REVISION

We thank the editor and the reviewers for giving us the opportunity of revising and substantially improving our manuscript.

# Reviewer #1

We thank the reviewer for providing these thorough comments that helped us clarify and address several important points in the manuscript and the abstract.

I do think the strenuous efforts in the analyses reduce the concern that findings might be Type I errors. However, these limitations of the study are only addressed at the end of the methods section. I feel that a different title (e.g., "a preliminary investigation into …") and a mention of the small sample could be made in the abstract.

p. 3, lines 21-27: You describe that previous literature has focused on deception detection, but not deceiving ability. I think this is logical, as it may seem more relevant or important to detect lies than to generate them. Could you add an argument why it might (sometimes) be necessary to be able to deceive others?

p. 4, lines 44-48 and 53: is interoceptive awareness one of three facets describing "one's ability to identify and appropriately respond to their perceived interoceptive state" (as argued first) or a "metacognitive dimension of interoception"? These do not sound like the same thing to me.

\*       p. 4, lines 53-64: what kind of association between interoception and rational decisions is posited? And are lies risky? Are lies rational or irrational? Could it be that people high in interoception show more immoral behaviour (as detailed in the paragraph below this one), and then choose a rational path of lying because it less risky? Is this literature really contrasting and mixed?

\*       p. 5, lines 71-72: Could you explain (or cite literature as support) that a short RT indicates a better lying ability? It is somewhat logical for higher confidence and lower physiological arousal to indicate a better lying ability, but could it not be that a longer RT means the lie is better formulated and thus reflects a better lying ability too?   
  
What was the rationale for the direct and indirect phrasing of the questions? There seems to be no hypothesis or reason for this, but then it is considered a factor in (some of?) the analyses. What are the results when this factor is not considered?

Control for phrasing effects and increase incentive to lie

# In the discussion (p. 19, lines 311-313 and further), the validity and reliability of some of the measures is questioned. While I understand that composite scores from the factor reduction are used, it could be good to include information about the self-reported measures (e.g., Lie Scale) in terms of their validity and/or reliability (e.g., Cronbach's alpha). There are some language mistakes throughout, for example: \*       p. 4, line 43: "Garfinkel et al. (2015)'s conceptualizes" ◊ no apostrophe S, and I guess 'conceptualized' or 'conceptualize' (plural) \*       p. 5, lines 70-73: "We expected ToM and interoception … in particular in the Polygraph and the Interrogation condition, respectively." ◊ is it not the other way around, that ToM is linked to the Interrogation condition and interoception to the Polygraph condition? \*       p. 5, line 78: "The final sample consists 26 participants" ◊ consists of \*       p. 9, line 142: "All participants reported believing that it was indeed their own signals being shown" ◊ that their own signals were shown \*       p. 10, line 184: "As all the analyses and data has been made available" ◊ have been \*       Abbreviations are inconsistent (TOM vs ToM; Yoni Task vs YONI task) and sometimes used without the full term written out first, so that the reader does not know what the abbreviation stands for (p. 8, line 139: ECG, EDA; explained on p. 9, lines 152-153). Reviewer #2

We thank the reviewer for the constructive comments as well as the suggestions made to include other works that had prompted us to delve deeper into the extensive literature of deception, interoception and theory of mind, and how their underlying processes may overlap and complement one another.

1.      In their Introduction the authors state that the ability to lie has been always studied in term of lie detection instead of paying attention to the factor that can make people able to lie. This is true, but there are also studies showing that some personality traits are particularly related to deception. For example, people showing traits of the dark tetrad adopt manipulative and deceptive behaviours, especially those showing psychopathic traits. There are quite some studies on this, a very recent one is  Eric Rassin, Carmen Sergiou, Dimitri van der Linden & Josanne van Dongen (2023): Psychopathy as a predisposition to lie hedonistically, Psychology, Crime & Law, DOI:10.1080/1068316X.2023.2213802  
I think that considering that this literature should be at least cited to provide a comprehensive picture of the state of the art. I think it could also be relevant -on a very general note- for discussing the results achieved.  
  
2.      On p. 4, the authors mention the link between interoception and individual differences  "Interoception has increasingly been tied to subjective perceptual experiences (Connell et al., 2018; Seth et al., 2012), as well as individual differences in executive functions, emotional processing, and decision-making (Barrett & Simmons, 2015; Murphy et al., 2019; Petzschner et al., 2021)." This made me think to the link between lying and executive functions and the possibility that because of the link of EF with interoception and lying  (separately) it is reasonable argue a link between interoception and lying.  
  
3.      I found a bit unclear the hypothesis. Did the authors have any hps on possible differences between the two manipulations (polygraph and interview)?  
  
4.      Did the authors carried out an a priori analysis to determine their sample size? Please, report this information and if so, please specify on which parameters they based their power analysis. If not, I think the authors should include a sensitivity/posteriori analysis.  
  
5.      I am a bit confused concerning the final sample size. Was it composed of 26 or 25 participants?  
  
6.      Was the study pre-registered?  
  
7.      I believe more clarification is needed concerning the exact design used in the study. Was it a between design? Which were the exact dependent variables?  
  
8.      Similarly, the tools used need much more elaboration. In particular, the Yoni Task and MAIA-2 need to be better explained (scales, some examples of items, etc).  
  
9.      I was left unclear why the authors included the BES questionnaire. I do not see any hypothesis on empathy and lying, so I wonder why this tool was administered (even if I understand is related to TOM).  
  
10.     More elaboration is needed also regarding the deception instructions provided to participants. Maybe it could be helpful that the authors add them as a Supplementary Material.  
  
11.     Also, I was surprise to see the authors used physiological measures to understand deception but I understand that the authors mainly used physiological measures to assess interoception and also as an additional measure to self-report ones. However, in their discussion they also claim that physiological measures are good lie detection cues. But, there is a large amount of studies showing that actually physiological measures are not good indicators of deception as they are influence by several individual and situational factors. I think this literature needs to be take into consideration at least while discussing the achieved results.  
  
12.     I also wonder whether the authors took into account the possibility that participants while deceiving were actually reporting the truth as I guess the authors did not have a ground truth allowing them to check this. Could this have affected the results? Also, considering the task proposed I have doubt on people's engagement with lying. Considering the questions were neutral ones, I question how motivated people were to deceive (as I understand they also did not receive any compensation for participating in the study).  
  
13.     On p. 8, the authors refer to "behavioural tasks", please clarify in brackets which task are referring to.  
  
14.     Also, the authors recorded people's reaction times, but did they also have a baseline phase during which they recorded individuals' reactivity?  
  
15.     The two conditions created remind me the paradigm of challenging memories with social feedback. Did the authors consider such studies? Maybe it could be helpful also for better justify their results.  
  
16.     I found the discussion of the results a bit hard to be followed. I think the authors should try to clarify how the achieved results fit or are in contrast with previous findings. The did try to but I honestly got lost in the link made and how they explained their results.  
  
17.     I think the authors need to do a small effort in clarifying the practical implications of their work.  
  
18.     Finally, a very general comment. Could it be that the concepts of interoception and TOM work together, that is could it be that they influence the ability to lie simultaneously and not in a separate way as it seems was conceived by the authors?