Supplementary Material: Calling for Backup: How Children Navigate Successive Robot Communication Failures

ANONYMOUS AUTHOR(S)

1 Video List

Table 1 displays sources and metadata on the videos shown to the participants.

Table 1. Video dataset. H: Human, R: Robot, the main actor in the video.

Description	Source	Video Type	Failure Time
Human crashes into a Christmas tree on a hover board	OOPs Dataset	Failure (H)	5s
Human crashes into inventory with fork lift	Youtube	Failure (H)	6s
Robotic arm placing pieces on top of one another	Youtube	Failure (R)	8s
Humanoid robot losing balance while standing	Youtube	Failure (R)	17s
Robot climbs down a staircase	Youtube	Control (R)	
Human avoids being drenched in water from a wave	Youtube	Control (H)	

2 Codebook

Table 2 displays the codes used for annotating the successive error dataset. Annotations were created using ELAN [1].

3 Robot Perception Questions

Table 3 presents the results of the Mann-Whitney U tests for each questionnaire dimension. After applying Bonferroni correction for multiple comparisons ($\alpha = 0.01$), none of the five questionnaire items showed statistically significant differences between conditions.

Additionally, Table 4 shows the results of Wilcoxon signed-rank tests examining pre-post changes across all participants regardless of condition. These analyses revealed no significant changes in any perception dimension from pre- to post-interaction.

Table 3. Mann-Whitney U test results comparing perception changes between interruption and control conditions.

Dimension	U Statistic	p-value	p-corrected	Significant
Willingness to interact	398.0	0.095	0.476	No
Competence	313.0	0.522	1.000	No
Trust	301.0	0.317	1.000	No
Social acceptance	340.0	0.958	1.000	No
Likeability	345.0	0.846	1.000	No

4 Error Codes

Table 5 shows behavior codes across all three errors in the performance error stage.

[1] hhttps://archive.mpi.nl/tla/elan

Table 2. Codebook used for annotation of successive error dataset. Colors represent coding categories: reprompting strategies (blue), verbal tone changes (green), emotional displays (red), and disengagement behaviors (purple).

Code	Definition	Examples
repeats prompt	participant repeats the exact same prompt as previously given	"Call the researcher" followed by "Call the researcher" again
more specific / longer prompt	participant's prompt included more words and description than the previous prompt (exclude filler words)	previous prompt: " Call the researcher"; current prompt: "Can you call the researcher?" previous prompt: "Call the researcher"; current prompt: "Can you call the researcher now"
swaps terms in a prompt	participant's prompt has same meaning as the previous prompt but some words are substituted with similar meaning words	previous prompt: "Call the researcher"; current prompt: "Get the researcher"
simpler prompt	participant's prompt included less words and/or syllables than the previous prompt	previous prompt: "Can you call the researcher"; current prompt: "Call researcher"
makes new prompt	participant creates an entirely different prompt with new phrasing or approach	previous prompt: "Call the researcher"; current prompt: "I need help"
'please'	participant adds "please" to their prompt	"Please call the researcher"
slower speech	participant's prompt was spoken slower or choppier than the previous prompt	previous prompt: "Call the researcher"; current prompt: "Call… the… researcher"
demanding tone	participant's prompt was directed at Nodbot and was spoken louder or more forcefully	
interrogative tone	participant's prompt has rising intonation (ends with a higher pitch)	"Call the researcher?"
filler words	participant uses filler words at the beginning of prompt	"Oh… call the researcher" "Um… can you call the researcher"
assertive tone	participant speaks with confidence and authority	
moves closer to robot	participant physically moves closer to Nodbot while speaking	
amusement / humor	smile, chuckle, speaking to Nodbot with humor	
frustration	frown, pursed lips, scrunched face, clenched jaw, utterance, sigh, eye-rolling, annoyed, glaring, looking away quickly	
confusion	awkward smile (corners of lips pulled to side), head tilt, furrowed or raised eyebrow, darting or widened eyes, looking up and rightward, staring at camera, looking around room	
stands up	participant rises from their seat	
leaves room	participant exits the experimental room	
quitting	participant stops interacting with Nodbot	
looks for researcher	participant looks around room searching for the researcher or calls the researcher directly	
no prompt	participant does not give any prompt to the robot during the interaction	

Table 4. Wilcoxon signed-rank test results for pre-post perception changes across all participants (N = 52).

Dimension	Mean Difference	p-value	p-corrected	Significant
Willingness to interact	-0.038	0.564	1.000	No
Competence	+0.038	0.405	1.000	No
Trust	+0.019	0.739	1.000	No
Social acceptance	-0.077	0.405	1.000	No
Likeability	+0.058	0.454	1.000	No

Table 5. Error Annotations by Error Type.

ERROR I (total 150)		ERROR II (total 127)		ERROR III (total 148)	
repeats prompt	24	no prompt	22	frustration	15
looks at robot	22	looks at robot	13	repeats prompt	14
slower speech	13	repeats prompt	13	no prompt	14
more specific/longer prompt	12	looks for researcher	12	looks at robot	13
confusion	11	looks at PC	12	looks at PC	10
looks at PC	11	confusion	8	simpler prompt	10
demanding tone	9	demanding tone	7	confusion	9
looks for researcher	8	slower speech	6	demanding tone	9
no prompt	8	frustration	6	looks for researcher	9
moves closer to robot	7	simpler prompt	6	slower speech	8
filler words	5	more specific/longer prompt	6	more specific/longer prompt	7
interrogative tone	5	quitting	4	moves closer to robot	6
amusement/humor	4	amusement/humor	3	amusement/humor	4
'please'	3	moves closer to robot	3	interrogative tone	4
makes new prompt	2	changes position	2	makes new prompt	3
simpler prompt	2	leaves room	2	filler words	3
changes position	2	filler words	1	quitting	3
stands up	1	swaps terms	1	'please'	3
swaps terms	1			leaves room	2
				changes position	1
				assertive tone	1