

**Supplemental Material S3.** Fleiss’ kappa statistics for individual test sessions are provided for the variable-level condition. In addition to the overall values computed by including all trials, three comparisons were computed based on classifying trials based on the signal intensity level relative to the estimated threshold that corresponds to the 71%-correct point (at or above versus below), the type of child behavior (full play-based response versus all other behaviors), and the real-time observer’s confidence level (“confident” versus “not confident”).

Subject Number	Overall	Signal Intensity		Child Behavior		Real-time Observer Confidence	
		<i>At or Above Threshold</i>	<i>Below Threshold</i>	<i>Full Play-based Response</i>	<i>Subtle Behavior</i>	<i>Confident</i>	<i>Not Confident</i>
21	0.608	0.808	0.125	0.896	0.058	0.896	0.058
58	0.566	--	--	0.889	0.154	0.941	0.138
66	0.729	0.789	0.443	0.968	0.35	0.885	0.38
89	0.638	0.863	0.14	1	0.142	1	0.142
90	0.602	--	--	1	0.513	1	0.464
91	0.589	0.59	0.585	0.894	0.206	0.892	0.242
93	0.596	0.774	0.311	0.968	0.235	0.973	0.097
94	0.69	0.841	0.371	1	0.191	0.957	0.243
97	0.56	0.756	0.349	0.965	0.22	0.963	0.253
100	0.806	0.83	0.54	1	0.394	1	0.332
102	0.543	0.711	0.353	0.958	0.191	0.917	0.228
103	0.62	0.794	0.37	0.881	0.34	0.884	0.322
107	0.624	0.875	0.201	0.957	0.133	0.957	0.133