Turn-taking, Children, and the Unpredictability of Fun

Jill Fain Lehman and Iolanda Leite

Abstract

When the goal is entertainment, designing language-based interactions between characters and small groups of young children is a balancing act. On the one hand, an autonomous character should support the freedom of expression and natural behaviors of children having fun. On the other hand, an autonomous character is only capable of supporting the activity it's designed for and the behavior it anticipates. In the last five years we have watched this tension between freedom and constraint play out in hundreds of small groups in a variety of activities. Using two of the activities as examples, we chart the ups and downs of turn-taking and other language behaviors along the Fun Curve.

Unrestricted, open-ended, speech-based interaction between humans and machines requires vocabulary and semantics as broad as human knowledge, mechanisms for resolving ambiguity and reference to the physical environment, and intricate rules for turn-taking based on a rich model of the situation, social roles, prior context, history, and culture. The practical alternative to such complexity is to design more narrowly focused dialog agents based on task-specific constraints: the vocabulary and semantics of an agent that gives directions (Bohus et al. 2014), the resolution of reference against a small set of physically-available objects (Smith et al. 2015), the turn-taking rules of a tutor and student (Swartout et al. 2013), the conversational simplicity of search (Bellegarda 2014). Task constraints provide the interaction with structure and predictability, creating a kind of pact between human and non-human conversational partners. The more you (human) limit your behavior to what I (agent) expect, and the better I have anticipated what you want to do, the more successful the interaction will be. When the underlying assumptions of commonality of purpose and content break down, the interaction does as well. A great deal of the art of interaction design lies in minimizing what is, from the agent's point of view, out-of-task behavior, both by anticipating natural in-task communication and by providing cues to lead participants down the predicted paths.

Anticipation and cueing are particularly important in designing interactions for young children, a population that is limited in its ability to understand and adapt to the bounds of a system when things go awry. We design for children in the age range of four to ten years old – participants with functional language competence but enormous variability in all aspects of language behavior. Most speech and natural language research that focuses on this population has pedagogy (Ogan et al. 2012; Gordon and Breazeal 2015) or therapy (Vannini et al. 2011) as an overarching goal. In contrast, we are interested in *language-based character interactions* (LBCI) that entertain. Such activities should be novel enough to engage children across the age range, but accessible enough to demand little instruction. They can be brief, but they should be fun. Within these loose criteria, we are free to explore the design space, limited only by imagination, the technologies we can compose or create, and, of course, our ability to predict and channel the language behavior of our young participants. The two case studies presented

here might, at first glance, seem to represent very different points in that design space, but they are highly related with respect to the turn-taking problems and challenges they expose.

Case Study 1: *Robo Fashion World*

Robo Fashion World (RFW) is an animated game in which children dress up an on-screen fashion model by calling out the names of visually-available clothing items and silly accessories. Figure 1 shows a typical moment during play. Edith, an animated robot character, hosts the game; children participate side-by-side in small groups of two to four players, which might or might not include adults. As explained briefly by Edith, there are two main game actions: effecting a change to the model by naming one of the clothing items or accessories on the board, and requesting a picture of the increasingly crazily-clad model to be printed and taken home afterward. The majority of the interaction consists of 20 *choice cycles* during each of which a valid reference to a board item is made, the model changes, and a replacement item appears.

Figure 1: ***A small group playing, and screenshot of, the animated game*** **Robo Fashion World.**

Between 2011 and 2013, the LBCI Group ran 3 data collections that included RFW, with a total of 177 children, 8 adults and 3 experimenters playing across 60 sessions. To collect natural data to test and develop the technologies needed for an autonomous version of the game, a human performed all of Edithʾs language understanding tasks using a Wizard-of-Oz design (Kelley 1984). The interface allowed the wizard to signal a clear request for a board item or picture, an unclear utterance that was nevertheless directed to Edith, a long silence, or multiple people speaking at the same time. The resulting corpus contains 9597 utterances, of which 9039 (94%) were spoken by children. Although there were some systematic differences in the participants across the years (for example, adults joined the children primarily in 2011), the behavioral characteristics discussed here hold throughout. Additional details about the data collections, data labeling, and participantsʾ behavior can be found in (Lehman 2014).

Turn-taking in *Robo Fashion World*

The fundamental rule of turn-taking in adult conversation is ‘‘no gap, no overlap’’ (Saks et al. 1974), an injunction to take oneʾs own turn in a timely manner and avoid speaking over others. Even if such rules were desirable in non-conversational settings like RFW, it would be unwise to expect that young children would have mastered them (Ervin-Tripp 1979). The wizards who

performed all of Edith's language processing defined an implicit turn-taking policy by deciding whether and when to make a selection at the interface. That decision was based on a normative understanding of group dynamics during play and the general instruction that the game should both move along effectively and be fun. For Edith to act as an autonomous host in RFW she must also implement a turn-taking policy, preferably one with the same goals. Doing so requires overcoming the effects of two complex out-of-task phenomena: participants' side talk and overlapping speech.

Whenever there are two or more human participants in a character interaction, there is the possibility of side talk between them and addressee identification becomes part of the turn-taking problem. If the vocabulary in the side talk is distinct from the vocabulary of the gameplay – which Edith must understand in any event – the addressee problem can be solved by assuming that in-task language is always directed to the character and out-of-task language is not. Unfortunately, as Table 1 shows, our corpus does not admit the simple solution; neither the presence nor the absence of task vocabulary accurately predicts addressee.

Addressee	**With Task Words** %corpus (%addressee)	**Without Task Words** %corpus (%addressee)
Edith	53% (78%)	15% (22%)
Another player	7% (22%)	25% (78%)

Table 1: ***Cued/expected task vocabulary does not predict addressee.*** **More than 20% of utterances meant for Edith do not contain expected task words while more than 20% of utterances between players do.**

It is not especially problematic that 22% of the utterances addressed to Edith do not contain anticipated task words: less than a fifth of those utterances (2% of the entire corpus) are actual item requests that use unexpected vocabulary (e.g., "the king hat'' rather than "crown''). The remainder are either evaluative statements ("that's my favorite''), partial phrases without actionable meaning, or unintelligible. On the other hand, the 22% of between-player utterances that do contain task words is not so easy to dismiss. Those utterances are either side conversations about naming ("what's that?'' "a mermaid tail'') or negotiations about the next action ("the fairy wings?'' "no, the bat wings''). If Edith were to assume that she is the addressee based on vocabulary alone, she would take the turn in these instances, despite the child's intent. An inability to distinguish between true item requests and side conversations about items will make Edith seem at best incompetent, at worst malevolent.

Of course, distinguishing addressee does not have to be a function of vocabulary alone. Working with data from an early version of RFW called *Mix-and-Match*, we were able to classify utterances as either *to-character* or *not-to-character* with almost 80% accuracy when we included multimodal features like head turn, pointing gestures, and volume in a time- and group-based Support Vector Machine model (Hajishirzi et al. 2012). *Mix-and-Match* was designed with adult and child players standing side-by-side to force a detectable head turn, in service of the eye contact that typically precedes speech. RFW preserved this design element in part because head turn was an important feature in the addressee identification model. As

adults were phased out of RFW, however, that environmental engineering became less successful because children find the game board to be a strong *situational attractor* (Bakx et al. 2003); unlike adults, they tend to stay visually focused on the board even when talking to other players. Without strong verbal or non-verbal signifiers for addressee, the problem of reliably distinguishing side talk from task talk remains, and Edith's ability to act as the child intends is compromised.

The second challenging language behavior that is ubiquitous across sessions is overlapping speech. This problem shows up in two forms: overlap between a player and Edith, and players overlapping with each other. The children speak over Edith constantly, probably as an inadvertent consequence of the need to make the game easy to understand. Edith's behavior in the choice cycle is purposefully formulaic: she displays idling behavior until the wizard selects an item, then acknowledges the choice to the players ("that's genius!"), turns her back to them as she pushes the red button, names the item and watches the model change, then faces the players to release the turn with or without additional comment and/or gesture. Nothing Edith says during this sequence is critical to the successful execution of the child's request – after the first few cycles, players don't need to hear her to know what is going to happen, so there is no practical consequence to talking over her. Because the wizard is shut out from any additional interface actions until Edith is finished, she appears to simply ignore everything that is said once she has taken the floor.

Ignoring the players' speech is a viable strategy for an autonomous RFW host only if the character takes the floor at appropriate times. Judging the right moment is made more difficult by overlapping speech among the players. Because the wizard was instructed to keep the game moving and fun, Edith rarely expressed that there were "too many voices" and the participants, themselves, were left to self-organize their turn-taking during the requesting portion of the choice cycle. The lack of formal structure affected sessions unevenly. Some groups organized their turn-taking with a simple round-robin and had as little as 10% of their utterances overlap, while others were boisterously chaotic and had more than 80% of their utterances all or partially obscured. Not surprisingly, chaos correlated strongly to group size, but almost all groups had some chaotic moments. Figure 2 shows the variability as experienced by individuals.

As in the case of addressee identification, our goal is not just to describe the language behavior elicited by the wizard's actions but also to replace the wizard's turn-taking behavior with an autonomous capability. To that end we explored the performance of three turn-taking models with respect to a subset of the data (Leite et al. 2013). The *Baseline Model* followed a rule to wait until the end of the first request in the choice cycle, respecting the first speaker's turn boundary but potentially interrupting other players. The *Wizard Model* was a Support Vector Machine that made *take/wait* decisions based on the wizard's actual performance at the interface and the same kind of hand-labeled, multimodal features that were useful for addressee identification. A potential problem with the wizard's data was that it reflected the features that existed at the moment he eventually took the turn, rather than the features that existed at the moment he formed the intent to take it – a variable delay that might be significant in a chaotic environment. In order to control for the variable delay, the *Annotator*

Model was based on data collected from a set of coders given videotape of the game board and players. Each video segment started at the beginning of a choice cycle and stopped at either an end-of-utterance or within-utterance moment, at which point the coders indicated whether or not Edith should take the turn.

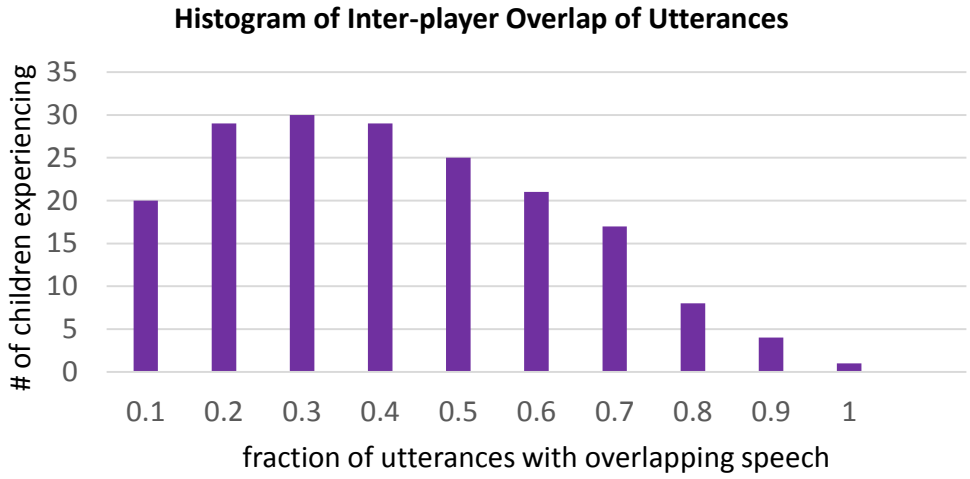

Figure 2: *Quantifying boisterous play*. Almost 60% of the children (105/177) had at least 40% of their utterances overlapped by another speaker. While there are promising new results in sound separation for two voices (Tu et al. 2015), current off-the-shelf speech recognizers do poorly with overlapping speech, even with adult voices. Children's voices are typically harder to recognize, even without overlap.

The three models differed on which turn-taking decision to make about 40% of the time. A new set of coders was asked to judge a sample of 1000 decisions where one of the models differed from the others. In general, the judges preferred decisions by the Wizard and Annotator Models. In particular, they preferred decisions that demonstrated aggressive interrupting by the character, thereby minimizing stretches of inter-player overlap and keeping the game moving with rapid changes to the visual display. In other words, the judges voted to have Edith add her voice to the din.

Knowing to take the turn as soon as a valid request has been made is not the same as being able to recognize when that event has occurred. The general problem of separating multiple excited utterances into speech content that can be recognized accurately is unsolved for adults, and undoubtedly made more complicated by the increased variability of pronunciation in young children. Thus while overlapping speech is not, in and of itself, out-of-task behavior, its significant presence has the same deleterious effect of breaking the pact between the human and non-human conversational partners. You (human) may be producing the language I (agent) expect, but I am unable to produce a meaningful action in return.

Problems in the Large

On the surface, the language understanding task for RFW seems quite tractable: there are only two meaningful actions the character must respond to and the in-task vocabulary for specifying

those actions is fairly small and visually cued. Side talk that contains expected in-task vocabulary complicates the possibility of autonomous performance, however. The better speech processing is at accurately picking out in-task vocabulary, the more likely that the character will act on a misinterpretation of a clarification dialog between players.

The children's natural turn-taking behavior, while easy to anticipate, is still more problematic. When we watch children play RFW, it is clear that most of them are having fun, and it is particularly clear in the groups that are the least orderly in their turn-taking. Even occasional use of Edith's "too many voices'' response had a sobering, albeit temporary, effect on spirited game play. A pilot study that attempted to use a hierarchy of proxemic, gestural, and verbal cues to make turn-taking less chaotic (but still fun) looked promising (Andrist et al. 2013), but in the subsequent, larger data collection, the modified game showed no appreciable effect on overlap.

Obviously there are design decisions we could make – single child interaction, push-to-talk interfaces, etc. – that would constrain away the effects of side talk and overlapping speech completely. Instead, we choose to try to tame them.

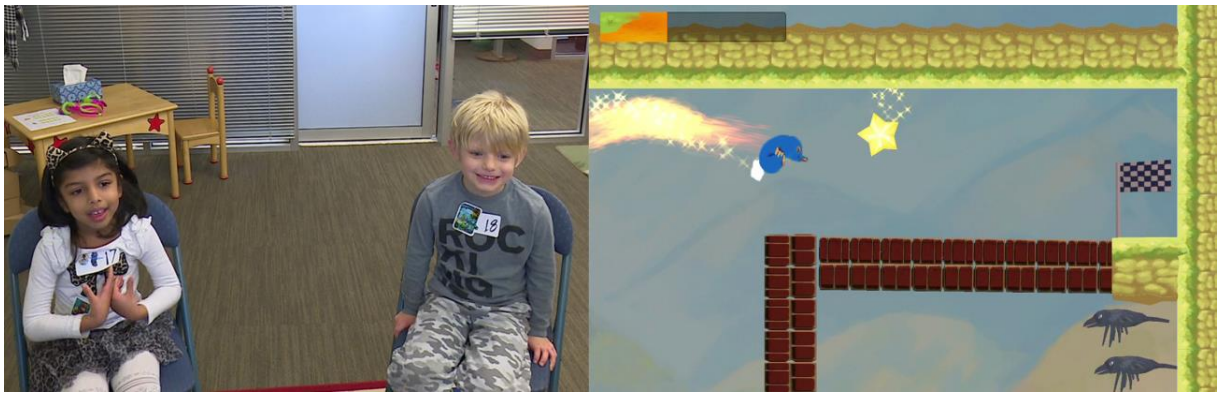

Figure 3: ***Two children playing, and a screenshot from, the animated side-scroller*** **Mole Madness.**

Case Study 2: *Mole Madness*

Mole Madness (MM) is a two-dimensional side-scrolling platform game, a scene from which appears in Figure 3. One player controls the mole's horizontal movement with the keyword *go*, while the other player controls vertical movement with the keyword *jump*. Without speech, the character gradually slows, falls to the ground and spins in place. The mole's environment contains objects that are typical for this type of game: walls arranged as barriers to go over or between, items that increase health (cabbages, carrots, tomatoes) or decrease health (cactuses, birds, bees), and a special object (star) that boosts the character's speed. Although players are not given any specific instruction other than to move the mole through the world to the flag at the end of each level, the health bar in the upper left corner of the screen updates as

the various kinds of objects are touched. Whether through convention or visual affordance, players seem to adopt maximizing speed and/or health as a goal.

We created MM in reaction to the problems in RFW. Where RFW has at least two players, MM has exactly two, the minimum number required to produce speech overlap. Where RFW has 20 or more potentially confusable words and phrases that are meaningful in every choice cycle, MM has exactly two phonemically distinct task words, the maximum necessary to provide each player with uniquely recognizable speech. Where RFW's Edith might add her voice to the acoustic confusion, MM's mole takes its turn through silent action. And where RFW has no disincentive for non-character-directed conversation, MM is fast-paced, with an obvious visual consequence when task talk is supplanted by side talk. In short, an autonomous RFW entails solving hard versions of hard problems, while an autonomous MM entails solving the same problems in their easiest forms – a degenerate point in the same part of the design space for LBCI.

Turn-taking in *Mole Madness*

Like RFW, our understanding of *Mole Madness* has grown through multiple data collections over time. Between 2013 and 2015, the LBCI group had 182 children play MM in pairs under a variety of conditions (most of the children also played MM one-on-one with a robot co-player, but our remarks here focus on the child-child games). In the early pilot games (34 pairs), children used Wii controllers in conjunction with *go* and *jump* to move the character. In the next two data collections, 45 pairs of children used only their voices, with the mole's movement generated by a wizard with a two-button controller who was listening out of view of both the children and the game screen. The most recent 12 pairs of children interacted with the mole directly in an autonomous version of the game. Additional details can be found in (Lehman and Al Moubayed 2015).

Both the issues of overlapping speech and out-of-task behavior should be greatly simplified by MM's design. The mole's world is arranged to elicit specific patterns of speech – if children play strategically, then turn-taking should be almost completely predictable. There are flat stretches to evoke repeated, isolated *gos* by one child, steep walls to produce repeated, isolated *jumps* by the other, and crevasses to get through and items to avoid that require coordinated, overlapping, and orchestrated sequences of the two commands by both voices. Together with the rapid pace, the everyday vocabulary and simple semantics of the keywords should make the game accessible to even the youngest players, without the desire or need for side conversation.

Despite such anticipation and cueing, almost none of the predictability that should have followed from the design decisions outlined above actually occurs during gameplay. Overlapping speech is not limited to areas where it is required to maneuver the character because most children discover, to their great delight, that sequences of overlapping *go*s and *jump*s make the mole fly. As a result, overlap can occur anywhere and does so, almost 40% of the time.

Even the keywords, themselves, defy expectations. All players start the first level with well-articulated, sensible employment of their individual keywords, but as confidence grows, language behavior changes, and children throughout the age range seek to increase the expressivity of the task vocabulary via elision, repetition, and elongation. A clearly pronounced instance of *go* or *jump* takes about 300 milliseconds and has a straightforward cause-and-effect meaning. To get faster movement than full word pronunciation allows, all children spontaneously create *fast speech* forms ("g- g- g- guh go," "jumjumjumjumjump''), crowding multiple commands into the same amount of time. Most children also create *slow speech* forms through elongation ("gooo!" "juuuuuuuuuump") when they want a single, bigger movement, a movement right away, or steady movement at the typical pace. The existence of these different forms, all unquestionably in-task from the child's point of view, adds not only to the complexity of recognizing each command *per se*, but also to the problem of handling speech overlap.

The presence of out-of-task speech is the final complication. The pacing of the game did have an effect – most inter-player speech was both brief and non-conversational ("oh nice start,'' "ah a crow," "yay I got that''). It was also different from side talk in RFW in that the amount was significantly correlated between players, that is, children tended to adopt more or less the same degree of sociability during play. In other respects, however, the phenomenon was quite similar: the overall amount was highly variable across pairs, but almost every pair had some, and about 25% of the utterances contained at least one instance of *go* or *jump*. Side talk with keywords was virtually always about the gameplay itself, encompassing both instructions where misidentification of the addressee would be advantageous ("jump he's falling") and admonitions where misidentification would exacerbate the problem ("no stop saying go").

Solutions in the Small

Despite the reappearance of the very phenomena we wanted to eliminate, MM did prove to have easier, more tractable versions of RFW's problems. As a result, we were able to build an autonomous version of the game by extending classic, example-based keyword spotting to handle overlapping speech and historical context. The implemented system has been trained on almost seven hours of hand-labeled data from the children in the last wizarded data collection. It contains separate models for non-overlapping *go*, non-overlapping *jump*, overlapping keywords, speech in social utterances, and background noise. It calculates whether to send a *go* and/or *jump* command to the mole every 150 milliseconds, based on the pattern of posterior probabilities for the full set of models over the last 450 milliseconds of game play (for details, see (Sundar et al. 2015) and (Lehman et al. 2016a)).

An evaluation with 12 pairs of previously unseen children showed that the system was more responsive and accurate than a human wizard for all of overlapping, non-overlapping, fast, regular and slow speech (Lehman et al. 2016b). Most important, children were judged to have enjoyed the game. We asked three mothers of young children to code the video of each player in both the wizarded training set (31 pairs) and the autonomous test set using a seven point scale with labeled values at 1 (*ready to do something else*), 3 (*could take it or leave it*), 5 (*very much into the game*) and 7 (*can't drag him/her away*) and unlabeled values at 2, 4, and 6 (Al Moubayed and Lehman, 2015). The average mean across all players in the training group was

3.64, 3.88, and 3.67 (coders 1, 2, and 3, respectively), while the average mean across all players in the test group was 4.68, 5.06 and 4.91. In other words, players who interacted with the wizarded character were judged to feel less than halfway between *could take it or leave it* and *very much into the game* (on average), while players who interacted with the autonomous mole were judged to be solidly enjoying the gameplay.

Engagement Isn't Fun

Every interaction is a concrete design problem, an attempt to find enough constraint to make what the human does align with what the technology can handle. When natural behavior is inconsistent with those assumptions, it exposes the hard edges of the design. Children who play *Robo Fashion World* have discussions among themselves about which item to choose next, shout their choices out at the same time, and make up "king hat" rather than saying "crown". Children who play *Mole Madness*, on the other hand, tell each other what to do and not to do before and after doing it, take actions that make no strategic sense, and invent new pronunciations of common, everyday words. On the surface, problem behaviors in the two games appear distinct and specific to their respective interactions, but in reality they differ in degree rather than kind. The most important thing about MM is not that it can be implemented as an autonomous system, but that it demonstrates that certain challenges are likely to arise whenever young children are having fun.

More often than not, young children are accompanied by others – parents, babysitters, siblings, and friends. When a character interaction is in danger of breaking down because what is expected is unclear to the child, someone who is older and more capable is always a potential source of guidance for problem solving. Although getting that guidance almost inevitably results in meta-conversation that includes task vocabulary, being able to get it may be the only way the child can successfully re-engage in the interaction. More importantly, children don't just use others for information – they use them to make an already fun experience more fun. In RFW, children who were shouting their choices over each other were smiling and laughing as they did so. In MM, the children who coders judged to be having the most fun were the children with the largest amount of side talk and the ones who were the most synchronous in their volume, pitch, and use of alternate word forms (Chaspari et al. 2015).

The *Fun Curve* in Figure 4 captures this dilemma: the very behaviors that signal we've achieved our entertainment goal appear to be the most problematic for autonomy. In our part of the design space, children's actions become unpredictable at both ends of the curve – when they are disengaged from the task and when they are so engaged that they, essentially, act like children: creative, boisterous, and unreservedly social. Adults can act this way as well, but adults can also diagnose the effect of their behavior on the quality of the interaction, modify their behavior to bring it back in-task, and find enjoyment despite the self-restraint. Most language-based interaction technologies and agent implementations have been the result of anticipating the natural communication of adults; adult self-correction and adaptation is a source of constraint that they assume. When basic capabilities are designed and combined into agents for children, it is typically done in the context of education or therapy, where

engagement is the focus, the efficacy of fun may be debated, and the idea of ‘‘crazy fun’’ is antithetical to the more fundamental requirement of time-on task.

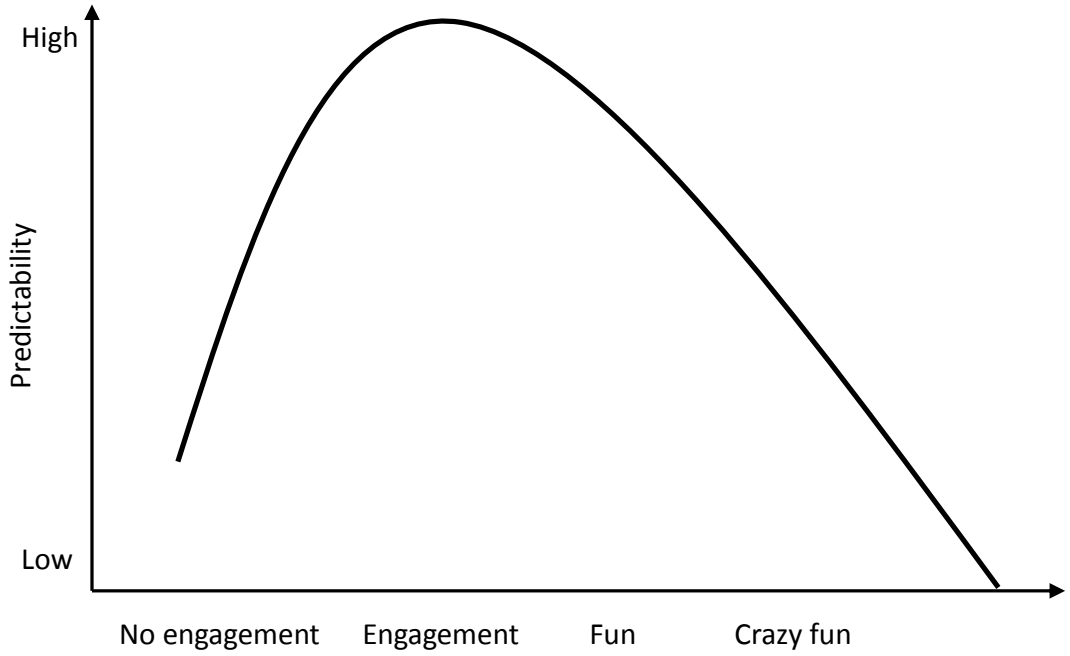

Figure 4: The *Fun Curve*. Unpredictability, from the system's point of view, occurs, by definition, when the child is not engaged in the interaction. It also occurs, at least for systems we can build with current technologies, when the child is having too much fun.

The solution to the dilemma is a science of fun. Our characters should be able to anticipate what form fun will take and recognize when children are having it. They should have weak methods for easing its most extreme expression back toward a state where the main activity can resume. And, eventually, they should have strategies for actively joining in. If a great deal of the art of interaction design lies in minimizing what is, from the character's point of view, out-of-task behavior, then a character that supports fun as its most fundamental in-task behavior will open up a new part of the space of interaction design.

References

Al Moubayed, S., and Lehman, J. F., ‘‘Toward Better Understanding of Engagement in Multiparty Spoken Interaction with Children.’’ 17th ACM International Conference on Multimodal Interaction (ICMI), 2015.

Andrist, S., Leite, I., and Lehman, J. F., ‘‘Fun and Fair: Influencing Turn-taking in a Multi-party Game with a Virtual Agent.’’ Interaction Design and Children (IDC), 2013.

Bakx, I., Turnhout, V. T., and Terken, J., ‘‘Facial orientation during multi-party interaction with information kiosks.’’ Proceedings of INTERACT, IOS Press, 2003.

Bellegarda, J. R., ‘‘Spoken language understanding for natural interaction: The siri experience.’’ Natural Interaction with Robots, Knowbots and Smartphones. Springer New York, 2014.

Bohus, D., Saw, C. W., and Horvitz, E., ‘‘Directions robot: in-the-wild experiences and lessons learned.’’ International conference on Autonomous Agents and Multi-Agent Systems (AAMAS). International Foundation for Autonomous Agents and Multiagent Systems, 2014.

Chaspari, T. and Lehman, J. F., ‘‘Exploring Children’s Verbal and Acoustic Synchrony: Towards Promoting Engagement in Speech-Controlled Robot-Companion Games.’’ 17th ACM International Conference on Multimodal Interaction (ICMI), Workshop on Interpersonal Synchrony, 2015.

Ervin-Tripp, S., ‘‘Children’s Verbal Turn-taking.’’ Developmental pragmatics, 1979, pp. 391-414.

Gordon, G., and Breazeal, C., ‘‘Bayesian Active Learning-Based Robot Tutor for Children's Word-Reading Skills.’’ Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI’15). AAAI Press, 2015, pp. 1343-1349.

Hajishirzi, H., Lehman, J. F., and Hodgins, J. K., ‘‘Using group history to identify character-directed utterances in multi-child interactions.’’ Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, 2012, pp. 207–216.

Kelley, J. F., ‘‘An iterative design methodology for user-friendly natural language office information applications.’’ ACM Transactions on Information Systems (TOIS), 2(1), 1984, pp. 26-41.

Lehman, J. F., ‘‘Robo Fashion World: A Multimodal Corpus of Multi-child Human-Computer Interaction.’’ Proceedings of the 2014 workshop on Understanding and Modeling Multiparty, Multimodal Interactions. ACM, 2014, Istanbul, Turkey.

Lehman, J. F., and Al Moubayed, S., ‘‘Mole Madness – a Multi-Child, Fast-Paced, Speech-Controlled Game.’’ AAAI Symposium on Turn-taking and Coordination in Human-Machine Interaction. Stanford, CA. 2015.

Lehman, J. F., Wolfe, N., Pereira, A., ‘‘G-g-go! Juuump! Online Performance of a Multi-keyword Spotter in a Real-time Game.’’ 5th Workshop on Child Computer Interaction (WOCCI), 2016.

Lehman, J. F., Wolfe, N., Pereira, A., ‘‘Multi-keyword Spotting in a Rapid-paced Game for Child-child and Child-robot Pairs.’’ under review, 2016.

Leite, I., Hajishirzi, H., Andrist, S., and Lehman, J. F., ‘‘Managing Chaos: Models of Turn-taking in Character-multichild Interactions.’’ 15th ACM International Conference on Multimodal Interaction (ICMI), 2013.

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., and Cassell, J., ‘‘Oh dear stacy!: social interaction, elaboration, and learning with teachable agents.’’ Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012.

Sacks, H., Schegloff, E.A., and Jefferson, G., ‘‘A simplest systematics for the organization of turn-taking for conversation.’’ Language, 1974, pp. 696-735.

Smith, J. S., Chao, C., and Thomaz, A. L., ‘‘Real-time changes to social dynamics in human-robot turn-taking.’’ IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2015.

Sundar, H., Lehman, J. F., and Singh, R., ‘‘Keyword spotting in multi-player voice driven games for children.’’ Sixteenth Annual Conference of the International Speech Communication Association. 2015.

Swartout, W., Artstein, R., Forbell, E., Foutz, S., Lane, H.C., Lange, B., Morie, J.F., Rizzo, A.S. and Traum, D., ‘‘Virtual humans for learning.’’ AI Magazine, 34(4), 2013, pp. 13-30.

Tu, Y. H., Du, J., Dai, L. R., and Lee, C. H., ‘‘Speech Separation based on signal-noise-dependent deep neural networks for robust speech recognition,’’ 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, 2015, pp. 61-65.

Vannini, N., Enz, S., Sapouna, M., Wolke, D., Watson, S., Woods, S., Dautenhahn, K., Hall, L., Paiva, A., André, E. and Aylett, R., ‘‘FearNot!’’: a computer-based anti-bullying-programme designed to foster peer intervention.’’ European journal of psychology of education 26, no. 1, 2011, pp. 21-44.