Supplementary file 2

Within-interaction rating framework (W-IRF)

Category	Rating definitions and indicators		
Rating Values	-1	0	1
Coherence Does the turn respond to the topic of the previous turn?	**Definition:** Fragmented from prior turn. **Key indicators:** (1) Ignores previous turn (2) Sudden topic shift (3) Interruptions (4) Refusing to interact	**Definition:** Partial and ambiguous responses. **Key indicators:** (1) Any turn that is not clearly 'consistent' or 'fragmented' (2) Unintelligible speech (3) Yes/no answers to open questions (4) Statements addressed to self (5) Statement addressed to characters in the game	**Definition:** Consistent with prior turn. **Key indicators:** (1) Answering questions comprehensively, yes/no answers to closed questions (2) "OK" to affirm action (3) Continuing topic of discussion (4) Etiquette greetings/goodbye (5) Repetition of phrases or sounds made by partner
Affect What type of emotion is displayed?	**Definition:** Negative affect. **Key indicators:** (1) Extreme negative comments (2) Criticism (3) Frustration (4) Anger / Complaining	**Definition:** Affect ambiguous or not displayed. **Key indicators:** (1) Any turn that is not clearly 'negative or 'positive' affect	**Definition:** Positive affect. **Key indicators:** (1) Laughter (2) Complimenting (3) Encouraging (4) Supporting
Symmetry How assertive/submissive is the turn relative to prior turn?	**Definition:** Submissive. **Key indicators:** (1) Hesitation (2) Turn interrupted (3) Short turns in comparison to partner	**Definition:** Assertive/submissiveness ambiguous or not displayed. **Key indicators:** (1) Any turn that is not clearly 'assertive' or 'submissive'	**Definition:** Assertive. **Key indicators:** (1) Shouting (2) Interrupting (3) Imperatives (4) Very long turns in comparison to partner
Note. Our goal was to understand within-interaction variability, and the constructs of coherence and symmetry examine alignment with respect to the prior turn. However, affect is operationalised in terms of harmony, because using alignment would result in an inflated affect score, because interpreting affect is dependent on observable features which may not always be possible to determine in each turn, thus leading to many 0 scores.			