

Qualitative Analyses of the Experiment Data

Thematic Coding of the Initial Video-Based Study Free-response Data

We used thematic coding to sort the qualitative data gathered from the initial video-based study's free-response question into three distinct themes: 1) mentioning the cursing as funny/relatable, 2) mentioning the verbal response of the robot as funny/relatable, and 3) mentioning that the robot appeared aggressive/unsavory. A second trained annotator performed the same coding to ensure inter-rater reliability. The resulting Cohen's kappa value was 0.881, indicating a near-perfect agreement between the raters.

In this study, 41 people made a positive comment about the cursing as either being funny or relatable (e.g., "I thought it was funny when [the robot] used profanity because it makes them a little more lifelike," "The swearing seemed more personable than the other conditions when the robot said nothing or used a more tame [...] phrase," "I thought the human-like responses that included profanity were pretty funny, they made me chuckle out loud"). On the other hand, one comment actually favored the no-speech condition as most humorous, musing "I think the ones that didn't have any talking were funnier because it's the purest form of robots messing up."

21 people commented positively that any verbal response at all made the robot feel more relatable than the no-speech condition (e.g., "I was most influenced when the robot had a verbal response. This made the robot seem more lifelike," "I think that when the robots actually said words back it made them feel less scary and more interactive, which I liked better").

10 people commented that the cursing made the robot feel aggressive, unfriendly, or inappropriate (e.g., "If the robot said something with profane language then they seemed more aggressive and not as kind," "When the robot would say curse words, I would feel it was a bit more aggressive").

Six people also mentioned that the videos progressively decreased in the funny and shocking quality after the first one or two videos. For example, one participant mused that "The only thing that really swayed me on any of the videos was the use of vulgar language. And even that got old quickly. The first time I heard it was the most enjoyable because it was a shock." Another noted that "I think that the videos that used profanity were only funny the first time due to how unexpected they were."

Lastly, five participants commented on the movement that robot made after failing at a task. Specifically, the movement of the camera or "head" toward the object stood out (e.g., "When they moved the camera it also made it seem more human-like as they gave 'human reactions,'" "The delayed camera movement added to the more playful and 'lifelike'

response - when something happens, many people take a second to freeze and then look at what just occurred”).

These qualitative results overall suggested that finding the cursing robot humorous or entertaining was much more common than being offended or unsettled by the swearing. It may be possible to uncover key preferences related to robot cursing in early system setup steps, for example via a brief questionnaire to discern whether an individual may be in the minority (within the student sample considered here) who do not prefer robot profanity.

Thematic Coding of the Follow-up Prolific Study Free-Response Data

The qualitative data in the follow-up study had the same natural groupings as in the initial video-based study. A second annotator again coded the data to ensure inter-rater reliability. The resulting Cohen’s kappa value was 0.973, indicating a near-perfect level of agreement.

In the follow-up study, 44 people commented about the cursing as either being funny or relatable (e.g., “The robots who cursed seemed cheeky and funny,” “When the robot swore I found it especially funny. It also made the robot more relatable,” “It was quite human-like and very funny, hearing the robots be emotional and cursing”).

15 people in the follow-up study commented favorably on the general verbal response of the robot (e.g., “If it did speak then it automatically made it feel more warm and gave it some sort of personality,” “I discovered that I enjoyed the robots who spoke after their mishaps. They were more ‘human-like’ than the boring ones who didn’t acknowledge the mistake,” “Videos in which dialogue was added after the failed attempts by the robots garnered the most reaction from me”). On the other hand, one participant particularly favored the no-speech condition’s responses, sharing that “Since I found all of the talk from the robots to be unpleasant (even when it did not swear), I came to appreciate the quiet robot videos.”

15 people commented negatively on the cursing as either unpleasant or aggressive. For example, 10 participants had general negative comments (e.g., “The foul language I think is unnecessary not that foul language offends me but a robot has no use for it,” “I think I tend to be more averse to the robots that cussed in a more vulgar (non-funny) manner because it makes me think that the robots are angry. It did feel more ‘real’ in a sense though”). There were five people in this study who commented that the robot’s cursing seemed aggressive (e.g., “Some of the words seemed to come across more aggressively than others”).

Only 4 mentioned that the videos became less enticing as they progressed, clearly showing some novelty effect (“The videos were repeated and that got boring,” “It was less surprising as the survey continued as I realized the footage, not the audio, was being reused,” “What was funny the first time quickly became rote”).

Lastly, 10 people commented positively on the movement of the robot as a factor that affected how they completed the surveys (“The way the camera looked down at whatever it dropped or ran into was pretty funny to me,” “I also noticed the robot's ability to pan its camera towards the area where it's arm is and take note of what the problem is,” “The way it's head tilt if it dropped an item or hit the table but also reliability”). There were no negative comments about the movement.

Overall, the follow-up Prolific study results showed that as with the initial video-based study, the group of participants who favored the robot profanity was much larger than the group that was offended. At the same time, for those who do not prefer a swearing robot, settings or early customization steps may be needed to ensure a satisfying user experience.

Thematic Coding of In-Person Deployment Interview Data

As in the online deployment data analysis, we grouped quotes from the in-person study's interview into themes. This time, the natural three themes that emerged connected each considered condition to the descriptors of funny, relatable, or cute. A second trained annotator coded the data to ensure inter-rater reliability. The resulting Cohen's kappa value of 0.894 indicated a near-perfect agreement between the raters.

In the expletive condition, four people positively commented that the robot was funny specifically because of the cursing (e.g., “It was kind of amusing that I heard an expletive from a robot,” “It sounded like it swore, and that is funny”). On the other hand, one participant did not like the swearing, even though they thought it was funny. Seven people mentioned that the robot's cursing made it seem more human or relatable (e.g., “I would've said ‘dammit,’” “I think honestly it making a mistake and then dealing with that seemed relatable”). Three participants called the robot ‘cute’ (e.g., “Its cussing, I thought that was cute”).

Four participants in the non-expletive condition mentioned that the robot was funny based on its speech (e.g., “Its reaction to the knock into the table, it was funny,” “The way it said ‘oops’ I thought was hilarious”). Ten people mentioned that the non-expletive speech itself made the robot seem more relatable (e.g., “The fact that it said ‘oops’ gave it a little personal touch,” “It made a human error and then acted human”). Lastly, three participants called the robot ‘cute’ (e.g., “I liked its little ‘oops’. It was very cute”).

For the no-speech condition, only two people called the robot ‘funny’ (e.g., “I thought it was funny that it tried to run through the table to get to me”). Two participants commented negatively that there was nothing relatable about the robot (e.g., “I couldn't hear and it doesn't have a face,” “It was just something mechanical to me”). One person mentioned that the robot was ‘cute,’ stating that “I think it's cute, I don't mind interacting with robots.”

In general, these in-person qualitative results show that both the expletive and non-expletive responses to failures were (almost universally) seen positively, especially compared to the no-speech condition. The biggest counterexample was the participant who disliked the robot's swearing even though they were generally not offended by profanity. It was otherwise difficult to distinguish between the two conditions involving speech.