me one important approach to this question. Rather, the main test seems to be whether there are inverse associations of compassionate vs. uncompassionate responding with relevant variables (e.g., affect), or based upon the sign of fixed effect slope of compassionate and uncompassionate responding. However, such findings in and of themselves do not prove bipolarity, as constructs could have moderate associations in different directions and still be separable/two related dimensions. Overall, I found the design/hypotheses and interpretations to be very over-simplified  
   and over-interpreted, and the paper would need a much more nuanced and comprehensive argument and test for how to distinguish bipolar from two unipolar constructs, with clear a prior guidelines for findings that would support bipolarity or not. This would likely require extensive re-writing, reconceptualizing the study, and changes in corresponding analyses and interpretation.
2. In contrast to point #1, the logic of the moderation tests (under what circumstances or for whom is bipolarity stronger vs weaker) was much clearer and more compelling. Again, however, I don't think this bears strongly upon a nuanced interpretation of bipolarity (or not), as in reality the strength of any association, whether bipolar or not, is likely to vary across some conditions. I find the idea that there must be perfectly consistent associations across context and people for constructs to be bipolar to be a bit of a straw-person argument. In addition, I did not see hypotheses in some cases (e.g., decentering) for the expected pattern of moderation.
3. There was no review in the introduction of other EMA studies of state self-compassion and what is known about it (or its structure) empirically. As such, it was not clear how this study is situated in the larger literature or what its novel contribution is, as well as how to interpret results against relevant prior findings. Similarly, the discussion had little integration of other studies, with few citations throughout and none for the first three pages. This is problematic for fully understanding the results and their interpretive context.
4. I found the manuscript quite difficult to follow, particularly in the results. I think most of this has to do with the structure and narrative that could be improved, as well as excessive detail at times and inadequate detail at other times. Overall, it seemed longer than needed and would be clearer if more concise, with more focused aims and analyses. But I would also suggest the authors reduce the use of acronyms, as these are not necessary and increase the burden on the reader.
5. Several times, experimental design is treated as synonymous with observational EMA studies, which it is not.
6. Why were standard indices of model fit in Bayesian analyses, such as ppp, not considered here? And why was a fairly wide 89% credibility interval selected?
7. Numerous results are reported in the method and need to be moved to the results.