Polly Talks About the Weather: Toward Evaluating the Expressive and Enrichment Potential of a Tablet-Based Speech Board in a Single Goffin's Cockatoo

ANONYMOUS AUTHOR(S)*

Augmentative and alternative communication devices (AACs) are designed to assist humans with complex communication needs. Recently, AAC use has been reported in non-human animals. Such tools may potentially provide enrichment and increase interspecies connection. However, there is no evaluation framework and little data available to assess AAC potential. Here, we examine seven months of a single parrot's sustained use of a tablet-based AAC totalling 129 sessions within 190 days. After devising a coding schema, we propose a framework to explore the expressive potential and enrichment value for the parrot. Our results suggest that the choice of destination words cannot be simply explained based on random selection or icon location alone, and 92% of corroborable selections are validated by behaviors. The parrot interactions also appear significantly skewed toward social and cognitive enrichment. This work is a first step toward assessment of AAC for parrot enrichment.

CCS Concepts: • **Human-centered computing** → *Interaction design*; *Usability testing*; *Touch screens.*

Additional Key Words and Phrases: Animal-Computer Interactions, Touchscreen Interactions; Animal Enrichment

ACM Reference Format:
Anonymous Author(s). 2024. Polly Talks About the Weather: Toward Evaluating the Expressive and Enrichment Potential of a Tablet-Based Speech Board in a Single Goffin's Cockatoo . In . ACM, New York, NY, USA, 25 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Augmentative and Alternative Communication (AAC) encompasses a set of methods, devices, and communication strategies to support people with complex communication needs. Interactive AAC tools provide an important avenue in facilitating communication for individuals with speech and language disabilities. These systems range from low-tech picture boards ([21] to high-tech speech-generating devices ([4]) and have been at the forefront of efforts to empower those with communication impairments ([17])

The HCI community has long shown interest in improving the experience of augmented communicators (ACs) in various contexts including designing systems for greater agency [23], evaluating abilities to use AACs for telehealth [3], supporting communicative functions for children with impairments [12], or assessing the use of AI in reducing the need for typing in adults. [22] The ways to evaluate the potential of such research often involve questionnaires, interviews of ACs, AC parent or support persons, or quantitative video coding. [23]

The field of animal study has used AAC strategies with animals since the 1960s, often in a controversial manner, including the use of sign language with Washoe [10]) and lexigrams with Kanzi the Bonobo [19]

While animal use of AAC devices is not new, recently, there have been reports of the usage of digital AAC strategies and tools with non-human animals in private homes. For instance, humans are teaching their companion dogs to use a hexigon-shaped sound board made of word buttons [25]. A parrot learned to associate images on a touchscreen AAC device corresponding to specific foods, activities, objects, and interactions [5]. Much room exists to advance technological development in this area.

However, the application of interactive digital AACs with animals is lacking consistent evaluation frameworks. Indeed, the assessment of AAC use with animals causes a series of major challenges:

- First, the concept of animal communication with humans, or interspecies communication, is complex and has been debated for decades and no consistent methodology exists for AAC use in this area.
- In addition, intentionality in the use of AACs with animals is difficult to define and demonstrate and is only beginning to be explored [11].
- Very little AAC usage data exists in this emerging area, and no schema exists for how to best collect, code, or analyze data related to AAC use in animals.
- While many studies have been done on animal use of AACs, no framework exists for evaluating the expressive or enrichment value AAC devices could potentially offer in animal use, especially for captive or companion animals.
- AACs are typically designed for humans, which causes complications, gaps, and inconsistencies for use in animals. Much work remains to be done in the area of technological research and design.
- Finally, many tools exist for assessing understanding in AAC use for humans [13], but the assessments are based on standards and benchmarks for levels of human language which are not appropriate for assessing the use in animals as animals are not trying to achieve these benchmarks.

Tackling these challenges would require not only an extensive collection of AAC data, but also a comprehensive schema that takes into consideration not only the user selection but also the broader context of use, and finally a specific angle and set of metrics to make sense of the data.

In this context, the current project is an attempt to explore the use of digital AACs in the animal realm. Here we collected 7 months of use of a commercially-available AAC board by a Goffin's Cockatoo previously trained with associative conditioning who has utilized a speech board for four years. Based on over 39 hours of logged video data, we developed a systematic approach to begin to address the challenges in assessing animal use of AACs and evaluating their expressive and enrichment potential.

One of the largest problems of using AAC devices as an intermediary for interspecies communication is that of validating the animal user's intention in expression. This study first initiates a fundamental exploration into the "expressions" of animal users as a necessary precursor to intention studies by documenting and analyzing the use of a commercially-available AAC board by "Polly," a Goffin's Cockatoo over the course of seven months of logged video data. The cockatoo had been trained with associative conditioning on word correspondences for 11 years and utilized a speech board with over 200 icons for four years.

Importantly, this work does not aim to delve into the nuances of communication. Here we define 'expressive value' as the potential for the parrot's icon presses on the AAC device to be indicative of non-random internal states, preferences, or dispositions. Crucially, we draw a distinction between 'expression' and 'communication'. While 'communication' suggests a purposeful conveyance or exchange of specific meaning or intent between participants, 'expression'—as employed in our study—refers to the possibility that the parrot's interactions with the speech board might reflect a form of internal consistency or preference, the full nature of which remains to be determined.

The term 'expressive' acknowledges the intriguing prospect that these interactions could represent more than arbitrary behavior; they might indeed hint at some level of genuine expression. However, the nature and depth of this expression are not within the purview of the current study to determine. Our immediate goal is to ascertain whether these interactions deviate from random patterns, thereby suggesting that they carry potential expressive value.

If our findings indicate non-randomness, it would set the stage for subsequent research to delve deeper into the complexities of these interactions. Such future studies would grapple with understanding the implications of these patterns: Are they rudimentary forms of communication, mere conditioned responses, or manifestations of a deeper, yet-undefined cognitive process? The prospect of the parrot truly grasping certain aspects of its interactions remains an enthralling, albeit challenging, subject for future exploration.

We remain conservatively poised in our assertions, emphasizing that while our research may pave the way for deeper inquiries, it does not, in itself, confirm comprehension, communication, or a comprehensive understanding by the parrot. It is a foundational step, inviting further academic scrutiny into the nuances of avian interactions with AAC devices.

This paper contributes to the field of human-computer interaction (HCI) by adding to the empirical understanding of animal-computer interaction. We explore an innovative use for training with technology both as enrichment and as a way to enhance the bond between caregiver and animal. This study takes an exploratory approach toward evaluating the enrichment value of AAC devices by looking at data and developing an innovative labeling and coding framework for exploring the expressive and enrichment value of this tool. In addition, this work identifies new design and development opportunities for AAC technology.

To this end, we ask the following research questions:

- RQ1: To what extent can we validate a parrot's use of a speech board?
- RQ2: What theme distribution emerged from the use of the AAC device?
- RQ3: To what extent is the use of a speech board providing enrichment value?

We contribute to the HCI surrounding technology for AAC as an intermediary for humans and companion animals in four ways. (1) We created a labeling and coding schema for the raw video data to capture not only the parrot's interface interaction with the AAC device, but also the social and environmental interactions and cues that may influence or help corroborate the use of the technology. (2) We used this schema to label and code 7 months of data for a total of 39.2 hours of interactions across 129 days of interactions. (3) We examined existing AAC assessment methods and developed a framework adapted to humans and pets to explore the expressive and enrichment value of such a tool this tablet-based speech board. (4) Finally, we used the data to answer the three research questions above: can we work toward evaluating the use of the tool for the parrot, what did the parrot do with the speech board over time, and can we work toward evaluating the enrichment value of the tool for the parrot?

Our results suggest three primary findings: (1) The choice of destination words cannot be explained by random selection alone or solely based on luminescence or icon location ; (2) based on a subset of interactions that offer potential for corroboration, 92% of requests appear to be validated by engagement and time on task; and (3) regarding enrichment, the interactions selected by the parrot were categorized by topic and suggest that the use of the tool leads preliminary to time spent on social enrichment and bonding with the caretaker.

In the broader discourse surrounding AACs in non-human animals, the question of whether animals can truly 'communicate' using these tools often emerges as a central inquiry. It is important to delineate the scope of the present study from this overarching question. While the potential for genuine interspecies communication via AACs is a compelling and much-debated topic, the primary objective of this research is not to validate or refute such claims. Instead, our focus is on developing and applying a robust coding and labeling schema, coupled with an evaluation framework tailored for AAC use in non-human subjects and its application to a dataset on one individual animal. This work aims to shed light on the expressive potential and enrichment value of these digital tools. Our objective is for this methodological contribution to open new doors for nuanced and rigorous assessments of AACs in animal enrichment contexts without being blocked by the broader – and often contentious – debate on animal communication.

2 BACKGROUND

2.1 History of AAC

AAC devices have over 100 years of rich history, beginning in the 1920s with the F. Hall Roe communication board that had words and letters on a masonite tablet. Early developments in the late 19th and early 20th centuries involved the exploration of alternative communication methods, including manual sign language systems and simple picture boards [24]. World War I played a significant role in AAC evolution, leading to the development of mechanical boards to help injured soldiers communicate [20].

In the mid-20th century, electronic AAC devices emerged, incorporating technology switches, voice synthesis, and portable components [24]. Notably, the 1930s saw Dudley's creation of the Vocoder, an early electronic device to aid speech [8]. The late 20th century brought substantial advancements with devices like the DynaVox and the Liberator, which used synthesized speech, including eye-gazing devices - to assist those with communication difficulties [9, 16].

With the rise of computers and software in the late 20th century to now, AAC technology evolved further to include software-based systems and mobile devices with predictive text [14]. Thus, the history of AAC devices reflects a process of continuous technological advancements as well as a deepening understanding of the importance of communication for individuals with speech and language disorders.

2.2 AAC and HCI

AAC, both a practice and a research domain, employs devices to compensate for challenges in spoken and written speech production [1]. Its primary goal is to enhance communication and improve quality of life [13]. HCI methodologies, with their emphasis on user-centric design and iterative interaction, have been instrumental in refining AAC systems, ensuring they are intuitive, user-friendly, and compatible with assistive technologies such as screen readers and switch devices [12].

Innovations from HCI, such as gesture interfaces and eye-tracking technology, have been integrated into AAC devices, allowing users, especially those with motor impairments, to interact more efficiently both with the device [23] and in their relationships [6]. These advancements, coupled with HCI's focus on customization, have made AAC systems more adaptable to individual needs. Furthermore, HCI methodologies offer valuable data collection techniques, shedding light on real-world AAC usage patterns and potential refinements

HCI methodologies have played a pivotal role in shaping AAC systems through user-centered design and rigorous standards for usability testing [**?**]. Based on these methodologies, AAC interfaces are not only more intuitive, but they are also more user-friendly, facilitating accessible communication for individuals with disabilities. Moreover, the incorporation of HCI principles has resulted in AAC systems that are inclusive of a broader range of users, by focusing on compatibility with assistive technologies like screen readers, switch devices, and eye-tracking systems [?].

Additionally, HCI-driven user-centered design has resulted in more customizable AAC systems, ensuring that each user can tailor their communication interface to their own needs. User-centered design engages end users in the iterative design and testing process to ensure their feedback is considered and the systems are tailored to their needs [12].

2.3 Early use of AAC with Animals

Beyond their role in human assistive communication, the use of AAC methods and devices has been explored with animals, most notably with primates. In the late 1800s and early 1900s, attempts to formalize sign language, a manual form of AAC, emerged. American Sign Language (ASL) was one of the signed languages that provided individuals who were Deaf with a comprehensive means of expression. Researchers began teaching primates sign language in the 1960s. One of the most well-known examples is the case of Washoe, a female chimpanzee whose training involved exposure to ASL through interactions with her trainers and other chimpanzees [10]. Over time, she learned to use a variety of signs, which encouraged further research into using AACs with primates.

Lana, another chimpanzee with significant contributions to AAC research, played an important role in shaping subsequent studies involving speech board devices {. Lana learned Yerkish, a communication system based on lexigram symbols representing words or concepts, through a systematic training process. Initially, she was taught to associate specific lexigrams with corresponding food rewards. Lana's lexigram vocabulary expanded to cover a broader range of words and concepts. Over time, she was trained to construct simple sentences and requests by combining multiple lexigrams for more complex interactions.

Kanzi, a Bonobo, was trained on paper and digital AACs with lexigrams for vocabulary words in the 1990s. Under masked conditions, his vocabulary was tested to determine how accurately he could use and recall the words through over 600 execution tasks resulting in a success rate of 71% [?]. More recently, Kanzi correlated pictures with spoken verbal words at a rate of 80%. [19].

2.4 Continued Development of AAC with Animals

AAC training and interactions, especially those mediated by technology, have been hypothesized to potentially offer social and cognitive enrichment to companion animals and their caregivers [25]. Researchers have recently pointed toward a need for captive animals to experience a variety of modalities of sensory enrichment and cognitive challenges. Indeed, a recent study found that parrots who are given training at least once per week have lower incidence of plucking.

other types of tech enrichment include...

Recently pet owners have begun adapting AAC devices for use with companion animals, including paper cards, buttons, and touchscreen devices. Such use has come under intense scrutiny due to the Clever Hans effect, which refers to the 1900 research done with Hans, the horse who was not actually demonstrating the math skills his caregiver thought he had because the caregiver was unintentionally providing reactions that became cues for Hans to select the correct answers (The Clever Hans phenomenon: Communication with horses, whales, apes, and people).

Nevertheless, within and outside of laboratory settings (e.g., Koko, Bunny), skepticism remains due to the Clever Hans effect. Some argue that in each situation of caregivers training animals to use English words, the closeness of the relationship precludes purely analytical results, as the bond itself may interfere with results. [25]. This may underlie both the results, as Clever Hans was deemed to have a knowledge base he did not actually possess when tested without his caregiver present. It may also inflate actual results, assigning meaning that does not exist or assuming greater intelligence than the subject actually possesses.

Notably, some caregiver participation in research particularly on paper symbols and touchscreen devices has yielded interesting results, as a cockatoo using a communication board on a touchscreen device with analyzed behavioral correspondences (Cunha Rhoads, 2020) as well as grapheme-phoneme training with paper card symbols (Cunha et al, 2023), and horses using symbols to request blankets (MEJDELL). Beyond home settings, recently chimpanzees have utilized touchscreen devices to exhibit preference for food (The LANA project) and for interactive tablet games (CITE). This emerging work indicates that there may be promise in assistive technologies for animal communicative interactions.

Overall, while primates have shown some ability to use AAC for communication, there is no agreed-upon methodology for studying their use or collecting and analyzing the resulting data.

2.5 Enrichment Needs of Animals

It is now well established that captive animals need enrichment for mental wellness and to avoid stereotypy behavior. These needs include physical (exercise and diet), mental (cognitive challenge and sensory experiences), and social aspects (socialization with others). Cognitive stimulation is important for mental wellness and often involves activities that build challenges with problem-solving skills and memory (Washburn, 2015). Social interaction has been found to reduce stereotypies by up to 30% in captive animals (Washburn, 2015) and involves interacting with others of the same or different species. Finally, choice and control are primary reinforcers for animals, and the more control an animal has over its environment, the better its behavior outcomes. (CITE)

Human-animal relationship (HAR) plays a meaningful role in the bond between animals and humans and the potential social enrichment deriving from it [25] Through AAC devices, animals may be able to foster interactive exchanges with their caretakers, promoting bonding and reducing social isolation. Moreover, the cognitive challenge associated with learning symbol icons with specific action outputs offer problem-solving and memory mental exercises. AAC devices may be able to empower animals to express their preferences and make choices regarding food, activities, and environmental conditions, thereby instilling a sense of agency and control over their surroundings. If properly validated, AAC devices may also facilitate expressive interactions between caregivers and animals and improve the quality of care. Thus, AAC devices may have the potential to enrich the lives of animals by promoting social interaction, offering cognitive challenge, and enhancing preference expression with caregivers.

2.6 Difficulties in Evaluating Animal AAC Use

When evaluating and validating AAC device use in companion animals, current methodologies are difficult to deploy. Firstly, measures used in laboratory settings are largely inaccessible to lay owners (see, e.g., disembodied voice training with Kanzi and systematic, automated testing devices with chimpanzees). Secondly, beyond limited-scope laboratory studies, no validation measures currently exist for non-human animals using AAC devices.

In humans, most available AAC assessment tools focus on communicative competence, quantities of words in verbal production, assessing progress in human language levels, and measuring the amount of cuing needed by speech-language pathologists to show improvement over time in the person using the AAC, all of which are important when assessing human language abilities. Some assess the tools and features for the AAC apps or devices as part of a needs analysis, which is also beyond the scope of this study. Here we looked at eight tools for assessing AAC use listed on the AAC Community website from the Institute on Disabilities at Temple University to determine if any could inform our framework (https://aaccommunity.net/caac_slp/evaluation-tools/): the AAC Profile measures human linguistic, operational, social, and strategic communicative competence based on Janice Light's skill levels. the Communication Matrix, developed in the 90s by Dr. Charity Rowland of Oregon Health & Science University - this tool has potential to have parts adapted but would require significant work. the Dynamic AAC Goals Grid 2, to measure understanding, expression, social interaction, literacy, also not appropriate for this study, the Functional Communication Profile, which measures sensory/motor skills, attentiveness, receptive and expressive language, pragmatic/social language, speech, voice, oral, fluency, and non-oral communication; the Test of Aided-Communication Symbol Performance (TASP), which is a standardized assessment is $299 and is a paper tool for use in clinical settings to guide in selecting and designing AACs. It can be used to document benchmarks and note progress and looks at understanding of symbol size and number, grammar, categorization, and syntax, so is not relevant to this study. the AAC Evaluation Genie, which can be customized to measure a variety of things PrAACticalAAC Dr. Joy Zabala's SETT framework, which gathers information about the student, environment, tasks, and tools to help match user-needed features with assistive technology. Dr. Zabala's work is designed to improve accessibility for human-technology use - this is not useful for our project because the AAC app has already been selected

While it is helpful to use frameworks like these to consider future goals for animal AAC work and assess design challenges in AACs for animals, that is beyond the scope of this study. We were not trying to assess the parrot's language ability level or communicative competence. We were looking for a framework to assess the AAC's potential for expressive and enrichment value for animals.

While studies exist on using AAC devices with animals, limitations include not exploring animal expressive and enrichment potential with the AAC device. Cunha and Rhoads' study on Ellie included a more comprehensive use of expressive AAC use along with corroborations, but was limited in scope and time, as it lasted only 22 days and across 82 words. Koko was deemed to have learned vocabulary if she used it "spontaneously and appropriately" at least fifteen times during a month. One challenge with this evaluation model is the sheer number of icons an animal would have to select daily to maintain a word deemed part of its vocabulary. With a potential vocabulary of hundreds of words (see, e.g., Alex and Kanzi) or even a thousand, (e.g., Chaser the Boarder Collie), it would be difficult to determine the maintenance and knowledge base of the animal with regard to learned words.

While each propose some sort of validation measure, none provides a comprehensive look at: 1) AAC validation across a large vocabulary of words and different types of queries; 2) overall taxonomy of what vocabulary words an animal uses most often; and 3) beyond the use of words, in what way is AAC and vocabulary learning enriching to animals?

To use AACs for animal enrichment, the animals must be trained on the vocabulary and the functionality of the technology. In this work, we see training as another piece of enrichment. This Cockatoo has been utilizing AAC for over eleven years, first with symbol cards, and then through a commercially-available AAC speech board since 2019. The cockatoo was trained using a rigorous force-free training approach with associative conditioning, positive reinforcement, and the highest ethical standards.

3 TAXONOMY

In this section, we provide definitions for various teems used throughout our methods:

- **Icon:** a word square on the speech board containing most often an image (photograph, sketch or graphic representation) and a written word below

- **Branch:** a menu that opens into another menu with submenus
- **Leaf node:**a final destination word that does not open into another set of menus
- **Menu:**An icon that opens to another menu of icons
- **Main Menu:**The topmost menu on the speech board
- **Submenu:**a menu appearing below the main menu
- **Destination word:** An icon that does not open to another set of menus
- **Interaction Session:**the period of time when the parrot utilized the speech board and accompanying social/cognitive enrichment opportunities
- **Associative Conditioning:**reinforcing desired responses
- **Reinforcer:** something the animal likes that will entice continued behavior for a reward (could be food, praise, social activities, etc.)
- **Positive Reinforcement**
- **Bridge:**typically a sound to identify the exact moment of the desired behavior
- **User:**a non-human user of technology
- **Corroboration:** behavioral correspondences the parrot makes in connection with the icon selection

Here we present data that show the parrot learned to select images that represent (and result in) salient objects, people, and activities.

4 METHODS AND MATERIALS

4.1 Ethcial Statement

All applicable international and national guidelines for the care and use of animals were followed. This project received IACUC approval by XX Institution. All data labeling was base pre-recorded video data, and the animal is privately-owned and resides in the caregiver's home.

To answer the three research questions (RQ1: To what extent can we evaluate one parrot's use of a speech board? RQ2: What theme distribution emerged from the use of the AAC device? RQ3: To what extent is the use of a speech board providing enrichment value?), we examined seven months of existing videos of all speech board interactions with a Goffin's Cockatoo and the caregiver in a private home setting.

4.2 Video Recordings

The data included videos on December 24-26, 2021, March 1-3, and March 25 through September 30, 2022. For weekly session engagement average, we excluded times when the caregiver was away. In sum, a window of 164 days were analyzed and 129 days included interaction sessions, with an average duration of 24.1 minutes. 4,698 words were analyzed in the following ways: Randomness: We analyzed the data for patterns and randomness via XXX

Corroborations: For portions of the data that could be visibly observed as to the behavior outcome of the parrot, corroborations were analyzed with inter-rater reliability. Categories of expression: We categorized the words selected by the parrot into categories and reviewed their prevalence. Enrichment value: We reviewed social and cognitive activity durations resulting from the speech board use in terms of parrot-initiated interaction, frequency of activity, and durations of engagement.

4.3 Participant

Polly [named changed for anonymity] is an 11-year old female Goffin's Cockatoo and was acquired by her caregiver (first author) at 14 weeks. She is housed in a 7.3m x 7.3m indoor aviary with two other cockatoos.

For enrichment, Polly is provided with naturalistic, foraging and parrot-appropriate toys rotated daily for fresh enrichment. She has an outdoor aviary that is 6m x 6m adjacent to her housing. She is also provided with cartoons, music, tablet games, and children books that can be request through the AAC device. Besides her primary caregiver, with additional caregivers who maintain the same schedule, interaction, and enrichment when present.

Polly receives four hours of daily interaction, including voluntary 20-45 min training. The caregiver has been training Polly on word associations and discrimination using force-free positive reinforcement with associative conditioning methods since 2012. In 2019, the caregiver began migrating the words onto an AAC application and subsequently re-training Polly on each association.

Training sessions are initiated either by Polly via a bell ring, by walking up to her tablet, or vocalization, or they are begun by invitation of her caregiver who invites her to the tablet station. During the subject timeframe, a typical session occurs in the following way. First, Polly is invited to press any icons she wishes, and the caregiver responds with verbal acknowledgement, food reinforcers, and also attempts to fulfill any expressions Polly makes. After a period of time, the caregiver invites Polly to focus on cognitive skill development activities. After the cognitive skill building session, the caregiver invites Polly to press any other icons she wants, and then the session ends when Polly is finished engaging.

Each speech board interaction for a period of seven months was video recorded and analyzed, except for 6 occasions when the camera failed. This yielded over 39 hours of training videos, across 129 interaction sessions.

The commercially available AAC application allows for significant customization and was populated by the caregiver over four years. Because the icons and options were chosen by the caregiver, there could be human bias in location and options, which we will consider in the anaylis.

4.4 AAC System

4.5 Setup

Add lots of images from setup and Commboard and menus

The parrot's training took place in one of eight possible locations in the indoor aviary.) In each location, the tablet was either placed on a stand in a learning area or held in front of the caregiver so the caregiver could not see the parrot when she was touching the screen selections. **(See figure XX.)**

The tablet was a Samsung Galaxy Tab A, sized 208.4x137.90x7.50, protected by a safety case. The AAC application was the commercially available CommBoards AAC Speech Assist for Android [ref Shmoontz Apps]. Images on the speech board were 38.1 mm square and came from either the in-app picture library or royalty-free images, or were taken by the caregiver.

The app interface was custom-populated for the parrot. Upon opening the app, the main menu comprises 12 menus (branches) that open into sub-menus (leaves) sometimes four levels deep (ex. menu-branch "To Play"" opens, leading to a variety of activity options, one of which is menu-branch "Book" that also opens into a variety of options-leaves including touch and feel, pop-up, reading book, etc.), and destination words. A destination word does not open into any other option. Each sub-menu contains a back button and between one and eleven additional options. Each option has an associated icon, which represents a person, place, object, feeling/state, navigation option, or activity. During this study, the speech board available to the parrot contained a total of 28 possible menu-branches leading to a total of 148 possible single options-leaves and 28 back-to-home icons. Additionally, each sub-menu contained a blank box below the menu leading to it, to account for the cockatoo's double-taps and avoid false-positive touches. There were 16 such blanks. Thus there were a combined total of 220 icons. **See Appendix [X]** for a complete list of menus, sub-menus, and destination words.

The tablet capacitive touch screen can be engaged by an electrical conductor like a finger or the parrot's tongue. When an icon is engaged, a synthesised voice pronounces the name of the selection and opens the submenu, except in the case of destination words which do not open additional menus. The caregiver gave the parrot varied low-to-high-value reinforcers including half or whole nuts.

4.6 Training Protocol

The parrot was initially trained with positive reinforcement and applied behavior analysis to touch "Treat." When she touched the icon, the caregiver said, "Good" as a verbal bridge and gave her a food reinforcer. After three to four successful sequential touches, only voice-engaging touches were reinforced. After four voice-engaging "Treat" touches, the parrot was trained to touch "Nut" with the same procedure and given a nut. When success was again achieved, the caregiver taught a preference discrimination task by asking, "Which one do you want?" thereby training Polly that she would receive the food item she reached for.

At this phase of training, the caregiver held the tablet parallel to her eyes so she could see only the back of the tablet but not the parrot selections. Although spontaneous eye gaze in lab settings has been inconsistent, birds have been trained to follow eye gaze ([**?**]). Nevertheless, to prevent inadvertent influence on the parrot's selection through visual cues, the selections were generally hidden from the caregiver, unless Polly struggled with capacitive touch.

Discrimination training was deepened by holding a reinforcer and asking, "Which one is this?" and reinforcing the food item corresponding to her selection on the board (for example, if the reinforcer was a pine nut and she selected "pine nut," she was given the pine nut). In addition, the parrot had previously been trained in yes-no choice and preference discrimination using symbols to indicate agreement or not ("Yes" or "No") and those were also on the speech board. The full vocabulary list is in Appendix [X]. Sub-menu items can be seen in Figures X-X.

Using this associative conditioning method, the parrot was similarly taught to discriminate icons and associate them with places, objects, feeling labels, activities, or people. The speech board contained three main categories of menus: food and beverage ("To Eat," "Treats," and "To Drink"), social-cognitive enrichment activities (e.g., menus "To Say Hello To," "To Play," "Games," etc.), and Ambiance, Locations, and Objects ("Temp & Weather," "Music," and "Cartoons"). A full list of the categories and accompanying menus is seen in Table X.

4.7 Frustration Avoidance

Given the demands of beak/tongue manipulation, motor skill development, and icon association training, sessions incorporated protocols to maintain momentum with a high rate of reinforcement. If the parrot attempted to touch but did not trigger the speaking voice, on a 3rd attempt the caregiver held the back of her hand 15 cm in front of the parrot and prompted "Touch," which was reinforced with a treat. In this way, at no time did the parrot attempt three target behaviors without a food reinforcer to lessen the risk of frustration. They then returned to the associative conditioning task, repeating this pattern as needed. Once the behaviors were consistent, an incorrect answer was treated with Least Reinforcing Scenario/Stimulus, which involved a 2-second neutral pause and withdrawal of the tablet and a subsequent prompt to recue the target behavior (ScarpuPollyi, et al. 1991: Decreasing the frequency of behavior through extinction: An application or the training of marine mammals).

4.8 Corroboration

In the majority of instances of Polly's AAC use, the parrot was provided with the opportunity to engage with the selections. The caregiver observed that occasionally when Polly's AAC selection was offered as fulfillment by the caregiver, Polly interacted for a few seconds, such as touching the screen of a tablet game quickly or tentatively tasting the food item , but then failed to follow through with the selection in a way that correlated with the selection (i.e., she walked away from the tablet or dropped the food item immediately after tasting it).

Thus, for our framework, corroboration was measured positively when logging the video data by noting whether the parrot engaged with the selected activity (e.g, tablet game or book) for a period of 1 minute or longer, or whether consumed at least three bites/sips of the food or beverage selected, or in the case of a smaller food item, consumed it entirely.

4.9 Data Collection

The caregiver stored pre-existing data in the form of recorded videos of every speech board interaction with the parrot. Although Polly had used the AAC device for four years, not all of the sessions were recorded. Thus, data was logged from seven months of videos. All video data was included in the analysis, totalling 129 days and 192 videos and communication interactions. The next section will explain the logging, labeling, and coding schema the research team developed.

4.10 Randomness Testing

One goal of the present work is evaluating *expresiveness*: to determine whether the parrot's choices in the AAC interface can be explained by simple heuristics corresponding to random selection strategies. We seek to determine if the parrot's interface selections correspond to properties of the interface structure, or if the observed behavior indicates additional factors beyond this driving the parrot's selections. Responses to exogenous rewards, specific intentions, and a richer cognitive understanding of the interface's semantics are some factors that could cause significant deviations from interface-structure-dependent selections.

The interface presents an extensive set of choices organized in a tree-like structure of sub-menus. If the parrot is simply responding to the visual layout and availability of options, we would expect the behavior to follow certain patterns determined by the interface design. For example, leaf node (i.e. destination) icons that are easier to gain access to in interface traversals (such as by being closer to the home screen) could be visited significantly more often. Our objective is to quantify how much the observed behavior deviates from this type of random behavior within the interface structure.

4.10.1 Modeling Framework. We implemented two primary regression analyses:

***Model 1**: Predicting the parrot's behavior using observed data variables.* We fit a Gamma family generalized linear mixed-effects model (GLMM) with a log link function to predict the parrot's average icon press rate per session. This is computed as $\frac{N_{presses}}{N_{sessions}}$ for each leaf node icon. Since this quantity is in principle positive and not clearly bounded, yet in practice right-skewed towards zero, the Gamma family with log link is a natural choice. Though we did not observe any values close to or above 1, we selected the Gamma distribution to account for this possibility (if a button is pressed more than once per session in an alternative dataset).

The model included fixed effects for the icon's X and Y position trated as factors ($X \in \{-3, -2, -1, 1, 2, 3\}$ from left to right with center as reference, and $Y \in \{-1, 1\}$ from bottom to top), depth of its location in the interface tree (an integer), average pixel brightness of the icon (Luminance) (normalized $\in [0, 1]$), number of choices available on the screen in its context (an integer), and interaction terms for this last component with the X and Y positions. Different sub-menus contain different numbers of items, so adding this interaction term accounts for the possibility of this creating a bias in the position of the selected icon. We accounted for non-independence in the data by including a random intercept for each interface sub-menu. If interface structure has any explanatory power over the parrot's selections, we would expect position and choice count to be significant predictors in the model.

***Model 2**: Comparing observed data with simulations of random behavior.* We generated simulated data sets by programatically traversing the interface randomly, as described in the following section. This produced null distributions of behavior under basic heuristics of random choice within the interface constraints. We compared three variants: pure random traversal, adding ability to repeatedly press leaf node icons, and adding ability to abandon sessions at any point.

We then fit another similar Gamma family GLMM with a log link predicting icon press rate, including a fixed effect of Dataset (Observed, plus three heuristically simulated datasets we discuss in the next section: Random, +Leafs, and +Abandon). We also included interactions of Dataset and spatial position (X and Y) to account for differing spatial biases across simulation methods and observed data. We also dropped the luminance (average brightness) variable both based on model 1 results, and also as simulations do not have any way of accessing this information (no perceptual module which could plausibly be driving choices).

4.10.2 Simulated Dataset. To simulate interactions with the interface, we transformed the interface structure into a tree representation. In this setup, the interface's menu states $s_i \in S$ are nodes and icons are actions $a_i \in A_{s_i}$ leading to transitions. Random behavior through the interface can be modeled as a Markov process. This consists of a set of states S corresponding to interface screens, a set of actions A_s available in each state, and deterministic transition probabilities $T(s'|s, a)$ that depend on the selected action. Specifically, each action $a \in A_s$ either transitions to a new state $s' \in S$ or terminates the interaction, except for a special "back" action that transitions to the previous state.

To simulate random behavior through this model, we define a uniform random policy $\pi(a|s) = \frac{1}{|A_s|}$ that chooses actions independently and uniformly at each state. We generate multiple simulated trajectories by sequentially sampling actions from this policy and transitioning between states.

Comparing statistics of these simulated trajectories against the observed traversal data allows us to assess whether the parrot's behavior displays additional structure beyond random exploration of the interface.

To generate null distributions to test the observed behavior against, we simulate three variants:

- Random Traversal: In each state s, select a random action a uniformly from the available actions in that state. A session terminates when a leaf node icon is selected: $a_t \sim \text{Uniform}(A_{s_t})$ where s_t is the interface state at time-step t.
- +Leafs: As above, but selecting a leaf node icon does not necessarily terminate a session. Instead, either another leaf node icon or session termination is selected with uniform probability:

$$P(a_t|s_t) = \frac{1}{|A_{s_t}| + 1} \forall a_t \in A_{s_t};$$
$$P(term|s_t) = \frac{1}{|A_{s_t}| + 1}$$

- +Abandon: In addition to the +Leafs version, this version allows the agent to abandon the interaction at any point in the traversal, not necessarily after a leaf node selection. At each state s, the agent abandons the session with probability $\frac{1}{|A_{s_t}|+1}$, and takes further actions (either submenu navigation or a leaf node icon selection) with equal probability.

Producing Markovian simulations over this state space allows us to generate null distributions of behavior under basic heuristics. Comparing to the observed behavior then reveals if additional factors beyond interface structure are needed to explain the observed selections.

To compare with the observed behavior, we also transform the real data into this format. We iterate through icon presses and split them into sessions at points marked as such. We add further processing to reduce any potential biases for this portion of the analysis: when a caregiver action modifies the interface state, we discard all state-action pairs from this point onward until the next session. At the conclusion of this processing, we were left with 296 sessions of average length 7.21±6.38.

5 CODING AND LABELING SCHEMA

To encapsulate the holistic context in which the parrot interacts with the screen, we devised a schema to encode the video data. We identified three themes of coding:

- **icon presses**: every individual selection made on the board including menus, leaves, back button, and exit.
- **Environmental information**: time of day, location, and duration of session
- **Parrot and caregiver interactions:**including initiation and comprehensive details of the interaction such as speech board navigation, menu and word icons, caregiver responses, two-choice preference selection, behavior correspondences, as well as other AAC use such as symbol objects, parrot gestures, and word cards
- **Categories of icon presses:** Social Interaction, Self-Reporting, Activity, Food/Beverage/Treat, Request for Learning Theme, Ambiance - Music/Cartoon, Temperature, Menus Learning, Location, Environmental Information, and Request for Object
-

Table X of the Appendix summarizes all the logged entries and notes the entries that were used for this analysis.

5.1 Logging methods

Rather than a transcription of the videos into text, to encapsulate the full experience, the team opted for documenting as much information as possible from each video. Two loggers were trained on consistent documentation practices for logging the data in a spreadsheet with links to each video. Sessions could occur multiple times per day and were divided in the spreadsheet with a solid black line between them. Multiple interactions could occur in each session and were separate videos and indicated with a gray line between them.

The two loggers independently watched each video and logged information from the sessions including details about the caregiver, the parrot, the speech board, and the resulting activities including the following: Date and Time of Day of the interaction Duration of recorded interaction Interaction initiation (who - parrot or caregiver and how - bell, etc.) Which caregiver was present and the location of the caregiver in the room Interaction location in the room Caregiver prompts, questions, responses during the interaction Menus and submenus pressed by the parrot Words pressed in the submenus by the parrot, including multi-words, which occurred when the parrot would tap multiple words rapidly Parrot corroboration: Two-choice options, yes-no options, and the parrot's selection (of relevant) Duration of engagement in the activity (of relevant/or for which category) Interaction closing (who ended the activity - the parrot or the caregiver - and how)

To log the parrot presses on the speech board, “Menu” refers to a press that results in navigation to a speech board menu. “Word” presses were logged only if the press resulted in audio voicing from the speech board application. Multi-word presses resulted in incomplete audio voicing due to interruption by subsequent presses.

5.2 Inter-rater Reliability

The loggers both completed 30% of the video logging independently, and interrater reliability was scored at 90% before completing the remaining logging and spot-checking IRR after every two hours of logging over a two week period to ensure high rates of consistency and IRR. Logger One logged March-July videos and Logger Two logged May-September videos. IRR scores were analyzed on the overlapping 30% of th May-July videos. Any variation in menu or word logging was noted. IRR was 95%. Three researchers analyzed corroborations on 50% of the data for IRR and then separately logged the additional 50% of the data.

5.3 Emergent themes

Once we had this large number of AAC-mediated interactions logged, to answer RQ 2: “What did this parrot do with the speech board? What themes emerged?,” the data was organized into themes for the interactions so as to identify the areas most often selected by the parrot . We created themes as this was difficult for two reasons: 1) there is not a widely accepted consistent list of themes for AACs, and 2) AACs are highly personalized. This particular speech board app came with some predetermined categories (e.g. “Food”) but has been customized over the four years it has been in use to include the words most salient to this parrot. Generally, the words on this speech board are treated as either requests (for someone or something) or reports (about self or environment). The researchers discussed this extensively and ultimately included the following overarching themes to organize the available selections:

- Activities: request for games, books, crafts, or other activities
- Ambiance: request for specific music or cartoons to be on in the room
- Environmental information: report about environment temperature in the room or outside
- Food/Beverage/Treat: request for food items (e.g. toast, eggs), beverages (e.g. tea, juice) or treats (e.g. seedball, pistachio)
- Location: request to move to a different location
- Learning theme: request for a “theme” or topic (e.g. animals, fantasy)
- Request to learn something new:
- Request for object
- Self-Reporting Feelings: report about emotional states and feelings (e.g. happy, mad)
- Self-Reporting Temperature:
- Social interactions: request for
-

These categories begin to address RQ2; however, because of the complexities in some of these categories, they will not all be discussed in detail in this paper but will be examined more thoroughly in future work. In addition to categorizing the data into these themes, we needed an approach to label and code each element of each interaction.

5.4 Expressive Evaluation Framework

The question of the expressive value for animal speech board use is complex. In this work we do not aim to prove that the parrot selections are what the animal truly intend to express.

However we wish to work toward identifying possible.. as a first step toward assessing

5.5 Categorization of Selections

The researchers independently analyzed the data and noted categories comprising Polly's selections. The researchers then came together and agreed upon a list of categories that reflected the types of selections Polly made.

The categories were: Activity Ambiance (music/cartoon) Environmental Information Food/Beverage/Treat Location Request for Learning Theme Request for Object Self Reporting -Feeling Label Self Reporting - Temperature Social Interaction Temp - Menus and Environment

These categories were then automated, except for "Talk About" and "Yes and No" which were context dependent and then manually entered. As part of the evaluative framework, we categorized each word into this list to analyze the themes she exhibited in her communication in a quantifiable manner, and compared against the actual distribution of the categories on the AAC device (i.e., % of food selection vs. total % of food selection on the device). A table of the words comprising the categories is located in Appendix X.

We then narrowed this list to three primary categories: Food/Beverage - nutrition related expressions Social-Cognitive Interaction - social and learning interaction expressions Request for Object, Location, or Ambiance - expressions relating to her environment

A table of the words comprising the categories is located in Appendix X.

5.6 Enrichment Evaluation Framework

To assess the enrichment value of the interaction sessions independently from expressive value, we analyzed each of the activity selections Polly engaged in over the course of the seven months.

The activities included: Tracing and Writing Tablet Phonics and Typing Letters Tablet Games Book (Picture) Outside Walk Reading Video Calls to Family and Friends Car Rides Crafts Visit With Another Bird

We analyzed the type of activity, the frequency of engagement, the overall quantified duration of engagement, as well as average duration of the activity.

6 RESULTS & ANALYSIS

After developing a framework and logging and coding schema, the researchers were able to use that framework to examine what the parrot did with the speech board and what themes emerged in the logged data. We note here, the distribution of words selected, the categories incorporating those words, the corroboration rates for the interactions, the randomness of the selections as compared to computational models, and the potential for enrichment and expressive value. The parrot's average speech board interaction rate over the 129 days was 5.3 days per week with an average of 18 minutes each day (morning, evening, or both).

6.1 Randomness Testing

6.1.1 Insights from Observed Data (Model 1).

Table 1. Analysis of Deviance results for Model 1 predicting button press rates, using type II Wald chi-square tests. : indicates interaction between two predictors. * indicates p<0.05, ** shows p<0.01, and *** shows p<0.001.

Predictor	χ^2	Degrees of Freedom	Pr(>χ^2)
X	24.05	5	0.0002 ***
Y	26.82	1	0.0000 ***
Depth	1.07	1	0.3019
Average Nrightness	0.84	1	0.3606
Total Number of Buttons	5.20	1	0.0226 *
X:Total Number of Buttons	4.38	5	0.4965
Y:Total Number of Buttons	0.56	1	0.4523

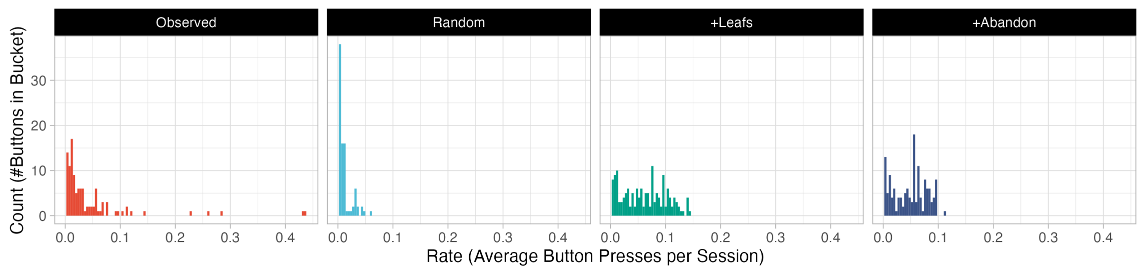

Fig. 1. Histograms of observed vs. simulated button press rate distributions. All three simulated datasets partially reproduce different attributes of the observed distribution, such as a bias towards low rates (all), the bounds of the bulk of the distribution (+Leafs, +Abandon), and the skew towards near-zero rates (Random). However, the observed distribution still differs from these in the specific skew, and the long tail with outlier buttons.

Interface Structure Provides Limited Explanatory Power. The marginal R^2 for Model 1 was 0.282, and the conditional R^2 was 0.661, based on calculations using the Nakagawa method. The marginal R^2 represents variance explained by the fixed effects alone, while the conditional R^2 includes both fixed and random effects. The marginal R^2 indicates substantial unexplained variability in press rates. The conditional R^2 is largely driven higher by including sub-menu random intercepts. The large gap between marginal and conditional R^2 reveals that while accounting for submenu dependencies improves model fit, much variation remains unexplained by the predictors. Sub-menu effects may also reflect differences in the content of different sub-menus, rather than solely dependencies in the interface. In summary, the Model 1 fixed effects provide limited explanation of variability in the observed press rates, indicating that unobserved factors may drive button selection.

Assessing Individual Predictor Contributions. We can further analyze results from this model by dropping each predictor one at a time and recalculating the model's goodness of fit. The resulting change in deviance indicates the variable's contribution to explaining variance in the response. We performed this analysis of deviance using type II Wald chi-square tests, and the results are shown in Table ??.

The results from this analysis show that the largest deviances come from the X position ($\chi^2 = 24.1$, $P(>\chi^2) < .001$) and the Y position ($\chi^2 = 26.8$, $P(>\chi^2) < .0001$), and also show a smaller significant effect of the total number of buttons in a given leaf node button's sub-menu ($\chi^2 = 5.2$, $P(>\chi^2) < .05$).

6.1.2 *Comparing Observed Data with Simulated Behavior (Model 2).*

Simulated vs. Observed Press Rate Distributions. The observed press rate distribution displays a skewed, power law-like shape from near zero to 0.1, with outlier rates exceeding 0.4. The simulations do not fully reproduce these features, though each one mimics particular features of the observed distribution:

- The random traversal simulation mimics but exacerbates the steeper decrease and higher concentration at near-zero rates. As such, it also does not extend as high as the observed distribution.
- The +Leafs distribution produces a similar extent and shows some decrease from the very low rates to slightly higher rates. However, it has a flatter, less skewed shape than observed.
- The +Abandon distribution spans a similar main range, up to 0.1, but lacks the skew.
- Additionally, all of these simulations fail to show any outlier button presses, suggesting that random interface traversals do not emphasize any specific buttons to the degree observed in the data.

In summary, no simulation accurately captures the specific skew shape, heavy tail, and presence of outliers seen in the observed data. Certain buttons are pressed far more frequently than expected under random traversal constrained by interface structure.

Table 2. Analysis of deviance table for Model 2 predicting button press rates with simulations, using type II Wald chi-square tests. : indicates interaction between two predictors. * indicates p<0.05, ** shows p<0.01, and *** shows p<0.001.

Predictor	χ^2	Degrees of Freedom	Pr(> χ^2)
Dataset	274.98	3	0.0000 ***
X	17.98	5	0.0030 **
Y	16.46	1	0.0000 ***
Depth	10.12	1	0.0015 *
Total Number of Buttons	0.24	1	0.6247
Dataset:X	27.78	15	0.0230 *
Dataset:Y	10.74	3	0.0132 *
X:Total Number of Buttons	3.22	5	0.6664
Y:Total Number of Buttons	0.21	1	0.6504

Simulations Remain Distinguishable from Observed Data. The inclusion of simulated data along with observed data in Model 2 resulted in a marginal R^2 of 0.375. Though higher than Model 1, over 60% of variance remains unexplained by the predictors including dataset type. An Analysis of Deviance shown in Table ??, akin to Model 1, shows statistically significant effects of Dataset (χ^2=274.98, p<0.0001), X (), Y (), Depth, and significant interaction effects between Dataset and X and Y dimensions (). Post-hoc contrasts using estimated marginal means (EMMs) reveal each simulation dataset differs significantly from the observed data ($p < 0.01$ for all, adjusted with the Benjamini-Hochberg method for 3 comparisons). However, there are notable interactions with position.

To account for these interactions, we compute EMMs for each dataset by both X and Y position, and plot the results in Fig. 2. These results show that the simulated data differs significantly from the observed data across most X and Y values. Specifically, we see significant differences for the +Abandon simulation at $X \in \{-1, 1\}$ (p<0.05), for the Random simulation at $X = -1$ (p<0.01), the +Leafs simulation at $Y = 1$ (p<0.01), and the +Abandon simulation at $Y = -1$ (p<0.01). Even more highly significant differences are seen for the Random simulation at $X \in \{-1, 1, 2, 3\}$ and $Y \in \{-1, 1\}$ (p<0.001), the +Leafs simulation at $X \in \{-1, 1, 3\}$ and $Y = -1$ (p<0.001), and the +Abandon simulation at $X = 3$ (p<0.001).

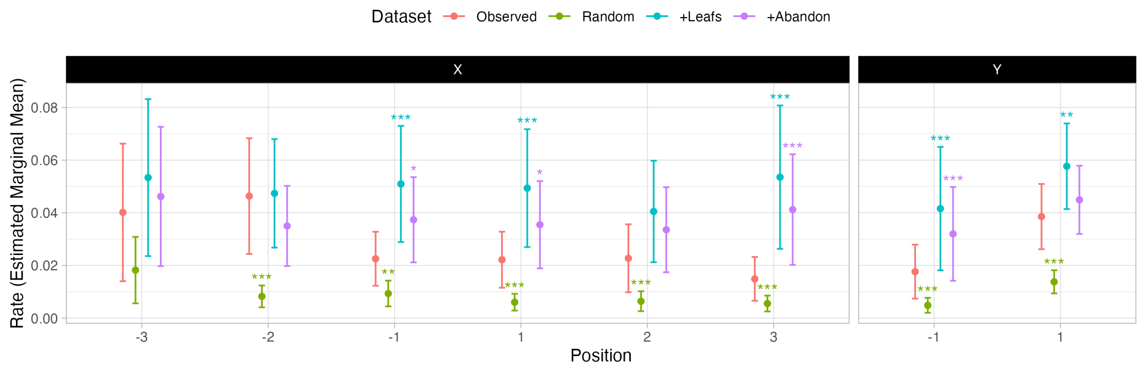

Fig. 2. Estimated Marginal Mean (EMM) rates for button presses by button XY position and dataset (observed vs. the three simulations). Though in some cases the simulations are close to the observed rates, several significant differences remain. * indicates p<0.05, ** shows p<0.01, and *** shows p<0.001. Overall, the observed behavior deviates from all simulations, suggesting factors beyond the interface structure may play a role in determining observed behavior.

These results indicate that none of the simulation methods fully capture the observed nuances of button press rates across different (X,Y) positions.

6.2 Corroboration

A total of 463 words were deemed corroborable, and three of the researchers independently rated the corroborable interactions, with an inter-rater reliability rate of 95%.

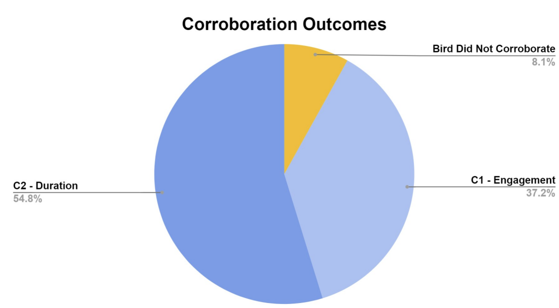

Fig. 3. Cororboration outcomes

Polly's positive corroboration rate was 92%, comprising engagement and duration behaviors. 54% of her corroborations were made with duration (e.g., activities), and 37.2% with engagement (e.g., foods and beverages) (Figure 3) Of those that were not corroborable, 37.5% were unable to be corroborated due to the caregiver declining the selection, 47.1% were due to the caregiver not following up on the request, and 15.4% were because the technology failed so it was not possible to observe whether or not the parrot corroborated (Figure 3).

Category	Total # Bird Selected	% of Bird Selection	% on AAC Device
Social Interaction	1254	24.5%	17.6%
Self Reporting - Feeling	1013	19.8%	5.8%
Activity	770	15%	16.1%
Food/Beverage/Treat	716	14%	13.5%
Request for Learning Theme	344	6.7%	12.5%
Self Reporting - Temperature	333	6.5%	4.7%
Environmental Info	325	6.3%	7.9%
Ambiance (Music/Cartoons)	274	5.3%	16.7%
Location	61	1%	3.5%
Request for Object	12	.2%	1.2%
Totals	5102	100%	100%

Three Enrichment Themes	Category	% Parrot Selected	% AAC Distribution
Socio-Cognitive	Social Interaction	72.5%	56.7%
	Self Reporting-Feeling		
	Activity		
	Request for Learning Theme		
	Self Reporting - Temperature		
Food/Beverage/Treat	Food/Beverage /Treat	13.5%	14%
Environmental Information	Environmental Info	12.8%	29.3%
	Ambiance (music/cartoon)		
	Location		
	Request for Object		

Fig. 4. Enter Caption

6.3 Category Selection and Emergent Themes

6.3.1 Category Selection. There were many menus and words that fit into categories based on similarities. The categories that were initially selected are shown in Table X with word counts, the percentage of times the parrot selected that category, and the category distribution on the AAC device.

Table X lists each of the themes, the number of times selected by the parrot, and the percentage of time it was selected by the parrot. The listing of words for each category is located in Appendix X.

Three Emergent Themes

To assess the potential enrichment value of the speech board, we categorized the AAC words in the above categories into three themes: Social-Cognitive Interactions, Food/Beverage/Treat, and Environmental. The categories comprising these themes, their use by the parrot, and the board distribution are demonstrated in Table X.

Enrichment Activities

Activity	Total time (min/sec)	Total count	Avg time (min/sec)
Tracing and Writing	228:24	27	8:45
Tablet Phonics Typing Letters	224:11	40	5:60
Tablet Games	147:40	39	3:78
Book (Picture)	139:30	49	2:84
Outside Walk	130:00	11	11:82
Reading	75:73	12	6:31
Video Call	55:55	4	13:89
Car Ride	55:00	3	18:33
Craft	20:00	1	20
Cards	18:73	5	3:75
Visit with Bird	18:38	4	4:60
Look at Art	11:37	9	1:32
Total	**18.78h**	**204**	-

Enrichment value Polly initiated interactions 85% of the time, and the caregiver initiated 15% of the interactions. Polly initiated sessions in one of three ways: walking up to the tablet 44.9% of the time, ringing a bell 37.8% of the time, and vocalizing 2% of the time.

Use of the speech board resulted in a variety of activity affordances to the parrot based on her selection behavior. Independent from her use of the AAC device, we analyzed what social-cognitive activities she engaged in, the frequency of their use, and the duration over the course of the 7 months as seen in Table X.

7 DISCUSSION

The goal of this study was to investigate the expressive and enrichment potential of AAC devices with animals. To this end, we propose a framework of evaluation predicated on a coding/labeling schema and, in our case, applied to seven months ' worth of data from the sustained AAC use by a parrot.

In order to make sense of the large quantity of data, we developed a coding and labeling schema that included documenting various aspects of the bird-human-computer interaction. We applied the schema to the video recordings to encapsulate relevent information related to the details of the parrot's use of the speech board. We the ran the models and determined whether or not the parrot's interactions with the AAC device could be predicted by computational models. Here, we reflect on the framework and resulting observations focussing around the research questions for :

RQ1: To what extent can we evaluate one parrot's use of a speech board? RQ2: What theme distribution emerged from the use of the AAC device? RQ3: To what extent is the use of a speech board providing enrichment value?

7.1 Randomness Analysis

Our goal in modeling the observed data and conducting random simulations was to determine if the parrot's interface selections follow simple heuristics or suggest additional factors have a substantial influence.

Our analysis showed that spatial biases exist but do not fully explain the observed behavior. This hints at preferential selection of certain destination words over others. Though our simulation methods suggest that heuristics such as arbitrary abandonment and repeated presses or sequential destination word selections may be at play, all simulations remained significantly different from observed behavior.

The inability of interface structure alone to explain selections motivates researching other factors like intentions, goals, and comprehension. Our work showing deviation from random simulations provides a foundation for future studies illuminating potential cognitive capacities underlying selections. For example, preferential selection of certain buttons could reflect learned associations between buttons and exogenous rewards.

Our simulations were also limited by available data and necessary choices. For example, our predictors and simulation heuristics were based on a combination of information available and observations of real data. However, the design space of possible simulation methods is large and complex, and future work could explore strategies such as weighted probabilities for different actions. Such experiments, though challenging to parameterize, could further illuminate different mechanisms involved in the parrot's icon selection behavior.

In prior HCI work, techniques like reinforcement learning have been applied to learning user policies based on expected rewards [2, 7]. Explicitly modeling expected rewards associated with destination words could help derive more accurate and complex policies to simulate the parrot's interface behavior, and in turn allow examining hypotheses related to these rewards. Other possible directions include inferring cognitive models from observed data, an emerging paradigm in HCI research [15, 18].

Building on our results, focused experiments varying interface structure and rewards can be used to test motivational hypotheses. Incorporating timing and sequence data could also reveal how selections change over time. More sophisticated simulations might capture additional factors like perceptual biases and reward effects. For example, in this work we operationalized a simple proxy variable for visual salience, the average brightness of a button. We found this did not significantly predict the button press rate. Though it's possible that visual salience of different buttons has minimal effect on the parrot's choices, more sophisticated visual salience metrics tailored to the parrot visual system could more accurately detect any underlying effects. Elucidating additional factors behind observed selective patterns through developing such designs remains an exciting open challenge on the path to better understanding animal-technology interactions.

7.2 Corroborations of Expressions

Animal behavior observation through time-on-task and engagement with activities and physical objects yielded a high percentage (92%) of corroboration on a subset of corroborable leaf word requests (492 corroborable selections). This is obtained in a blinded caretaker paradigm where the human was standing behind the tablet, and was supported by a high IRR of 95% reflecting consistency between coders.

This level of agreement between the parrot's choices and her subsequent actions suggests a strong alignment between the use of the AAC device and some level of expressive intent, defined as the sense of agency provided by the system . Ensuring that a greater portion of selection are corroborable through behavior could help improve our understanding of the expressive enrichment provided by AACs. By looking more in-depth into the type of selections that didn't lead to corroboration analysis, 37.5% were due to the caregiver actively declining the parrot's selection (i.e.: additional treats, going outside during rain, etc) . While not analyzed here, future work could explore the parrot's use of words that are declined to observe the parrot's response to and use of the declined words over time. Additionally, 15.4% of non-corroborable instances were due to technology failures, emphasizes the need for ease of access for caregivers to reliable technologies in documenting the use of AACs.

7.3 Enrichment Value

In our schema design, the categorization of Polly's word selections into ten distinct themes sheds light on the variety of enrichment provided by the AAC interaction. Such categorization could potentially lead to a better understanding of animal preferences and intentionality. Companion parrots have a variety of enrichment needs, including cognitive and social that technology could help tackle.

Among the themes identified, "Socio-Cognitive Interaction" emerged as the mode frequently selected, constituting 72.5% of the icons the parrot chose. Although the bird is already provided by default with four hours of supervised play activities on a daily basis, the AAC device appears to be used primarily to request some of such specific activities

The parrot often rang her bell or walked to the AAC device as soon as the caregiver arrived, and Polly averagely remained in engaged in the interactions for over 24 minutes per session.

These findings may highlight a potentially meaningful role the AAC device played in social and cognitive engagement for the parrot. Regardless of the parrot's capacity for intention, it served as a portal for interaction between the parrot and her human companion.

Moreover, the variety of social-cognitive activities available to Polly and that she engaged in demonstrates the diversity of enrichment that is available through AAC devices. Such activities included playing tablet games, exposure to new vocabulary, and engaging with picture books, providing varied cognitive and sensory experiences. These activities also offer avenues for bonding and social interaction. This aligns with previous literature on animal enrichment that emphasizes the importance of providing a variety of sensory, attentional, motor, and cognitive experiences.

Guidelines: The enrichment value of our study offers insights for HCI, particularly as it relates to a diversity of enrichment selections. Interactive systems should encourage caregivers to offer diverse affordances to the non-human users to contribute to varied attentional, sensory, multi-modal enrichment opportunities. While our study did not include an evaluation of agency, the variety of selections also points toward the need for agency-based preference expression in technology for non-human users. Technology should be respectful of the user's natural behaviors and needs, and systems should prioritize low-frustration interactions and safety.

PHATIC: While we cannot definitively determine whether the parrot was genuinely experiencing temperature changes, or recognizing them in her environment, nor do we know her emotional states, it is plausible that her use of temperature- or feeling-related terms mirrors human conventions of phatic expression. Phatic expression refers to ritual speech acts used primarily for social purposes rather than conveying substantive content. For instance, in human interactions, phrases like "how are you?" or "nice day, isn't it?" are often employed as greetings or conversation starters to foster social connections. Similarly, the Goffin's Cockatoo may be using sub-menus related to temperature to initiate and reinforce social bonds with her caregiver.

Phatic expressions: HCI designers and researchers should be attuned to the potential role of phatic expressions in animal communication, acknowledging that they serve not only functional but also social and emotional purposes.

7.4 Living Project

The research with Polly represents a living project of interaction. Over the course of seven months, the parrot's engagement with the speech board evolved. Menus and words were added and deleted. Her preferences and interactions changed organically. The dynamic nature of the project demonstrates the importance of understanding technology interactions as living processes.

Based on the "living project" nature of the work, users' preferences and behaviors may change over time. The HCI community can develop systems that can personalize content to users' preferences and unique interests. We recommend designing technology interfaces that are flexible and adaptable to user-initiated changes, allowing users to modify the interface as much as may be possible.

Guidelines: Based on the "living project" nature of the work, we recommend the following: Longitudinal Engagement: Recognize that users' preferences and behaviors may change over time, and consider ways to adapt the technology to meet their changing needs. Personalization: Develop systems that can personalize content to users' preferences and unique interests. Agency: Although we did not investigate findings around the agency, allow users to express their choices and potential intentions, creating a sense of empowerment and control over their interactions. Record Personal Histories: Incorporate features that record personal histories of user interactions to enhance future interactions and personalization. Flexible Interfaces: Design technology interfaces that are flexible and adaptable to user-initiated changes, allowing users to modify the interface as much as may be possible. User Development: Support user growth and development, allowing for simpler interfaces at the beginning and increasing in complexity in time, with an interface that makes such changes accessible and simple for caregivers. Playfulness: Encourage playfulness and curiosity within HCI interfaces. The cockatoo's expressions often had a playful and exploratory nature. This approach can make interactions more engaging and enjoyable.

7.5 Technological Considerations

This study raises the question of designing technology interfaces for non-human users. While HCI traditionally focuses on human-centered design, this study demonstrates the potential for extending HCI principles to benefit caregivers and, by extension, their non-human animals. Due to the enrichment this brings to both caregiver and the parrot, future HCI research could explore the design aspects required to create technology interfaces specifically for non-human animals, taking into account their limitations and body orientations.

Guidelines: Based on the foregoing, our technology recommendations are: Interface Customization: Providing caregivers (and possibly non-human users) the ability to customize their interface, such that it can adapt to the specific needs and preferences of the user.. Safety: Interfaces should be designed to afford high levels of safety for non-human users, and consider opportunities for them to be able to interact with the interface without supervision. Independent User Use: Interfaces should emphasize the ability of non-human users to access technology in a meaningful way potentially independent of the caregiver, and even facilitate remote transferring of information from the interface to the caregiver's technology devices.

7.6 Future Considerations and Research Opportunities

Many open questions remain after observing the data from Polly...

7.7 Limitations

This work took a large team of researchers many hours and over a year to begin examining the data, and we have not yet completed all the possible analyses or work in this data set. One interesting consideration for CHI, however, includes the design considerations for building a speech board. The current speech board has some attractive features such as the low cost and highly customizable interface, however it does not have internal assessment options, which could be a valuable design addition. AAC assessment options exist for humans (primarily assessing language abilities) and could be adapted, or new interfaces developed, to target considerations for animal use. All of these elements would be valuable contributions to the field of CHI.

Examining data from one parrot over seven months, while producing an enormous amount of data, is also a limitation. Future research could include looking at data over an extended time period, which might elucidate patterns and activities for additional consideration. Looking at additional parrots' use of AAC devices could produce more generalizable data. The speech board application evolves regularly with new words and phrases being added and locations changing, which is both a limitation for data collection variability and an asset to show the advanced abilities of the parrot. The large quantity of data made it impossible to delve into every interesting aspect of this project, so we hope to examine other areas within this data set, including the navigation into sub-menus, the selections made in specific contexts, and more.

This is an enormous amount of data and our process evolved with our knowledge. While we did a rigorous job developing this system and are refining it as we progressed through the data, we will need to continue to refine this iterative and continuously evolving process. For example, although some of the categories were easy to choose (To Drink menu items easily fit into Food/Beverage category), others were more complex and we had to make decisions that could be considered subjective (Talk About menu items are able to fit into multiple categories). Our continued work in this area may elucidate additional clarity.

Despite the limitations and the need for future research in this area, we do think the work done here is useful and recommend taking a systematic approach similar to this when looking at animal AAC use.

8 CONCLUSION

This study examined seven months of video data from a Goffin's cockatoo's sustained use of a table-based AAC speech board totalling over 200 sessions on 127 different days. Videos were logged, labeled, coded, and analyzed to find the answers to three questions. The value of this work includes the creation of a labeling and coding scheme and an evaluation framework for exploring the use of tablet-based AAC data in parrots in addition to exploring the expressive and enrichment potential. Contributions from this work include

Results suggest that the choice of destination words cannot be simply explained based on random selection, luminescence, or icon location alone, and 92% of corroborable selections are validated by either engagement or time on task, indicating both expressive and enrichment value. AACs can potentially help reduce boredom, provide social and cognitive enrichment, and improve animal agency and control over their environment.

While this work shows value for using a speech board to enhance enrichment opportunities in this parrot, similar to the AAC work in humans, the technology itself cannot be the only source of enrichment. A variety of stimulation is crucial for animals in captivity, including parrots. The social and cognitive enrichment appeared more frequently than other selections, potentially indicating that those are important areas to focus on in future work. Tablet-based AAC use shows promise for improving captive animal wellbeing with increased choice and control and enhancing social bonds between animals and caregivers.

REFERENCES

[1] American Speech-Language-Hearing Association. 2023. American Speech-Language-Hearing Association. Retrieved Sep 7, 2023 from https://www.asha.org/public/speech/disorders/aac/

[2] Gilles Bailly, Mehdi Khamassi, and Benoît Girard. 2022. Computational model of the transition from novice to expert interaction techniques. *ACM Transactions on Computer-Human Interaction* (2022).

[3] Erin Beneteau, Ann Paradiso, and Wanda Pratt. 2022. Telehealth experiences of providers and patients who use augmentative and alternative communication. *Journal of the American Medical Informatics Association* 29, 3 (2022), 481–488.

[4] David Beukelman and P Mirenda. 2012. *Augmentative and Alternative Communication.*

[5] Jennifer Cunha and Carlie Rhoads. 2020. Use of a Tablet-Based Communication Board and Subsequent Choice and Behavioral Correspondences in a Goffin's Cockatoo (Cacatua goffiana). In *Proceedings of the Seventh International Conference on Animal-Computer Interaction*. 1–9.

[6] Jiamin Dai, Karyn Moffatt, Jinglan Lin, and Khai Truong. 2022. Designing for Relational Maintenance: New Directions for AAC Research. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[7] Seungwon Do, Minsuk Chang, and Byungjoo Lee. 2021. A simulation model of intermittently controlled point-and-click behaviour. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[8] WH Dudley. 1939. The vocoder. *Bell. Labs. Rec.* 18 (1939), 122.

[9] Susan Fager, Lisa Bardach, Susanne Russell, and Jeff Higginbotham. 2012. Access to augmentative and alternative communication: New technologies and clinical decision-making. *Journal of pediatric rehabilitation medicine* 5, 1 (2012), 53–61.

[10] R Allen Gardner, Beatrix T Gardner, and Thomas E Van Cantfort. 1989. *Teaching sign language to chimpanzees.* Suny Press.

[11] Emilie Genty, Raphaela Heesen, Jean-Pascal Guéry, Federico Rossano, Klaus Zuberbühler, and Adrian Bangerter. 2020. How apes get into and out of joint actions: Shared intentionality as an interactional achievement. *Interaction Studies* 21, 3 (2020), 353–386.

[12] Seray B Ibrahim, Asimina Vasalou, and Michael Clarke. 2018. Design opportunities for AAC and children with severe speech and physical impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[13] AAC INSTITUTE. 2023. Resource and Advocate for the AAC Community. Retrieved Sep 7, 2023 from https://aacinstitute.org/what-is-aac/

[14] Simon Judge and Mark Landeryou. 2007. Disambiguation (predictive texting) for AAC. *Communication Matters Journal* 22, 2 (2007), 37–41.

[15] Antti Kangasrääsiö, Kumaripaba Athukorala, Andrew Howes, Jukka Corander, Samuel Kaski, and Antti Oulasvirta. 2017. Inferring cognitive models from data using approximate Bayesian computation. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1295–1306.

[16] Oihana Leonet, Maria Orcasitas-Vicandi, Argia Langarika-Rocafort, Nahia Idoiaga Mondragon, and Gorka Roman Etxebarrieta. 2022. A systematic review of augmentative and alternative communication interventions for children aged from 0 to 6 years. *Language, Speech, and Hearing Services in Schools* 53, 3 (2022), 894–920.

[17] Janice Light and David McNaughton. 2014. Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication? , 18 pages.

[18] Hee-Seung Moon, Antti Oulasvirta, and Byungjoo Lee. 2023. Amortized inference with user simulations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[19] Andrea Rabinowitz. 2016. *Linguistic competency of bonobos (Pan paniscus) raised in a language-enriched environment.* Ph. D. Dissertation. Iowa State University.

[20] Lucia Reily, Helena Panhan, and Ariane Tupinambá. 2009. Early Evidence of Low-Tech Communication in an Otto Dix Painting of 1920. *Augmentative and Alternative Communication* 25, 4 (2009), 217–224.

[21] Rose Sevcik, Mary Ann Romski, and Bronwyn Robinson. 1996. Longitudinal research: Considerations for augmentative and alternative communication. *Augmentative and Alternative Communication* 12, 4 (1996), 272–276.

[22] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K Kane. 2023. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[23] Stephanie Valencia, Amy Pavel, Jared Santa Maria, Seunga Yu, Jeffrey P Bigham, and Henny Admoni. 2020. Conversational agency in augmentative and alternative communication. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[24] Gregg C Vanderheiden. 2003. A journey through early augmentative communication and computer access. *Journal of rehabilitation research and development* 39, 6; SUPP (2003), 39–53.

[25] Lindsay Wilson. 2021. From Clever Hans to Bunny the TikTok Dog: An Exploration into Animal-to-Human Communication. *Johns Hopkins University* 2, 1 (2021).

Received ; revised ; accepted