REVIEWER 2

1. The report includes many different comparisons and statistical tests (noted; with adequate Bonferroni corrections). Based on my reading however, the experimental contrast of interest is between the interactive and non-interactive conditions, separately per grip action. The reported figures could be restructured to more adequately reflect this (i.e. grouping baseline, non-interactive and interactive data next to each other, structured per grip condition), ideally with an indication of level of significance as reported via the planned comparisons.

R1. We thank the Reviewer for this suggestion. To improve the readability of the graphs the results for precision grip and whole-hand grasp conditions are now separately grouped, and the comparison between Interactive and Non-Interactive conditions are easier to appreciate. In addition, based on the comment on point 8, we now divided the figures more thematically, following the in-text descriptions. We believe that this will make it easier to match the graphical results with the corresponding text in which statistical details are given. Please, see the revised Figures 2, 3, 4, 5.

2a. For some statements in the Discussion, it is ambiguous to which specific outcome variables they refer, and thus unclear how they map onto the obtained statistical results, e.g.: "Extending this line of research, here we also observed that in real-time scenarios, reduced MEPs during action preparation were followed by efficient performance during interactive actions execution, in terms of faster and **more efficient** motor responses." > What is meant by more efficient motor responses? The more since: (i) no differences in EMG activity during actual execution of hand actions between interactive and non-interactive conditions were noted; and (ii) the correlation analysis only demonstrates associations between MEPs and RT's, and no other variables of action execution. This notion is also mentioned in the abstract.

R2a. We thank the Reviewer for giving us the opportunity to better clarify these aspects. With ‘efficient motor responses’, we are referring to the results obtained from the kinematic analysis of the performed actions when comparing Interactive and Non-Interactive conditions.
As described in the Discussion, we interpret the efficiency as follows: “the grasping component of the Interactive precision grip movement was characterized by a longer closing of the hand while approaching the spoon, as indicated by the earlier time of maximum grip aperture in the Interactive compared to the Non-Interactive condition [48,69,70]. This indicates that participants chose an anticipation strategy and allocated more time to correctly calibrate fingertip placement on the object and to firmly finalize the grasping movement for the Interactive condition. Notably, this movement was also performed within a shorter time window of grasping time with respect to the Non-Interactive condition. This suggests that the movement was overall quicker and better calibrated, prompting the individual to respond efficiently to the interactive request.”
“No differences were found when performing a whole-hand grasp on the cup full of coffee in the Interactive and Non-Interactive conditions for any kinematic variable. This was likely due to the inherent difficulty of lifting a heavy cup without spilling the coffee. This ecological but challenging condition probably increased the accuracy level, bringing to a general ceiling effect. This is confirmed by the temporal aspect of the kinematic variables: grasping time and time to maximum grip aperture were double when performing a WHG with respect to a PG (see Fig. 5). This finding could be a useful warning for future ecological studies adopting real-life stimuli.”

Based on the Reviewer's observation, we now discuss also that EMG activity did not show to be modulated when the same action was performed under Interactive and Non-Interactive conditions (please see changes at p. 14, lines 607-612):

"Muscle activations also do not appear to differentiate between the same action executed under Interactive and Non-Interactive conditions, suggesting that EMG index may fail to reveal subtle changes in performance compared to the more detailed kinematic analysis of movement. This highlights the importance of acquiring different measures to investigate a phenomenon through different levels of sophistication."

2b. "We excluded this hypothesis by demonstrating that the co-experimenter always performed the Interactive and Non-Interactive action sequences in the same way, regardless of the subsequent response action." > How was this tested, which part of the data is used to back up this claim?

Note that these two examples are non-exhaustive. Since many different outcome measures are employed, please take care to direct the reader to the corresponding results throughout the Discussion.

R2b. We thank the Reviewer for having highlighted this shortcoming in the discussion of the results. Considering the many outcome measures that characterize our methodological approach, we tried to better clarify to which variables and corresponding results we are referring to in order to make the text more intelligible by modifying different parts of the discussion.

For example, at p. 15, lines 629-632 we now report:

"As demonstrated by the velocity results for the co-experimenter's wrist, we ruled out this hypothesis by showing that she always performed the Interactive and Non-Interactive action sequences in the same way, regardless of the subsequent response action."

3. To me, it is still a bit unclear how the participants' prompted responses (stirring the cup via precision grip or lifting the cup with a whole-hand grip) fit within the interactive request. In my experience (and also in the experience of the participants of the preliminary validation study), neither is a full, socially appropriate response to the request, which would involve lifting the cup *towards* the co-experimenter (not merely up), who can then pour the sugar in the participant's cup. In this respect, the precision grip and stirring of the cup may be interpreted as an incongruent social response (i.e. ignoring the interactive request), whereas the whole-hand cup lifting may be considered a (partially) congruent response (i.e. lifting the cup to facilitate the requested sugar pouring) (although again, it can be argued that both may be considered incongruent in a daily social setting). I think this social congruence/incongruence of the prompted action is an important conceptual factor here, which is not reflected in the analysis nor interpretation of the obtained results.

R3. We thank the Reviewer for giving us the opportunity to better clarify this issue. The gesture of lifting the cup was commensurate with the other's action: lifting the cup was in fact the best way to approach the teaspoon, which was almost on top of the cup. We have now made this clearer by specifying that the lifting gesture was used to bring the cup closer to the teaspoon full of sugar (please see changes at p. 4, line 157-158). As regards the different degrees of social appropriateness of the executed action based on the previously observed action, we agree that the lifting action was the more straightforward for the completion of the social interaction and the

achievement of the joint goal. This is also suggested by the responses to the validation questionnaire according to which, for the Interactive conditions, 73% of participants reported that lifting the cup was more socially appropriate, whilst 27% chose stirring the coffee. Nevertheless, we decided to keep different actions evaluated with different degrees of social appropriateness to test whether or not corticospinal responses reflected this explicit evaluation. As shown by the MEP results, different motor preparation emerged for both precision grip and whole-hand grasp performance when the Interactive and Non-interactive conditions were compared. However, only for responses with the whole-hand grip (i.e., cup lifting) did we find a correlation between MEPs and reaction times, with stronger inhibition associated with faster initiation of the action. This result is probably related to the fact that, in general, the lifting action is considered the most appropriate response (see also our response to point 4). We attempted to discuss the differences characterizing the adopted actions within the framework dealing with the completion of a social response in the paragraph “The encoding of incomplete joint actions”.

4. The comment above could maybe also explain why correlations between preparatory MEPs and action-related outcome measures (RTs) were only observed for the whole-hand grip, not the precision grip (i.e., only for the socially congruent response). Perhaps the authors have another explanation for this finding, which I’d be happy to read in the Discussion.

R4. Thanks for your observation, this aspect was indeed poorly discussed in the manuscript. As also reported in our response to point 3 we better addressed this issue in the Discussion (see change at p. 13, lines 561-567):

“When participants were requested to grasp and lift the coffee cup following the co-experimenter’s interactive request action, compared to the non-interactive action, a stronger CSE inhibition was associated with a quicker action initiation (i.e., shorter reaction times), as revealed by the correlation results. However, this effect emerged only for the action of lifting and not for the action of stirring, that is, for the action considered more appropriate in response to the interactive request gesture according to the validation study.”

5. Eye contact is known to be an important factor for conveying interactive requests, and is also known to modulate CSE, as shown in earlier work from the same group of authors. Was this factor taken into account and/or standardized across conditions? Or could the differences encountered in the interactive vs. non-interactive condition be due to differences in eye contact by the co-experimenter?

R5. We agree with the Reviewer that gaze direction is a crucial aspect in this type of experiment; thank you for pointing out this missing information in our stimuli description. In fact, we carefully considered this aspect when we instructed the co-experimenter on how to perform the actions in the different experimental conditions: she was asked to follow the movements of her hand with her eyes and to look toward the participant when her movement ended (i.e., when she reached toward the cup in the Interactive condition or returned to the initial position in the Non-Interactive condition).. We now specify this better it in the Stimuli section of the main text as follows (please see changes at p. 4, lines 162-170):

“Given the important role played by the agent’s gaze, along with limb movements, in providing a source of information for social interaction [18,19,31], we carefully controlled for the co-experimenter’s gaze direction at different stages of the observed action and experimental conditions. Because previous research suggests that gaze toward the observer at the end of the

action (i.e., direct gaze) increases perceived social engagement more than gaze toward the target object for the interaction [18], the co-experimenter was instructed to follow her hand movements in a natural way with gaze and, at the end of the movement, to look at the participant. In this way, the co-experimenter's gaze was comparable in both the Interactive and Non-Interactive conditions."

6. Page 8 final paragraph: when reporting statistics, sometimes the '*p* =' is missing.

R6. Thanks for spotting this, we amended it.

7. To increase readability, preference to write precision grip and whole-hand grip always in full, instead of the PG and WHG abbreviations, also in figures.

R7. Based on this suggestion, we tried to limit the use of acronyms (e.g., PG, WHG) in the most discursive parts of the text to improve readability and comprehension. Figure labels have also been revised in this direction.

8. Figures could be grouped more thematically, i.e. following the in-text descriptions of results distinguishing between outcome measures related to motor preparation vs execution.

R8. We thank the Reviewer for this suggestion. We divided the figures more thematically: now Fig. 2 refers to MEP results, Fig. 3 to reaction times results, Fig. 4 to EMG results and Fig. 5 to kinematics results.

9. Figure legend 3: "RT (A), GT (B) and TMGA (C) for PG (white) and WHG (black) actions..." is a bit unreadable with all abbreviations.
Similar comment about the figure axes in figure 4.

R9. We now reported in full all the abbreviations, thanks for the suggestion.

10. Preference to include the exemplary pictures of the experimental setup, now in the supplementary materials, in the main manuscript.

R10. In Figure 1 we show the experimental setup. In this schematic representation we have included all the necessary and relevant information through a bird's eye view. Such a schematization would be difficult to reproduce through photographs at the apparatus and techniques adopted.