Rebuttal for HRI Submission #135 – “Gestures for Industry: Intuitive Human-Robot Communication from Human Observation” by Gleeson et al.

We thank the reviewers for their time and effort in providing thoughtful feedback. We are happy to hear that all reviewers felt that the paper would be of value to HRI attendees and recognized the importance of interaction design for industrial applications. We are pleased to be able to address the reviewers’ concerns, particularly the key questions on our methodology.

Methodology:

The most significant objection of the reviews relates to the presentation of our methodology, specifically the lack of reported validation metrics for coding reliability. We also recognized this shortcoming and agree that this was a critical oversight. In the time since the original submission, we have conducted the validation suggested by Reviewer 3. A second researcher analyzed and coded a random subset (25%) of the human-interaction videos. In the coding of both the communication terms and the gestures, we found an acceptably high degree of inter-coder reliability (Cohen’s Kappa for communication terms, 0.76; for gestures, 0.71). Very few terms or gestures were coded differently than in our initial analysis. Most disagreements occurred when a coder missed a gesture all together, which would affect reported gesture frequencies but would not change our lexicon. If missed gestures are excluded from the analysis, the agreement between coders is almost absolute: Kappa = 0.97(terms) and 0.96 (gestures). The second researcher also coded the entire set of open-ended responses in the robot gesture identification experiment and showed excellent agreement with our initial analysis (Kappa = 0.91). This validation, following the standard procedure suggested by Reviewer 3, addresses the methodological concerns in the review and confirms that our results can be replicated. We will include a full documentation of this validation in the revision.

The other significant methodological concern regarded the nature of the questions answered by participants in the robot gesture identification experiment. In this case, it appears that the reviewer misunderstood our methods, due to a lack of clarity on our part. Participants gave open-ended responses to the question “What should the human/robot do?”, but were restricted to numerical responses when rating the ease and naturalness of the gestures. Responses to the open-ended question were indeed coded and have been cross-validated as suggested by Reviewer 3 (discussed above), but the responses to the numerical questions required no such coding. We will include this clarification in the revision.

Related Work:

Thank you for referring us to earlier studies of touch-surface interaction. We will cite and discuss the two papers suggested by the reviewers as examples of this body of work. Like both Wobbrock et al. and Micire et al., we analyze intuitive user actions in a simplified task to generate a natural and understandable gesture set. In our revision, we will compare our methods to those used in these papers. We agree that this will help to clarify how our study relates to important prior work.

Detail and Clarification:

Reviewers suggested that robot gesture identification experiments should be conducted with naïve users. We felt that experienced participants were the most appropriate for this experiment because they best represent industrial workers who are expert in their assembly tasks and in communicating about their tasks. The experiment conducted with non-naïve users is more representative of a real-world use case, in which a worker experienced with human assistance would be asked to conduct the same tasks with a robotic assistant. We agree, however, that in future experiments naïve users will provide an informative alternate perspective.

The variance data requested by Reviewer 3 and the gesture frequency data requested by the primary reviewer will be added to Figures 9 and 6, respectively, and reported in the text. Reviewer 3 noted that the lack of error bars on Fig. 9 left some uncertainty about the interpretation of the results. For these data, the standard deviations were acceptably small (min = 0, median = 0.97, max = 1.98 on a 0-7 scale). We will add this information to the Fig. 9 to better illustrate the validity of our results. We will also list those seldom-used gestures that were omitted from the final lexicon, as suggested by the primary reviewer.

We welcome the primary reviewer’s suggestion of a future experiment with participants facing each one another. We chose the side-by-side configuration for our current experiments because it was the most relevant for our industry partner’s industrial assembly environments, based on our observations during site visits, but we agree that a face-to-face configuration could produce more two-handed gestures and is an interesting avenue for future work.