Preliminary Assessment of Siamang (*Symphalangus syndactylus*) Cognition Using Digital Cognition Testing Software and Touchscreen Technology

Gina M. Munir[1,2,*], and Paul M. Nealen[1]

[1]Department of Biology, Indiana University of Pennsylvania
[2]Department of Anthropology, Southern Illinois University

*Corresponding author (Email: vaimun.gina@gmail.com)

Abstract – Studies of nonhuman primate cognition have traditionally employed subjects from the monkey (*Cercopithecoidea*) and ape (*Hominidae*) lineages, with relatively much less examination of the phylogenetically intermediate gibbons (*Hylobatidae*). To begin to rectify this gap in our understanding of primate cognition, we used a short-term training and testing method to preliminarily evaluate siamang gibbon (*Symphalangus syndactylus*) performance on a battery of standardized cognitive tasks, using Monkey CANTAB cognition testing software (Lafayette-Campden Neuroscience) and a computer touchscreen apparatus. Five different operant tasks examining distinct perceptual and cognitive functions were used to train and test each subject over a period of five weeks. Over 1900 cognitive testing trials were completed with three subjects. Despite limited training and testing opportunity, all siamangs achieved proficiency scores well above chance (> 65%) in both Concurrent Discrimination and Spatial Working Memory tasks. Over the limited training opportunities available, our subjects did not achieve proficiency in Delayed Match-to-Sample, Paired Associates Learning, or Conditional Visual Discrimination tasks; however, subjects did exhibit (to varying degrees) improved performance in these tasks, suggesting that additional training may lead to higher proficiency and that siamangs demonstrate effective, task-based learning capabilities. We present one of the first successful tests of a hylobatid ape using touchscreen cognitive testing methods and demonstrate that hylobatids can be evaluated using the same methods now commonly employed in tests of other nonhuman primates. Further, our testing regime proved to be behaviorally rewarding for our subjects, suggesting that hylobatids may benefit from advanced forms of interactive cognitive enrichment.

Keywords – Hylobatidae; Cognition; Siamang; Gibbon; Nonhuman primates

_____

Cognition is the study of the complex mental processes that guide advanced behaviors, such as the cracking of a nut with a stone, or evaluating variable "ripeness" among the thousands of food items found in tropical forests (Hill, 2006; Vodel et al., 2009; Zuberbühler & Janmaat, 2010). One of the best ways to observe and study these processes is by examining the cognitive behaviors of our closest living relatives, the nonhuman primates (NHP). Although comparative study of nonhuman primates is by no means a new endeavor, the scope of species involved has been relatively limited. Most studies of NHP cognition evaluate members of the catarrhines (*Cercopithecoidea*; Egelkamp et al., 2018; Joly et al., 2014; Munger et al., 2017; Nagahara et al., 2010; Ryan et al., 2019; Rodriguez et al., 2011; Taffe & Taffe, 2011; Vogel et al., 2009; Weed et al., 2008; Zurcher et al., 2010) or the great apes (*Hominidae*; Beran et al., 2016; Call, 2001, 2010; Cronin et al., 2017; Damerius et al., 2019; Hopper et al., 2019, 2021; Inoue & Matsuzawa, 2007;

Kanngiesser & Call, 2010; Köhler, 1925; Morimura & Matsuzawa, 2001; Osvath & Martin-Ordas, 2014; Rumbaugh & Rice, 1962; Vonk, 2013; Vonk et al., 2013; Vonk & MacDonald, 2004; Vonk & Rastogi, 2019; Wagner et al., 2016; Wobber et al., 2014; Yerkes & Yerkes, 1929). Despite certain species of hylobatid (e.g., *Hylobates lar*) being commonly housed in captivity, most studies exclude gibbons and siamangs (Abordo, 1976; Andrieu et al., 2020; Caspar et al., 2018; Cunningham et al., 2006; Sánchez-Amaro, 2021), which occupy a phylogenetic position between the two more-commonly studied groups (Hopper, 2017).

In many ways, hylobatids are unique when compared to other extant apes. They are the only obligatory brachiators (of the extant apes), and they are also much smaller in mass than other ape species (Bartlett, 2007; Brockelman et al., 1998). Behaviorally, one of the most defining characteristics of the hylobatids is their pair-bonded sociality. Hylobatids are the only apes that demonstrate long-term pair-bonding and live in small family groups, which most commonly consist of an adult breeding pair and their offspring (Leighton, 1987). This social structure includes close social bonds between the mating pair, vocal duetting between mates, lack of a prominent dominance hierarchy among group members, little to no sexual dimorphism, and a shared territory, which the pair defends together (Brockelman et al., 1998; Chivers, 1974, 1976; Liebal, 2016; Reichard et al., 2016). However, recently there have been reports of hylobatid group "plasticity," with family units consisting of two males, one female, and their shared offspring, along with extra-pair copulations by females living in one-male, one-female pair groups, suggesting that some female gibbons may be polyandrous (Barelli et al., 2008; Palombit, 1994; Reichard et al., 1995). Even with these findings, pair-bonding and small family group-living are staples of hylobatid sociality, tightly coupled to their ecology and behavior, and also, very likely to their cognition (Liebal, 2016).

Comparative Neuroanatomy is Predictive of Hylobatid Cognition

As hylobatids are underrepresented in cognitive studies, one can use comparative neuroanatomy to suggest how hylobatid cognition may compare to that of other NHP taxa. As noted by Jerison (1973) and others, brain volume varies widely across the hominoids, from 400-700 cm³ in gorillas (*Gorilla gorilla*), to 300-500 cm³ in orangutans (*Pongo pygmaeus*) and 300-400 cm³ in chimpanzees (*Pan spp.* Schultz, 1930). Siamangs (*Symphalangus syndactylus*) exhibit the largest body and brain masses (100-150 cm³; Tobias, 1968 among the hylobatids, with *Hylobates* and *Nomascus* gibbons having reported brain sizes ranging from 80-120 cm³ (Schultz, 1930; Tobias, 1968, 1971). That hylobatids have the smallest brain mass among the apes is expected, given that they also have the smallest body sizes within the Hominoidea (Jerison, 1973). Given the noted allometric relation between body and brain masses within the apes (Jerison, 1973, 1982; Smith & Jungers, 1997), multiple investigators have recognized the need to consider relative, rather than absolute, encephalization, suggesting that the amount of cortical expansion itself (measured by the overall surface area of the brain) may be more directly related to cognitive performance (Cunningham et al., 2016; Passingham, 1981).

In comprehensive studies by Rilling and Insel (1999) and Rilling and Seligman (2002), the neocortical volumes of 11 species of living haplorrhines were measured via MRI imaging. The neocortex volume (measured as neocortex ratio) for one species of gibbon (*Hylobates lar*) was reported to be 1.16, compared to those of the orangutan (1.14), chimpanzee (1.3), gorilla (1.0), baboon (sp. *Papio*, 0.8), macaque (sp. *Macaca*, 0.71) (Rilling & Insel, 1999). By this neocortex ratio measurement, hylobatids exhibit encephalization comparable to that of the great apes, and greater than that of baboons and macaques. However, these ratios are somewhat non-uniform across the primate lineage. For example, the measured neocortex ratio of the capuchin monkey (Family Cebidae) was 1.28 (Rilling & Seligman, 2002).

As an alternative measure of cortical enlargement, Zilles and Rehkämper (1988) proposed a "Gyrification Index" as an evaluation of the degree of neocortical folding. By this measure, gibbons, with a Gyrification Index of 1.9, exhibit lesser encephalization than do the great apes (orangutan (ssp. *Pongo*): 2.29, gorilla (sp. *Gorilla*): 2.07, chimpanzee (sp. *Pan*): 2.19; $M = 2.18$), and are more similar to monkeys (baboon (sp. *Papio*): 2.03, mangabey (sp. *Cercocebus*): 1.84, macaque (sp. *Macaca*): 1.73; $M = 1.87$; Rilling & Insel, 1999) in this regard. Likewise, Semendeferi et al. (2002) demonstrated that the frontal lobe

proportional volume of gibbon *Hylobates lar* (29.4% of cerebral hemisphere volume) was more similar to that of capuchin (30.6%) and macaque (30.6%) monkeys than to that of the larger apes (orangutan: 37.6%; gorilla: 35.0%; chimpanzee: 35.4%; bonobo: 34.7%; M = 35.7%; Semendeferi et al., 2002).

Hylobatids are represented relatively poorly in these neuroanatomical assessments, with most studies evaluating a very small number of individuals of a single genus, typically *Hylobates* (Cunningham et al., 2016; Semendeferi et al., 2002). If cognitive capability is indeed simply related to aspects of whole-brain or frontal lobe volume/folding, the limited data available suggest that hylobatids should be at least as cognitively capable as the monkeys (*Cercopithidae*). More comprehensive neuroanatomical data is needed for all species of hylobatids to gain a more complete understanding of their neurological structure, capability, and intelligence (Cunningham et al., 2016).

History of Hylobatid Cognitive Evaluation

Prior considerations of the cognitive abilities of hylobatids have often described their understudied and poorly understood nature as that of an enigma (Abordo, 1976; Cunningham et al., 2006; Rumbaugh, 1970). The French zoologist, Louis Boutan, noted learning with voluntary attention [analogous to what we today would term 'insight' (Abordo, 1976)] in one gibbon he tested with a series of puzzle boxes similar to those used by Thorndike (1911) in his tests of cat problem solving (Boutan, 1913, 1914). Additionally, while examining the problem-solving abilities of hylobatids using tests modeled after Köhler's "string" experiments, Beck (1967) reported that his hylobatid subjects performed on par with Köhler's chimpanzees.

Rumbaugh and McCormack (1967) reported very mixed hylobatid performances in their comparative studies on associative learning and object discrimination. While most hylobatid subjects tested poorly in comparison to other apes (*Pan, Gorilla, and Pongo*), Rumbaugh and McCormack (1967) also noted that modifications of their testing method (removing the reward from behind a glass viewing window) during their study allowed their gibbon subjects to perform at higher levels of proficiency, serving as an important reminder that not all forms of NHP cognitive evaluation are equally useful across study species (Abordo, 1976; Bates & Byrne, 2015; Beck, 1967; Beran et al., 2016; Tomasello & Call, 1997).

Despite limited studies, hylobatids have been reported to exhibit tool use (Cunningham et al., 2006, 2011, 2016), self-awareness (Fedor et al., 2008; Heschl & Fuchsbichler, 2009; Ujhelyi et al., 2000), and selective problem-solving abilities (Cunningham et al., 2006, 2011, 2016; Hill et al., 2011; Inoue & Inoue, 2002). Similarly, their unique-complex pair-bonded sociality, communicative duetting abilities, and neural encephalization may suggest a relatively high level of cognition (Bates & Byrne, 2015; Byrne, 2004; Dunbar, 2003; Freeberg et al., 2012; Reichard & Barelli, 2008). To generate data that would inform questions about hylobatid cognitive capabilities, we performed a preliminary evaluation of hylobatid cognition with siamang gibbons (*Symphalangus syndactylus*), using an array of five standardized cognitive tasks presented on a touchscreen.

Current Focus

To quickly identify which cognitive tasks may prove most worthwhile for hylobatid cognitive evaluation, we employed a "rapid assessment" study design, in that each subject received limited training on each cognitive task. The primary goals of our study were to (*i*) test whether touchscreen devices can be used for hylobatid cognitive assessment via operant techniques, (*ii*) perform a preliminary evaluation of the kinds of cognitive tests that may be useful in evaluation of hylobatid cognition (from a battery of standardized and commonly used cognitive tasks), and (*iii*) promote future cognitive testing and enrichment in hylobatids. To these ends, we show that (*i*) hylobatid cognition can be evaluated using touchscreen testing methods and digital task technology, and (*ii*) most subjects readily adapted to operant instruction, routinely performing at high work rates, and actively seeking additional instrument interaction. That our subjects actively engaged with our tasks and achieved test proficiencies > 65% on some tasks over relatively short training/testing periods demonstrates the potential of touchscreen devices in both hylobatid cognitive

evaluation and cognitive enrichment, while also providing evidence that hylobatid cognitive abilities in relation to learning may be more flexible than previously reported.

Method

Ethics Statement

This research was approved by all participating institutions, including the Indiana University of Pennsylvania (IACUC log #01-1617-R1), the Pittsburgh Zoo and Aquarium (via their IACUC Committee and Director of Research), and the Erie Zoo (Erie, PA; via the zoo director, with additional approval from the Association of Zoos and Aquariums [AZA]). The research protocol and procedures were approved by each zoological institution's management and keeper staff, with all training and testing sessions conducted in the presence of a primary keeper. The test subjects remained the property of, and under the care of, their home institutions at all times.

Subjects and Housing

A total of four siamangs (two adult, monogamous, male-female pairs) participated in this study. All subjects had been raised by their parents (not hand-reared), and, to our knowledge, were never exposed to or involved in prior cognitive testing. All subjects had been exposed to small amounts of positive reinforcement training via their keepers, including target, shifting, and injection site training.

Pair One consisted of one, 7-year-old male (Merlin) and one, 6-year-old female (Leela) housed at Pittsburgh Zoo and PPG Aquarium in Pittsburgh, PA, USA. Training and cognitive testing of Pair One occurred from January-March 2017. The male (Merlin) was eventually dropped from the study because of his failure to use the touchscreen without verbal cues from GM. Pair Two consisted of one, 19-year-old male (Dandy) and one, 28-old female (Deedee) located at the Erie Zoo in Erie, PA, USA. Their touchscreen training was conducted from May-July 2017. Their cognitive testing began in July 2017. However, due to construction on a nearby exhibit, testing was delayed and conducted from December 2017-January 2018.

All subjects remained in their normal housing areas during this study and remained on their regular husbandry and exhibit schedules. Training and testing sessions occurred outside of public exhibit and mealtimes, which limited the amount of testing time available. Subjects remained on their normal diets; water was available *ad libitum*.

Procedure

Testing Station

The touchscreen testing station (Figure 1) consisted of a 19" ELO AccuTouch pressure-sensitive touchscreen monitor (ELO Touch Solutions Inc., Milpitas, CA) and an ACER laptop computer (Acer Inc., San Jose, CA), affixed on a mobile, compact computer cart (model 698 Black, e-Joy, Orange, CA). The mobile unit allowed easy introduction and removal of the touchscreen from the outside of the test subjects' enclosures.

Touchscreen Training Procedure

Touchscreen training sessions for each subject occurred three days per week, with each day comprising two training sessions (typically, one morning and one afternoon session) lasting 15-30 min per session. The touchscreen training procedure (via Monkey CANTAB, Lafayette Neuroscience) was designed to enable subjects to gain familiarity with the touchscreen via presentation of visual and auditory stimuli and food rewards (one raisin or half of one grape, at each offering). The training procedure began with a presentation of a large (5 x 5 in), red, square screen icon. With each screen touch of the training icon,

an audio tone (indicating a correct response) would sound, and a food reward was given to the subject (by hand), followed by a brief (2 seconds) disappearance of the icon. A smaller version of the same icon would then reappear. Once subjects reached the smallest icon (ca. 1.5 x 1.5 in, after 10 touches), the touchscreen testing session would end.

Figure 1

Portable Operant Discrimination Training/Testing Station

Note. Station equipped with laptop and interactive touchscreen, in front of the testing bench at the siamang enclosure of the Pittsburgh Zoo & PPG Aquarium. The operator could stand at the front left corner of the workstation (at figure left) and conduct the experiment via the laptop screen, while the subject interacted with the touchscreen (at figure upper right) through the enclosure mesh. Photo courtesy of GM.

During touchscreen training, the experimenter was positioned next to the testing apparatus and often gave verbal and/or physical (pointing at screen) cues to encourage the animals to touch the screen, especially early in training. When a subject interacted with the touchscreen without verbal or physical cues from the experimenter, cognitive training began for that subject. All subjects were physically but not visually separated from conspecifics during training/testing, with the exception of Deedee, who was not separated from her one-year-old offspring, Damai.

Cognitive Training and Testing Procedure

Once subjects were familiar with the touchscreen, cognitive training and testing sessions began. Typically, each subject experienced cognitive training/testing three days per week, with each day comprising two sessions (typically, one morning session and one afternoon session). Cognitive training/testing sessions normally included ~50 operant trials over 15-30 min per session. During all stages of cognitive training and testing, subject responses that were 'correct' resulted in the automatic sounding of a 'chime' tone (as a reinforcement bridge) via the CANTAB software. GM then rewarded the animal with a small food reward (dispensed via a feeding pipe), and a verbal reinforcement of "good." Another

trial would appear on the screen within ~1 second of reward being given. If the subject selected the wrong answer(s) during a training/testing trial, a 'buzzer' tone was emitted from the testing software, informing the subject and GM that the subject had selected the wrong answer. When this occurred, the subject would not receive verbal or food reinforcement. Incorrect responses induced a brief 'timeout' between trials (3 s), after which a new trial initiated automatically.

During cognitive training and testing, GM was not able to see the touchscreen or the task on the computer monitor, to ensure that GM could not give any unintentional cues (for example, eye or body movements) to the subjects, to avoid a "Clever Hans" effect (Pearce, 2013). GM was always stationed behind the computer apparatus during training/testing sessions and moved only to provide food rewards, as required. With minor exceptions (as noted), each subject experienced the same schedule of cognitive training and testing (Figure 2). Each subject was trained, and then evaluated, on one cognitive task per week, with subjects tested on five different cognitive tasks, sequentially, over five weeks. Subjects experienced three days of cognitive training/testing per week. The first two days (four operant sessions) of each week were used as training periods, allowing subjects the opportunity to learn how to perform the week's task.

The two operant sessions that occurred on the third (and last) day of the week were used for evaluation of subject performance on that week's task. All data presented here were generated on subject task evaluation days, after two days (four sessions) of training on that week's cognitive task. This compact schedule provided subjects only ~200 operant trials (over four sessions spanning two days) to learn a cognitive task before being evaluated on it (during the third and final day of testing for the week, comprising two 30 min sessions of ~50 trials each). This time-constrained schedule was chosen both as a function of limited subject availability (due to facility schedules for animal husbandry and exhibition) as well as a desire to perform a brief, but intense, triage of subject performance across these tasks. Subjects were able to learn some, but not all, of the cognitive tasks during this short training interval. All cognitive training and testing was done without GM instruction or intervention. Throughout the study (~ 90 training/testing sessions), early termination of a session occurred only four times due to lack of interest/subject disengagement from the testing area (Dandy: 3 sessions; Leela: 1 session). All training and testing sessions were video recorded for additional documentation.

Cognitive Tasks Employed

The siamangs in this study were evaluated via operant techniques on five different cognitive tasks: Concurrent Discrimination (CD), Spatial Working Memory (SWM), Delayed Match to Sample (DMS), Paired Associates Learning (PAL), and Conditional Visual Discrimination (CVD). These tasks were chosen specifically because they explore several areas of core cognitive abilities (Cambridge Cognition, 2016), and because of their common usage in the testing of human, great ape, and monkey cognition (Call, 2001; Crofts et al., 1999; Fedor et al., 2008; Joly et al., 2014; Köhler, 1925; Leighty et al., 2011; Nagahara et al., 2010; Rodriguez et al., 2011; Rumbaugh et al., 1972; Rumbaugh & McCormack, 1967; Taffe & Taffe, 2011; Tomasello & Call, 1997; Weed et al., 2008; Zurcher et al., 2010). Successful completion of any task required visual evaluation of on-screen stimuli (of software-controlled number, shape, color, position, and duration) followed by subject physical pressing of touchscreen icon location(s) as responses. The location of each stimulus changed (in software-controlled order) between each trial/block to ensure that location bias did not occur. Subject responses and performance at each task automatically were logged by the CANTAB software.

With minor exceptions as noted, each subject was evaluated during two operant sessions, typically of 15-30 min duration, during the third and final day of cognitive task exposure for the week (Figure 2). Task evaluation sessions are defined here as contiguous periods of time during which operant interaction was available to the subject. Each task evaluation session began with a particular cognitive task, at a specific level of task difficulty. For some tasks, task difficulty automatically increased (by increments of stimulus number or delay) if subjects reached proficiency at a given task level, resulting in several trial "blocks" of increasing difficulty within a single training session, with each trial block constituting all operant trials at

a particular difficulty level for the task at hand. Subjects' task sessions were of similar length (# of operant trials) for all cognitive tasks employed; subject tasking performance was evaluated once subject testing was completed.

Figure 2

Schematic of the General Training and Testing Schedule Employed

Note. Each subject was trained and tested on five different cognitive tasks, sequentially over five weeks, and each subject experienced only a single cognitive task each week. Where CD = concurrent discrimination, SWM = spatial working memory, DMS = delated match-to-sample, PAL = paired associates learning, and CVD = conditional visual discrimination. Each week included two days (four sessions) of training on the task for that week, followed by one day (two sessions) of evaluation of performance on that week's task. All subjects were presented with the tasks in the above order except for the Pair 1 female who was trained and tested on the tasks in this order: CD, DMS, CVD, SWM, and PAL. All data presented herein result from the final day of testing within each week.

The form of each task was as follows:

       1. Concurrent Discrimination (CD; Figure 3): This task evaluates components of core cognition, including learning and habit-forming cognition (Hood et al., 1999). During each trial, subjects were presented with two visual stimuli that differed in color or shape (stimuli icons: purple bow tie, purple triangle, yellow star, purple hourglass, white circle). Only three of the stimuli (purple bow tie, yellow star, and purple hourglass (randomly selected) were the "correct" stimuli when paired with the three other visual stimuli ("incorrect" stimuli). Each correct and incorrect stimulus was used *once* for each testing block, which consisted of ~50 trials. However, it should be noted that, because of her proficiency with this task, the protocol for subject Leela eventually included additional "incorrect" stimuli (instead of one), in order to increase the difficulty of the task. Subjects had to learn which stimulus was the "correct" stimulus during the brief two-day training phase. Subjects were rewarded for touching the "correct" stimulus, but not for touching the distractor stimulus (Cardinal & Aitken, 2010). Proficiency (10 correct trials within a testing session) resulted in an automatic increase in the number of stimuli presented (up to 4 stimuli and 4 screen locations, for subject Leela) with the "correct" stimulus remaining the same. The stimuli were located in

either the right, left, or middle quadrants of the screen. Subjects completed an average of 295 CD trials (range: 286 – 300) during their evaluation sessions.
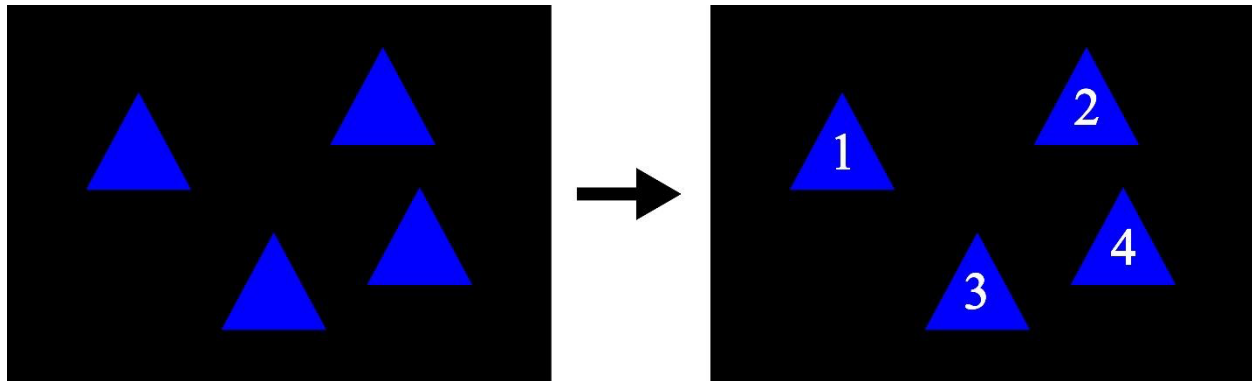
Figure 3

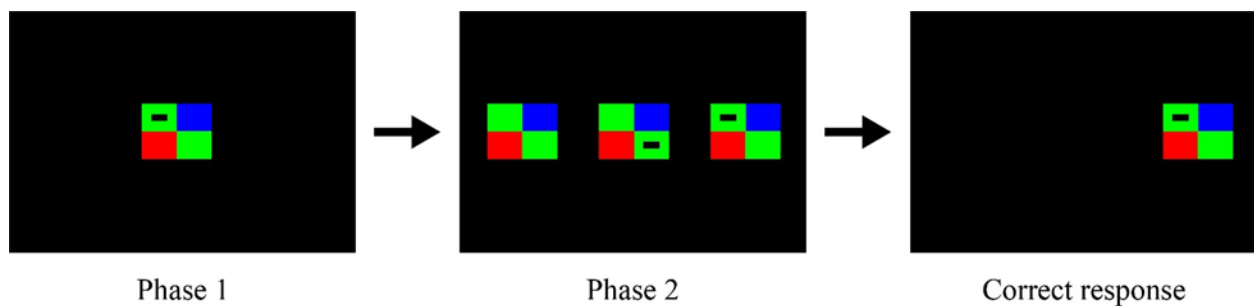*Example Trials for the Concurrent Discrimination (CD) Task*
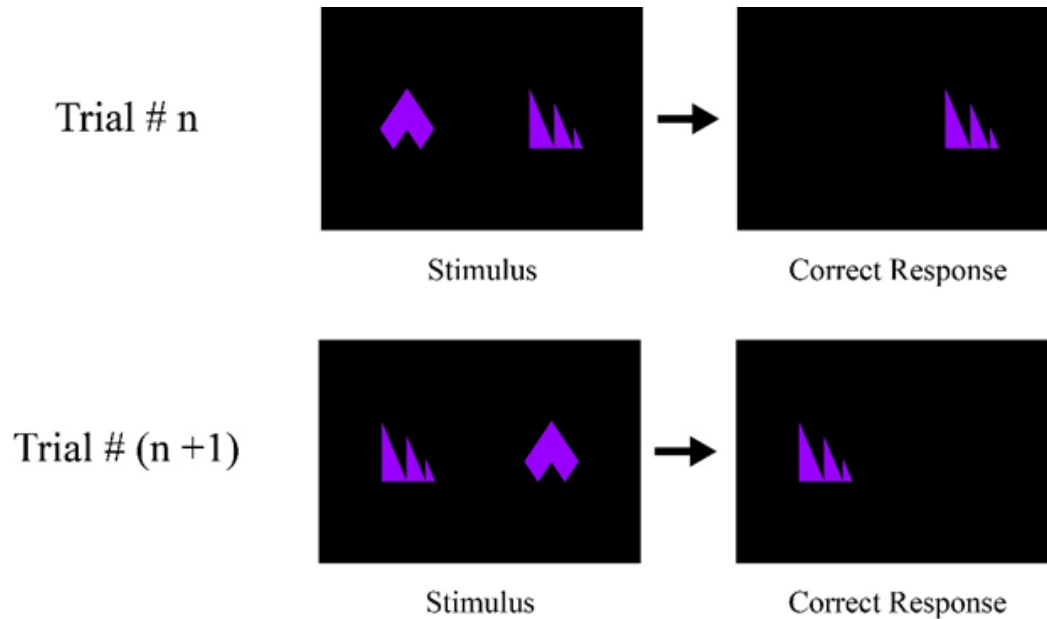
Note. Images modeled after Monkey CANTAB, Lafayette Instrument Company. Icon sizes exaggerated here for presentation.

       2. Spatial Working Memory (SWM; Figure 4): The "self-ordered" SWM task evaluates executive function, retention, and working memory (Cambridge Cognition, 2016). In this task, visually identical stimuli (yellow star icon; minimum of two) appeared on the screen in different locations (within a 16-way scattered pattern grid scheme), and the subject was required to select each stimulus in turn without revisiting a stimulus already touched (Cardinal & Aitken, 2010). Choosing the stimuli in any singular and complete sequence constituted a correct response. Failing to select all of the stimuli or "double touching" any single stimulus resulted in an incorrect response. Once a subject achieved 10 correct trials with a given number of stimuli, task difficulty automatically increased via increments of stimulus number (up to 6 stimuli for subject Leela). Subjects completed an average of 104 SWM trials (range: 70 – 125) during their evaluation sessions.

Figure 4

Example Trial for the Spatial Working Memory

Note. Images modeled after Monkey CANTAB, Lafayette Instrument Company. Icon sizes exaggerated here for presentation.

3. Delayed Match to Sample (DMS; Figure 5): This task evaluates both visual matching ability and short-term visual memory (Cambridge Cognition, 2016). In this DMS task, a target stimulus (one checkered square) appeared for a fixed duration of three seconds in the middle of the screen, and then disappeared for a three second delay interval. The target stimulus was selected by the CANTAB software for each trial from a test bank of three similar figures. Then, the target stimulus was displayed with two additional, similar but visually unique, distractor stimuli. Both distractor stimuli (non-identical checkered squares) were chosen randomly by CANTAB software from a bank of six similar stimuli, each having a different color pattern from the "correct" stimulus square. Subjects were required to identify the initial stimulus from this group (Cardinal & Aitken, 2010). The form of the target stimulus automatically changed between trials. As subjects gained proficiency within a test session, task difficulty automatically increased as a lengthening of the delay period from its initial value of 0 s to as long as 5 s. Subjects completed an average of 107 DMS trials (range: 88 – 125) during their evaluation sessions.

Figure 5

Example Trial for the Delayed Match-to-Sample (DMS) Task

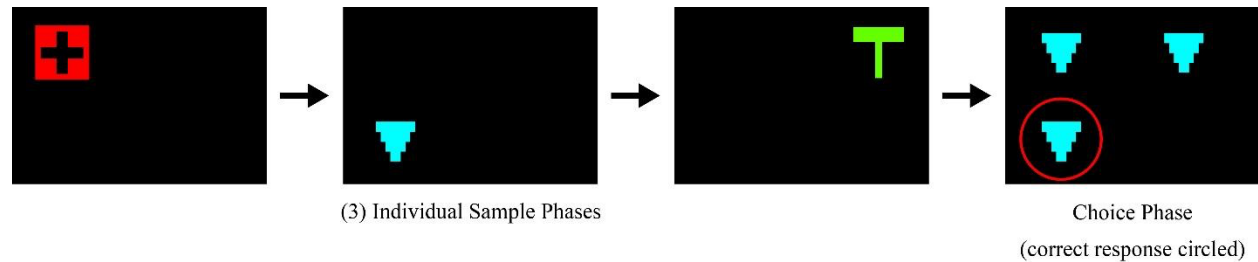Phase 1   Phase 2   Correct response

Note. Images modeled after Monkey CANTAB, Lafayette Instrument Company. Icon sizes exaggerated here for presentation.

4. Paired Associates Learning (PAL; Figure 6): This task assesses learning and visual memory (Cambridge Cognition, 2016; DeLuca et al., 2003). Subjects complete this task by responding to both the shape and the position of presented stimuli. At the start of a PAL trial, subjects were sequentially presented with three on-screen stimuli, of different colors/shapes/positions (stimuli were randomly selected by the software from those within the CANTAB icon library). After a three second delay, one of the initial stimuli was shown in all three locations at which the initial stimuli had been depicted. To correctly respond, subjects

were required to select the icon at the screen location that corresponded to the original location at which that same stimulus had been presented. Subjects completed an average of 88 PAL trials (range: 56 – 120) during their evaluation sessions.

Figure 6

Example Trial for the Paired Associates Learning (PAL) Task

(3) Individual Sample Phases
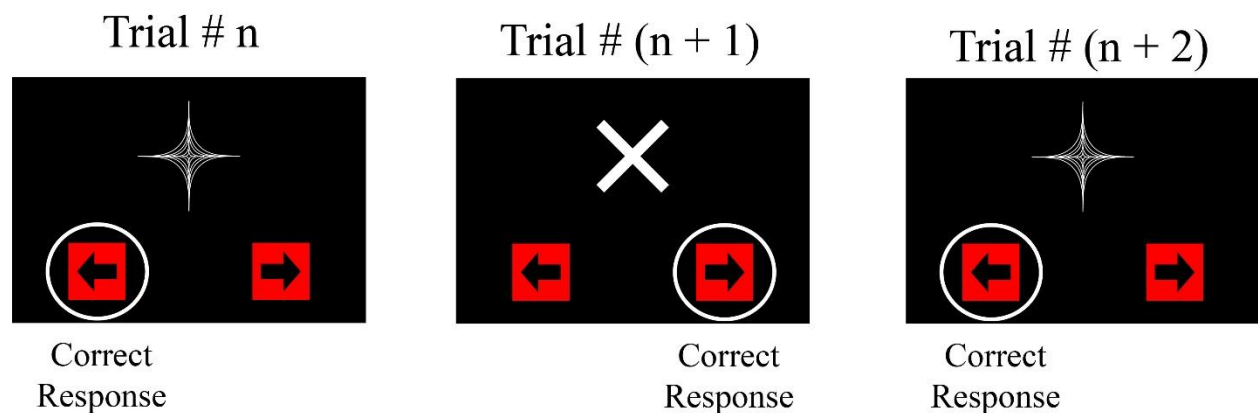
Choice Phase
(correct response circled)

Note. Images modeled after Monkey CANTAB, Lafayette Instrument Company. Icon sizes exaggerated here for presentation.

5. Conditional Visual Discrimination (CVD; Figure 7): This task evaluates retention, learning and working memory (Cambridge Cognition, 2016). CVD trials began with the onscreen presentation of three different stimuli (stimuli: red heart, blue rectangle, yellow "V") in a horizontal row. One of two possible stimuli occupied the middle position, and the identity of the middle stimulus determined whether the subject should choose the left-side or right-side stimulus as the correct response. Subjects completed an average of 116 CVD trials (range: 111 – 122) during their evaluation sessions.

Figure 7

Example Trials for the Concurrent Visual Discrimination (CVD) Task

Trial # n	Trial # (n + 1)	Trial # (n + 2)
Correct Response	Correct Response	Correct Response

Note. Images modeled after Monkey CANTAB, Lafayette Instrument Company. Icon sizes exaggerated here for presentation.

Data Analysis

Each cognitive trial was automatically scored for correct/incorrect responses. After all training and testing sessions were complete, CANTAB raw data files were exported to Microsoft Excel (Microsoft Corp.; Redmond, WA) for tabulation. Statistical analyses were performed in SPSS (v. 25.0, IBM, Armonk, NY). Analysis focused upon measures of subject performance during final (day three) evaluation sessions for each of the five cognitive tasks employed. Training data (collected during the 2 days each week prior to testing/evaluation day) were collected but unfortunately lost due to a malfunctioning hard drive. Because

we had no prior expectation for how hylobatid task performance would evolve during testing sessions, we examined subject task performance in terms of both end-of-session cumulative performance (cumulative % of trials within a testing session for which a correct response was given) as well as in terms of subject within-session peak performance (maximum cumulative performance within a session after a minimum of 10 trials). Subjects were noted as having achieved proficiency in a particular task when they achieved an end-of-test-session (or test block) cumulative performance of > 65% correct responses during the task, as determined during post-task analyses. This achievement threshold is somewhat lower than that used in other studies of primate cognition (e.g., Beran et al., 2016; Tomasello & Call, 1997; Vonk et al., 2013), and was chosen to reflect the preliminary nature of these investigations as well as the limited training available to subjects. As noted below, this *a priori* threshold of > 65% reliably separates subject cognitive performance between those tasks on which subjects performed well and those tasks which subjects failed to grasp. Performance measures were compared both across testing sessions and task types within subjects, as well as within task types across subjects. Correlations between task difficulty level and subject performance were evaluated using Pearson's *r* statistic. The distribution of subject performance variation among subjects and across task types was examined using a repeated-measures GLM. Any performance measure (cumulative or peak) that exceeded our *a priori* threshold of > 65% correct performance was evaluated for statistical significance using the binomial test. As a test of the degree to which task performance was related to task type or subject identity, we used repeated-measures ANOVA to evaluate end-of-session cumulative performance on the base (lowest) level of task difficulty, across all task types and subjects. Exact binomial probability values are reported when available; asymptotic estimates of them are reported otherwise. All statistical tests were two-tailed; the alpha threshold for the significance of statistical testing was set to .05.

Results

A total of 1,915 cognitive trials were completed by the three siamang subjects, across five different cognitive tasks. Cognitive performance varied between subjects and within subjects across tasks. The siamangs were able to achieve proficiency scores well above chance (> 50% cumulative performance) for some, but not all, of the cognitive tasks employed.

Overall, subject performance varied more strongly across tasks ($F(4, 8) = 7.484$, $p < .01$, $\eta p^2 = 0.789$) than among subjects ($F(2, 2) = 0.194$, $p = .838$). Post-hoc tests (Tukey's Least Significant Differences) across task types reveals that subject performance on CD and SWM tasks were statistically distinguishable from CVD, PAL, and DMS task performances (pair-wise task performance comparison *p* values from .058 – .079). Subjects were most successful at the Concurrent Discrimination (CD) task (Figure 8); all three subjects achieved proficiency (> 65% cumulative performance) on one or more blocks of CD test trials. All of the siamangs also exhibited proficiency in the Spatial Working Memory (SWM) task (Figure 9). Both Deedee and Dandy completed test blocks of two objects with > 65% correct performance, while Leela exhibited 100% correct performance in a test block of three objects. All subjects exhibited sufficient performance (e.g., ≥ 10 correct trials within a test block) during SWM testing to experience automatic increases in task difficulty. During both CD and SWM tests, subjects achieved sets of ≥ 10 successive testing trials that were entirely or mostly correct, clear evidence that subjects understood the mechanics of the tasks.

In contrast to their capability at CD and SWM tasks, the siamangs did not achieve proficiency on Delayed Match-to-Sample (DMS; Figure 10), Paired Associates Learning (PAL; Figure 11), or Concurrent Visual Discrimination (CVD; Figure 12) tasks during the limited training available to them. Subject end-of-session cumulative performances were near or below 50% for all of these tasks, and no subject exhibited a series of > 10 successive correct trials within these tasks.
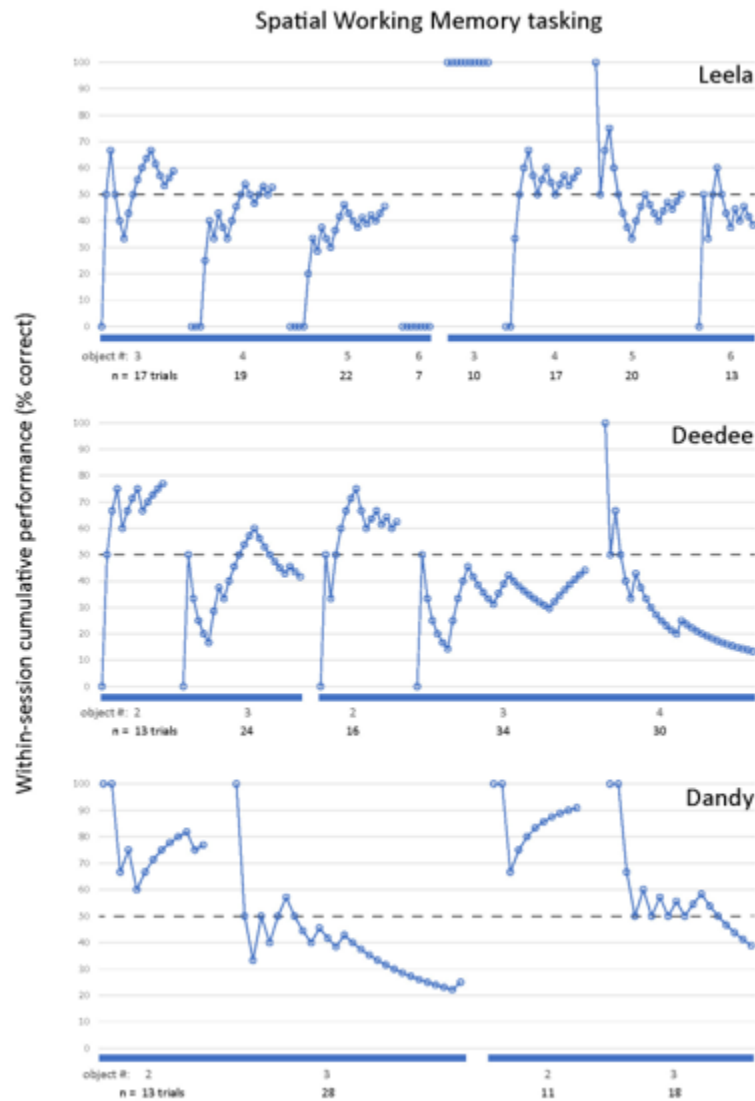
Figure 8

Cumulative Performances for the Concurrent Discrimination (CD) Task, Across Subjects
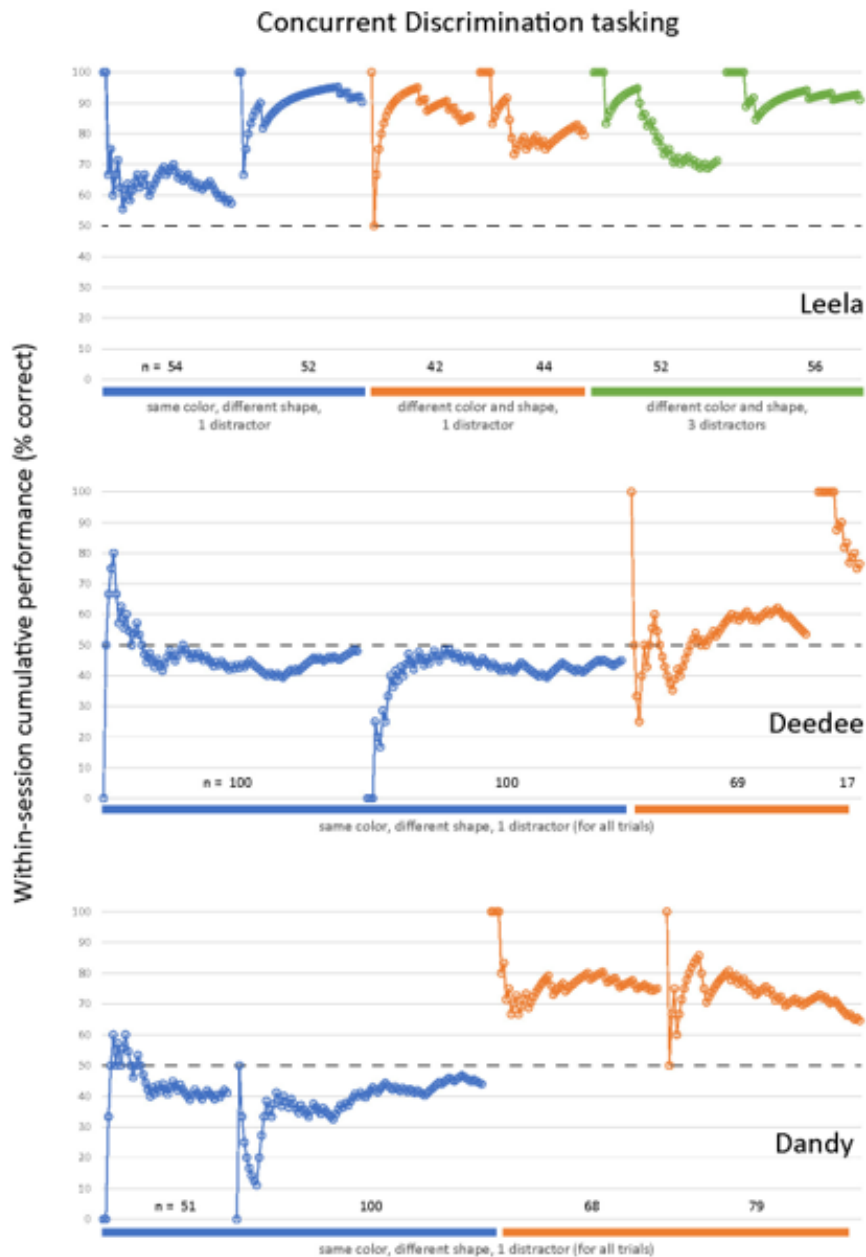
Note. For all cognitive test performance figures (Figures 8-12), discrete testing sessions are represented by solid lines at the bottom of each subject plot, with each color representing a unique testing day for each subject. Essential testing parameters are described at the bottom of each plot; dashed lines in plots represent 50% (chance) performance during testing. Some sessions contained only a single block of identical test trials (represented by a continuous set of data points), while others contained multiple blocks (especially if task difficulty was automatically increased due to subject proficiency). Distinct test blocks are separated by gaps in the data within test sessions and are labeled with the specific criteria for each block. The number of operant trials in each testing block or session is indicated within plots. Data are presented separately for each subject evaluated. Subject Leela completed six blocks of CD trials, over three testing days, while Deedee and Dandy each completed four blocks over two testing days. All subjects achieved proficiency at CD testing of base level difficulty (only two screen stimulus objects, of the same color), with Leela achieving task criterion at two additional levels of CD task difficulty involving a greater number of stimulus objects, or objects of different colors.
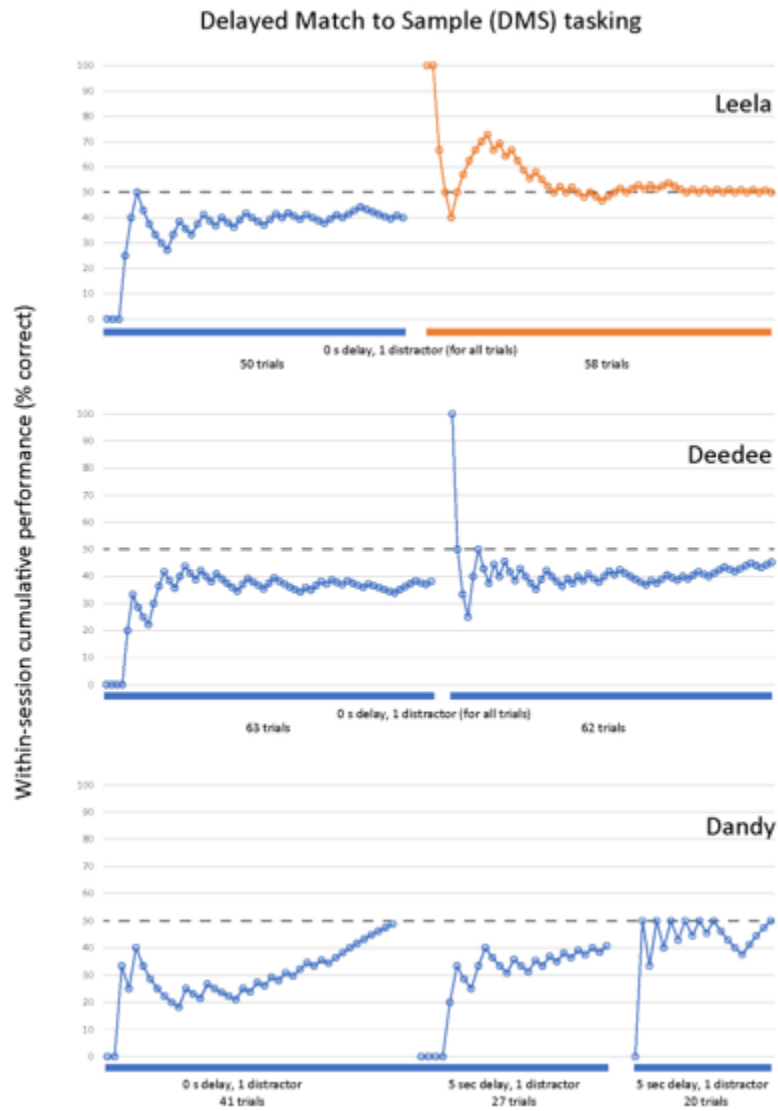
Figure 9

Cumulative Performances for the Spatial Working Memory (SWM) Task, Across Subjects

Note. Figure formatting is as described for Figure 8. Each subject was tested during two separate sessions within one day. Subjects initially were presented with a small number of on-screen stimuli (three for Leela, two for Dandy and Deedee); the number of stimuli increased automatically after 10 correct responses for a given number of stimuli. All subjects achieved proficiency for one of more sets of test stimuli, and all subjects exhibited declining performance as the number of test stimuli increased. Leela achieved both the highest session cumulative performance score (100%) as well as reached the greatest number of stimuli (6). Leela achieved performance criterion with three stimuli, while both Deedee (once) and Dandy (twice) reached criterion with two stimuli.
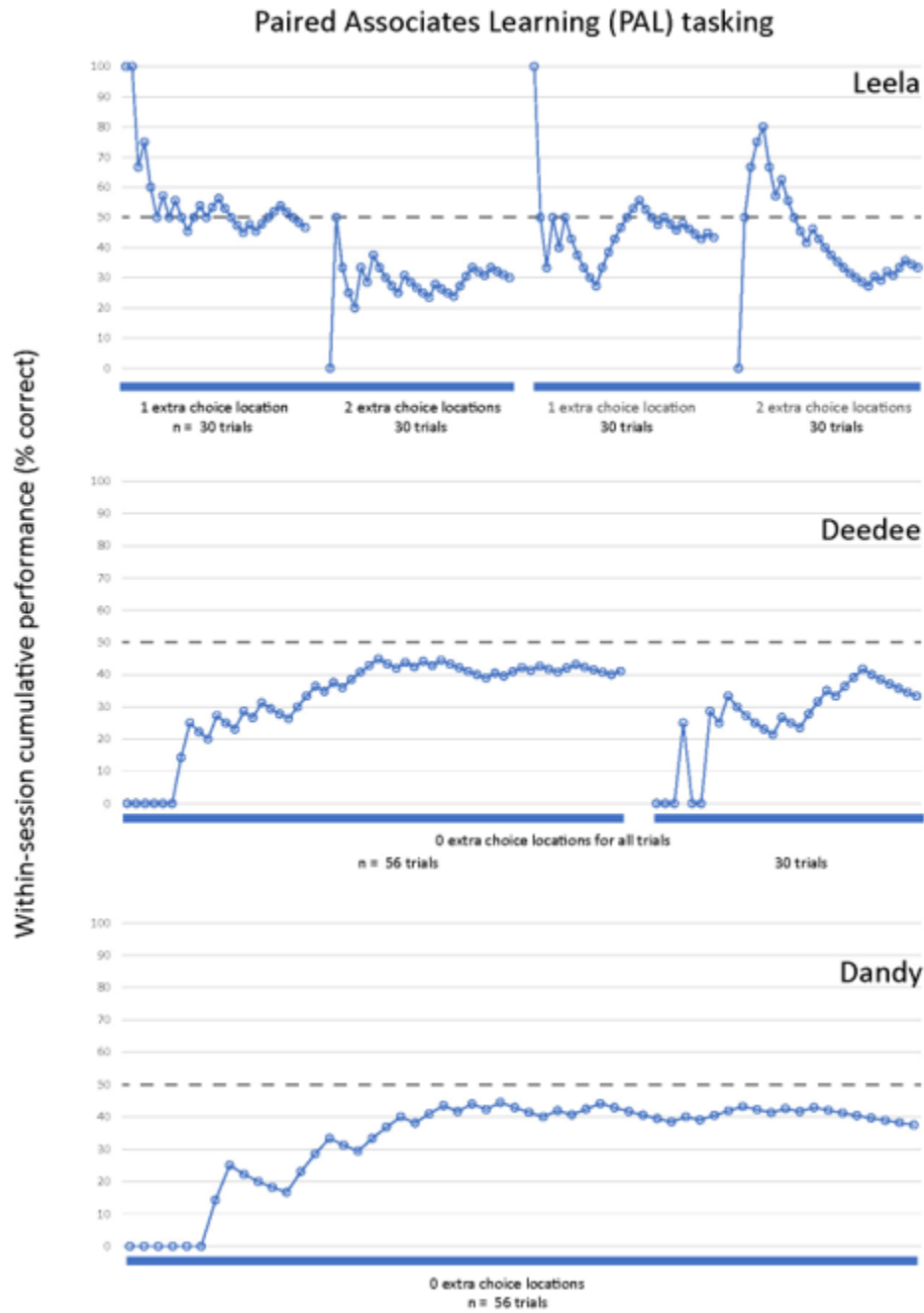
Figure 10

Cumulative Performances for the Delayed Match-To-Sample (DMS) Task, Across Subjects

Note. Figure formatting is as described for Figure 8. Each subject was tested during two separate sessions (on one day for subjects Dandy and Deedee and over two days for subject Leela). Leela exhibited a high level of performance during the middle of day two testing, while Dandy achieved enough successful trials during session 1 of testing that task difficulty was automatically raised (via extension of the inter-stimuli delay duration). No subject achieved end-of session cumulative performance criterion on this task.
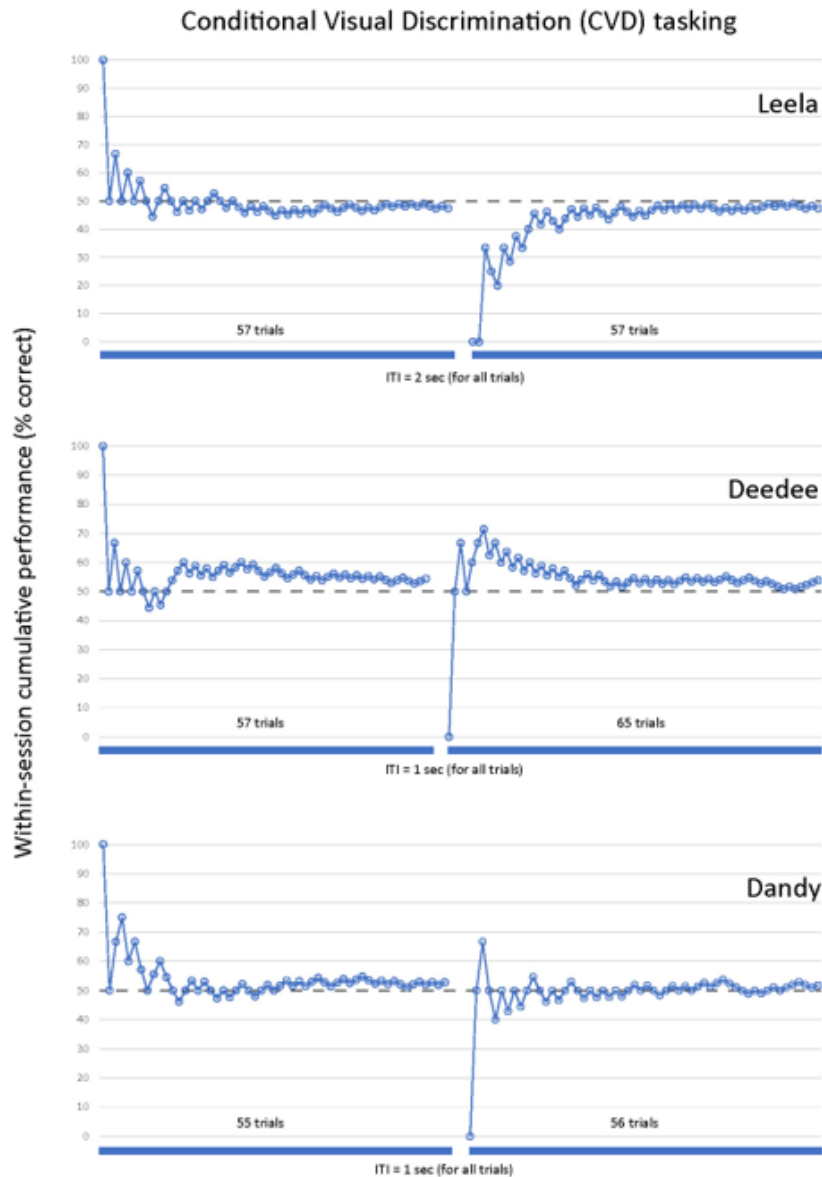
Figure 11

Cumulative Performances for the Paired Associates Learning (PAL) Task, Across Subjects

Note. Figure formatting is as described for Figure 8. Leela and Deedee were tested during two separate sessions on one day, while Dandy experienced only one test session. Leela achieved enough successful trials during both of her training sessions that task difficulty was automatically raised (via addition of extra choice locations). No subject achieved end-of session cumulative performance criterion on this task.

Figure 12

Cumulative Performances for the Conditional Visual Discrimination (CVD) Task, Across Subjects


Note. Figure formatting is as described for Figure 8. Each subject was tested during two separate sessions on a single day. No subject exhibited proficiency at this task.

Rather than representing a broad continuum of cognitive performances, subject performance on the five cognitive tasks used here was largely categorical, with relatively proficient performances at CD and SWM tasks (base difficulty-level average performance scores of 75% and 89%, respectively), and little or no demonstrated capability at DMS, PAL, and CVD tasks, for which the base difficulty-level average performance scores were 51%, 39%, and 37%. A frequency assessment of all subject performance scores across all tasks reveals that the distributions of subject task performance values (both cumulative as well as peak performances) were largely bimodal, with the majority of values lying at or below chance performance, with only those performances associated with CD and SWM tasks lying well above our *a priori* performance threshold of >65% (Figure 13).

Figure 13

Overview of Performance Scores
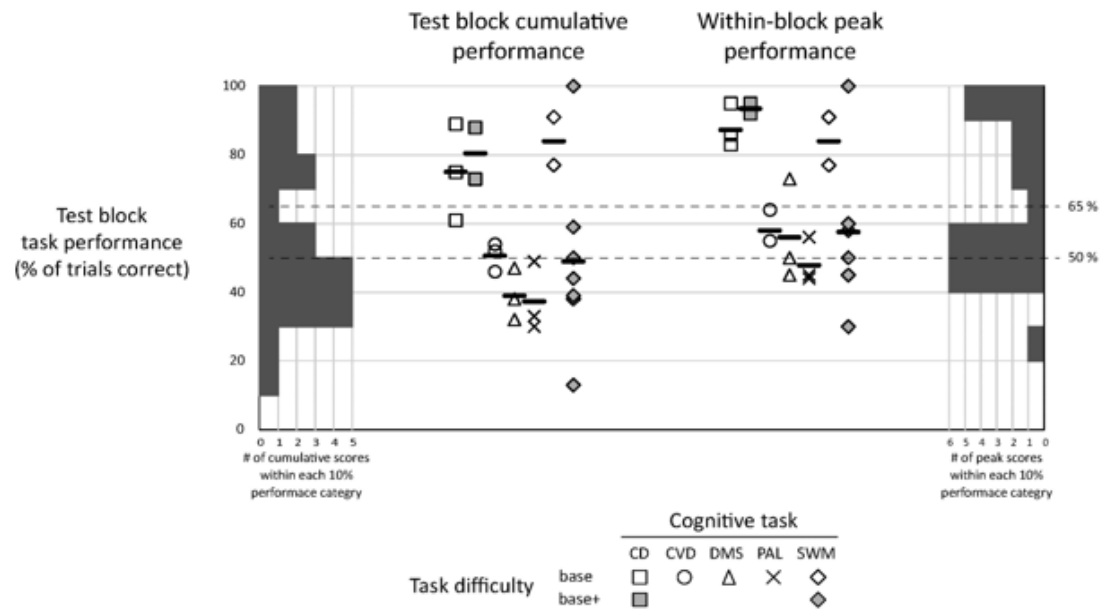
Note. The distributions of subject test performance scores suggest two general categories of subject cognitive task performance. Shown here are cognitive task performance scores from the three test subjects across all five cognitive tasks, representing 23 separate combinations of task identify and task difficulty (task key at figure bottom). Subject cumulative performance (computed as end of test block or test session overall performance for that block or session) is depicted in the left half of the figure, with subject peak performance (computed as subject greatest task performance within a task block or session, after a minimum of 10 operant trials) depicted at figure right. Insert histograms depict tallies of performance scores by decadal performance category. Group means are shown as bars within each column of symbols. Subjects exhibited little proficiency at CVD, DMS, and PAL tasks, and showed more general capability at both CD and SWM tasks. Subject performance scores largely clustered into two groups (those at or below chance level, and those at >70% task performance, with relatively few intermediate performance scores.

In 16 instances during cognitive testing, individual subjects experienced an increase in task difficulty (within the same task type) from one test block to the next. These adjustments included increased delay periods, an increased number of screen icons, or variation in an increased number of icon characteristics (shapes, colors, or positions). Some of these adjustments in task difficulty were automatic, due to subjects achieving a criterion number of correct trial performances within a given test block. In all 16 cases, subject cumulative performance (within task) decreased for the test block following an increase in task difficulty. This effect was seen both in task types in which subjects exhibited proficiency (CD and SWM tasks) as well as in task types not mastered by our subjects (DMS and PAL). For the SWM test in particular, changes in test difficulty were very linear (incremental increases in the number of test stimuli; Figure 9). Across all three test subjects, SWM task performance worsened steadily with an increasing number of stimuli in the SWM task (Pearson $r = -.615$, $n = 17$ test blocks, $p < .01$).

Comparison of End-of-Session Cumulative, Versus Within-Session Peak, Performance Criteria

At multiple points during subject testing, subjects were observed to exhibit a high level of performance during the middle of a testing session, only to then exhibit lower performance during the latter portion of the session, causing their cumulative performance measure to decline from its mid-session peak. Consideration of peak vs. end-of test session (or test block) cumulative performance revealed that within-session peak performance (after a minimum of 10 trials) exceeded end-of-session performance for 20 of the 23 different cognitive test iterations used here (Figure 13). That our data collection continued past subject peak performances allowed us to examine within session changes in performance in some detail.

Twenty-eight distinct testing sessions were conducted (Figures 8-12), and 11 of them would have been terminated earlier had we adopted a strict criterion of peak performance >65% as the rule for session termination. Interestingly, the sessions that could have been subject to early termination were almost entirely (10 out of 11 total) from CD (6 of 7 total) and SWM (4 of 4 total) sessions (Figures 8 and 9). One additional DMS session (Leela, day 2) could have been truncated at peak performance after just 11 trials (Figure 10). In all other test sessions during which peak performance eclipsed end-of-session cumulative performance (17 of 28 test sessions), neither peak nor cumulative performance was > 65% correct performance.

At select times during cognitive testing, individual subjects experienced two test blocks of identical format and difficulty within the same session ($n = 7$ instances), within different testing sessions on a single day ($n = 16$), or on different days ($n = 5$). The degree to which subject performance improved across these different iterations is a potentially useful indicator of the manner in which task comprehension progressed. Subject performance (measured as end-of-test block cumulative performance) improved unevenly across these different intervals. Performance gains across test blocks (average performance gain of 8.42%) and test sessions (7.09%) were of similar, small magnitude, whereas the improvement in subject performance across testing days (average gain of 20.29%) was more considerable.

The youngest test subject (Leela) had the highest cumulative performance rate for four of the five tasks (with CVD being the exception). She was the only subject to perform above a level of four stimuli in SWM tasks (achieving up to 6 stimuli; Figure 9), and was the only subject to perform at a level of three distractor stimuli during CD trials (Figure 8). Older subjects Deedee and Dandy had similar cumulative performance for all tasks except for CD, in which the male subject (Dandy) tested ~14% (M) higher than female Deedee (Table 1). On the basis of end-of session (or test block) cumulative performance, Leela achieved a total of four task proficiencies (Table 1) compared to two for Dandy and one for Deedee. On the basis of peak within-session (or test block) performance, Leela achieved a total of five task proficiencies (>65%) compared to two each for both Dandy and Deedee (Table 1).

Discussion

The siamangs exhibited general comprehension of cognitive tasks CD and SWM, but not of DMS, PAL, or CVD tasks. Thus, our data demonstrate that, with limited training, siamangs can demonstrate a level of comprehension with novel cognitive tasks, suggesting a form of flexible learning abilities, which replicates cognitive trends found in the primate evolutionary lineage. Overall, subject performances were relatively progressive across training and testing sessions, suggesting that individual performances were not spurious. As noted above, subject performance was responsive to task difficulty, in that subject performance consistently decreased following incremental increases in task difficulty, demonstrating that subjects were meaningfully engaged with the testing apparatus and procedure. These data suggest that the cognitive tasks employed were of appropriate difficulty, representing a meaningful and variable degree of cognitive load.

Together, these data indicate that (*i*) touchscreen computer methods can be used successfully for hylobatid cognitive evaluation, (*ii*) hylobatids are more than capable of reaching basic cognitive task comprehension and demonstrate successful learning with limited time/training, and (*iii*) the cognitive capabilities exhibited by the siamangs in this study, while not extensively explored, are at least minimally comparable to those exhibited by other NHP. A more extensive, noninvasive cognitive evaluation of these very understudied apes should be encouraged within zoos, sanctuaries, and *in situ* rehabilitation centers, to support efforts to fully understand how hylobatid cognition relates to primate cognitive evolution and promote better welfare and husbandry practices for organizations housing/rehabilitating hylobatids.

Table 1

Subject Cumulative and Peak Performances (Percentage Correct) During Cognitive Testing

Task[b]	Level[c]	Leela Cumulative[d]	Peak[e]	Dandy Cumulative[d]	Peak[e]	Deedee Cumulative[d]	Peak[e]
CD	Base: 2 objects	90 ****	95 ****	75 ****	86 *	76 *	90 *
	1 distractor of same color but different shape	86 ****	95 ****				
	3 distractors of different color but same shape	91 ****	95 ****				
CVD	(base)	46	55	52	55	54	64
DMS	(base)	47	73 --	32	50	38	45
PAL	(base)	49	56	33	44	30	45
SWM	Base: 2 objects			91 *	91 *	77 --	77 --
	3 objects	100 ***	100 ***	39	58	44	60
	4 objects	59	60			13	30
	5 objects	50	50				
	6 objects	38	45				

Note. Shaded results (values above 65% correct) were evaluated for statistical significance using the binomial test[a]. All subjects exhibited proficiency at CD and SWM tasks in the limited training/testing time that was available, with lesser capability at CVD, DMS, and PAL tasks (see Results). [a] Binomial test significance values represented as – for $p > .05$, * $< .05$, *** for $< .005$, **** for $< .001$. [b] Indicates the specific cognitive task employed. [c] Level of task difficulty; automatically advanced for some tests. [d] The highest end-of-test-session (or training block within a session) cumulative task performance achieved for a given subject at a given task level (normally out of 2 testing sessions). [e] The highest within-session cumulative performance achieved for a given subject at a given task level, after a minimum of 10 trials within a test session

Although our sample size and experimental design do not allow for assessment of the association between subject age and cognitive performance, it is worth noting that the youngest subject (Leela) exhibited better performance than the two older subjects on four of the five tasks used here. She also reached the greatest levels of difficulty in both the CD and SWM tests. Older subjects have been reported to exhibit lower levels of cognitive performance than younger subjects in other studies of NHP cognition (Munger et al., 2017; Nagahara et al., 2010). Whether this trend is due to cognitive decline with age, or due to extrinsic (non-neural) considerations cannot be determined; testing of more subjects over a broader age range would be necessary to provide further insight on this issue. Due to scheduling limitations, subjects in this study were trained and tested during relatively few operant sessions, across a small number of days. As such, our data were not collected in the fashion typical for most operant studies of NHP cognition, which generally assess subject performance over a much longer period of time and a much greater number of training trials, until performance criterion has been achieved (Joly et al., 2014; Leighty et al., 2011; Munger et al., 2017; Nagahara et al., 2010; Rodriguez et al., 2011; Rumbaugh & McCormack, 1967; Taffe & Taffe, 2011; Weed et al., 2008; Zurcher et al., 2010). It is common for NHP subjects to have hundreds to thousands of trials to learn cognitive tasks (Tomasello & Call, 1997), while these subjects were given only ~200 learning trials for a given task before their performance on that task was evaluated. This is an extremely strict learning protocol, representing only a fraction of the time and number of trials typically used in many other studies of NHP cognition (Call, 2001; Leighty et al., 2011; Rumbaugh, 1970). Interestingly, even with limited training, these subjects demonstrated proficient performance across multiple of the cognitive task presented in this study (Figures 8-12; Table 1). As such, we are disinclined to conclude that hylobatids are ultimately incapable of achieving proficiency on the more difficult tasks (DMS, PAL, CVD), as a full assessment of hylobatid capability on these tasks would require more training opportunity than was available here.

Our testing protocol included, in some cases, a subject being tested on a given task at a fixed level of task difficulty multiple times, at repetition intervals of differing length (e.g., twice within the same testing

session, across testing sessions, or across testing days). As described above, improvements in subject task performance were generally greatest *between*, rather than within, testing days. These data suggest that, for hylobatids, single operant sessions per day may be a more productive training/testing design than multiple operant sessions per day.

As noted above, the performance criterion used here as being indicative of subject proficiency at a given cognitive task was a cumulative, end-of-block/session performance level (> 65% correct). This is lower than the 70-80% threshold used in some other studies of NHP cognition, but was chosen here specifically because our test subjects were given a very limited amount of time to acquire these tasks. The subject proficiencies reported here are not attempts to estimate subject or species maximal capabilities at these tasks, but rather to identify those tasks that subjects were quickly able to learn, as a guide to future, more-extensive testing. In that light, several aspects of subject performance suggest that the >65% criterion used here was both appropriate as well as effective in identifying those cognitive tasks that siamangs could quickly acquire. Firstly, adopting a higher performance criterion of 70% produces equivalent results, which suggests that our results are robust relative to this performance criterion. Secondly, as noted above, relatively few individual performance scores were close to our performance criterion threshold, with the majority of scores lying well above, or below, this value (Table 1). This suggests that subject performance was largely bimodal (proficient, or not), and that our performance criterion accurately discriminated between these groups. In addition, a > 65% correct performance value is unlikely to be achieved by chance alone. For example, the 65% threshold lies outside of the $M \pm 2$ SD interval of performance values achieved by subjects on the PAL and CVD tasks. Collectively, these data support the use of a > 65% correct performance criterion in this preliminary study, as a way to identify those cognitive tasks that siamang gibbons could quickly master, versus those which they could not acquire with relatively limited training opportunity. Additionally, nearly all of the performance measures we identified as being greater than 65% correct performance were statistically verified as being unlikely due to chance alone (Table 1), providing further evidence that our 65% criterion for task proficiency was both valid as well as useful in identifying subject competencies and learning capabilities among the cognitive tasks being evaluated.

As described above, most studies evaluating NHP cognitive capabilities test subjects until a specific performance criterion is reached. Typically, it is subject cumulative performance within a testing session that is tracked, until it reaches a maximum at the end of the session, when performance criterion has been achieved and the testing session is concluded (Beran et al., 2016; Cronin et al., 2017; Rodriguez et al., 2011; Rumbaugh et al., 1972; Tomasello & Call, 1997). Our method employed a data collection window that was defined temporally, rather than by performance, such that we were able to track subject task performance within test sessions beyond that which is normally observed (Table 1).

Our method of extended data collection is somewhat conservative, as we identified 9 of 23 test sessions as having reached criterion according to end-of-session (or test block) cumulative performance, whereas the use of within-session peak performance would result in 11 of 23 test sessions being identified as having achieved performance criterion (Table 1). Using peak performance as the metric by which task proficiency was judged would not, however, have changed the general result of subject understanding of CD and SWM tasks relative to DMS, PAL, and CVD tasks.

It can be seen across multiple test types and multiple test subjects that cumulative (within-session) performance often peaked within the middle of testing sessions, rather than at the end. This late-session decline in performance may be due to the fact that our testing sessions typically contained 50 trials per session, which may promote cognitive fatigue or "lack of interest" in hylobatids (Benoit et al., 2019). A variety of factors could cause subject performance to wane over the latter half of testing sessions, including subject 'boredom' due to decreased novelty (Ramirez, 1999; Tomasello & Call, 1997), prolonged attentional demand (Lacreuse et al., 2020; Tomasello & Call, 1997), or decreased subject motivation due to reward satiety (after receiving multiple rewards for correct trial performances; *pers. obs.*). If one considers peak task performance (after a minimum of 10 trials), rather than cumulative within-session task performance, our data suggest that hylobatid cognitive performances on the tasks in this study may be greater than previously suggested. Subject Leela achieved a peak performance rate of 72% correct partway through DMS testing, while subject Deedee achieved 63% correct performance midway through CVD

testing, although our subjects did not collectively perform well at either DMS or CVD tasks. As described above, these data suggest that expanded training/testing opportunities may reveal additional performance capabilities not observed in our initial sample.

Comparable Cognitive Studies and Results

Although no previous studies used the exact testing parameters and methods that we employed in our study, a few studies that used similar parameters exist, allowing us to compare our results on some tasks. In a study by Leighty et al. (2011), eight mandrills (*Mandrillus sphinx*) of mixed sex and ages 6-19 y/o were tested with a simple concurrent discrimination task involving two experiments: color and shape discrimination. The color discrimination task consisted of one correct stimulus and three distractor stimuli, while the shape discrimination task consisted of one correct stimulus and three distractor stimuli. The testing platform was touchscreen-based, and all subjects were physically isolated from conspecifics. In their study, subjects needed an average of 819 trials to reach criterion (>80%) for the color discrimination task and 1,176 trials (on average) to reach criterion (>80%) for the shape discrimination task. Our female subject Leela achieved similar performance here after ca. 300 trials. Comparably, Rumbaugh and Rice (1962) tested three female subjects of three different ape species (*G. gorilla, P. troglodytes, P. pygmaeus*) considered to be of "young" age in a simple discrimination task consisting of one correct and one distractor stimuli, of the same color, but of different shape. Testing was conducted with physical objects and all subjects were visually and physically isolated from other subjects. It took the gorilla 139 trials, the chimpanzee 273, and the orangutan 280 trials to reach criteria (>80%). In our study, Leela reached and maintained task proficiency in less than 100 trials with the same testing parameters (one distractor, same color stimuli) used in Rumbaugh and Rice (1962). Additionally, subject Leela achieved task proficiency under conditions (three distractors, variable in both color and shape) more difficult than those used by Rumbaugh and Rice (1962). With this evidence, one can argue that at least one of our siamang subjects (Leela) tested at the same level or perhaps higher than that of the great apes that participated in the Rumbaugh and Rice (1962) discrimination study.

Additionally, this study may provide direct evidence for siamang long-term memory. Here, CD training and initial testing were conducted with subjects Dandy and Deedee in July 2017. As described above, testing was halted due to environmental disruption (nearby construction) and was not resumed until Dec 2017. During the five-month interim, the siamangs had no interaction with the testing apparatus or GM. However, when testing resumed in Dec 2017, both Dandy and Deedee immediately sat at their testing bench and began to interact with the touchscreen, immediately choosing the same "correct" stimulus that was used previously. Both siamangs ultimately performed above-chance level on the CD task (mean cumulative performances of 71.1% [Dandy] and 61% [Deedee]). The siamangs not only remembered the mechanics of the apparatus but also remembered the specific task and the stimuli associated with that task, after being completely isolated from it for five months. This, in itself, can be considered evidence of higher cognitive ability, specifically pertaining to memory (Hopper, 2017; Martin-Ordas et al., 2010).

Implications for Enrichment and Welfare

Although enrichment is heavily practiced in most accredited facilities housing nonhuman primates, and is considered to be a vital component of animal welfare, often enrichments are limited in form to environmental (physical) and food-based (dietary) supplements (Swaisgood & Shepherdson, 2006). Cognitive enrichment programs, while increasingly common, remain relatively scarce (Clark, 2017; Kim-McCormack et al., 2016). This is despite the fact that those facilities that do utilize cognitive enrichments overwhelmingly report positive benefits from them (Clark, 2011; Clark et al., 2013; Coleman & Maier, 2010; Dolins et al., 2017; Markowitz, 1982). Computer-based, interactive touchscreen cognitive enrichment is becoming very popular for animals in captive environments because of its success rate in decreasing stereotypic behaviors, aggression, and signs of boredom in captive animals (Clark et al., 2013; Coleman & Maier, 2010; Fagot et al., 2014; Swaisgood & Shepherdson, 2006). This is especially true for nonhuman

primates, with particular relevance to great apes (Clark, 2011; Egelkamp et al., 2016; Fagot et al., 2014; Kim-McCormack et al., 2016; Leighty & Fragaszy, 2003; Perdue et al., 2012; Tarou et al., 2004). Unfortunately, as in all other aspects of primate cognition, hylobatids continue to be underrepresented not only in comparative cognition studies, but in cognitive enrichment programs as well (Gronqvist et al., 2013).

We want to emphasize the importance of studying hylobatid behavior and cognition not just for their utility in helping us to understand primate evolutionary history, but also to improve our proficiency in providing stress-reducing and pleasurable forms of enrichment for these and other nonhuman primates. Video recordings of our testing sessions provide strong evidence that these hylobatids readily (and rather eagerly) interact with cognition testing software and touchscreens, in some cases holding onto the test platform to prevent its removal. Not only is this form of cognitive enrichment mentally stimulating to the animals, it is also cost-effective for the facilities using them. Instead of purchasing a stream of novel objects as enrichment (which often imperfectly capture animal attention and interest; Tarou et al., 2004) touchscreens and interactive software are infrequent purchases, require little or-no maintenance, and permit high degrees of flexibility. These devices also give subjects a sense of "control" of their environment which is believed to greatly enhance the effectiveness of cognitive enrichment for animal welfare (Dolins et al., 2017). Our subjects' eagerness to interact with the testing apparatus suggests that cognitively interactive enrichment will positively contribute to hylobatid mental and physical states in captivity, ultimately improving their welfare.

Acknowledgements

References

Abordo, E. (1976). The learning skills of gibbons. *Gibbon and Siamang, 4*, 106–134.

Andrieu J., Penny, S. G., Bouchet, H., Malaivijitnond, S., Reichard, U. H., Zuberbühler, K. (2020). White-handed gibbons discriminate context-specific song compositions. *PeerJ, 8*(e9477), 1–23.

Barelli, C., Boesch, C., Heistermann, M., & Reichard, U. H. (2008). Female white-handed gibbons (*Hylobates lar*) lead group movements and have priority of access to food resources. *Behaviour*, 145, 965–981.

Bartlett, T. Q. (2007). The hylobatidae: Small apes of Asia. In C. J. Campbell, A. Fuentes, K. C. MacKinnon, M. Panger, & S. K. Bearder (Eds.), *Primates in perspective* (pp. 274–289). Oxford University Press.

Bates, L. A., & Byrne, R. W. (2015). Primate social cognition: What we have learned from nonhuman primates and other animals. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology*, *Vol. 1. Attitudes and social cognition* (pp. 47–78). American Psychological Association.

Byrne, R. W. (2004). The manual skills and cognition that lie behind hominid tool use. In A. E. Russon & D. R. Begun (Eds.), *The evolution of thought: Evolutionary origins of great ape intelligence* (pp. 31–44). Cambridge University Press.

Beck, B. (1967). A study of problem solving by gibbons. *Behaviour, 28*(1), 95–109.

Benoit, C. E., Solopchuk, O., Borragán, G., Carbonnelle, A., Van Durme, S., & Zénon, A. (2019). Cognitive task avoidance correlates with fatigue-induced performance decrement but not with subjective fatigue. *Neuropsychologia*, *123*, 30–40.

Beran, M. J., Menzel, C. R., Parrish, A. E., Perdue, B. M., Sayers, K., Smith, J. D., & Washburn, D. A. (2016). Primate cognition: Attention, episodic memory, prospective memory, self-control, and metacognition as examples of cognitive control in nonhuman primates. *Cognitive Science*, *7*(5), 294–316.

Boutan, L. (1913). Le pseudo-langage. *Actes de la Société linnéenne Bordeaux, 67*, 5–80.

Boutan, L. (1914). Les deux méthodes de l'enfant. *Actes de la Société linnéenne Bordeaux, 68*, 3-146.

Brockelman, W., Reichard, U., Treesucon, U., & Raemaekers, J. (1998). Dispersal, pair formation and social structure in gibbons (*Hylobates lar*). *Behavioral Ecology and Sociobiology, 42*(5), 329–339.

Call, J. (2001). Object permanence in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and children (*Homo sapiens*). *Journal of Comparative Psychology, 115*(2), 159–171.

Call J. (2010). Do apes know that they could be wrong? *Animal Cognition*, *13*(5), 689–700.

Cambridge Cognition. (2016). CANTAB test selector, choose the right test for your research. Retrieved from https://www.cambridgecognition.com/resources/cantab-test-selector.

Cardinal, R., & Aitken, M. (2010). Whisker: A client–server high-performance multimedia research control system. *Behavior Research Methods & Instrumentation, 42*, 1059–1071.

Caspar, K. R., Mader, L., Pallasdies, F., Lindenmeier, M., Begall, S. (2018). Captive gibbons (*Hylobatidae*) use different referential cues in an object-choice task: insights into lesser ape cognition and manual laterality. *PeerJ*, 6(e5348), 1–27.

Chivers, D. (1974). *The siamang in Malaya. A field study of a primate in tropical rain forest. Contributions to Primatology* (Vol. 1). Karger.

Chivers, D. (1976). Communication within and between family groups of siamang (*Symphalangus syndactylus*). *Behaviour, 57*(1), 116–135.

Clark, F. (2011). Great ape cognition and captive care: Can cognitive challenges enhance well-being? *Applied Animal Behaviour Science, 135*(1), 1–12.

Clark, F. (2017). Cognitive enrichment and welfare: Current approaches and future directions. *Animal Behavior and Cognition, 4*(1), 52-71.

Clark, F., Davies, S., Madigan, A., Warner, A., & Kuczaj, S. (2013). Cognitive enrichment for bottlenose dolphins (*Tursiops truncatus*): Evaluation of a novel underwater maze device. *Zoo Biology, 32*(6), 608–619.

Coleman, K., & Maier, A. (2010). The use of positive reinforcement training to reduce stereotypic behavior in rhesus macaques. *Applied Animal Behaviour Science, 124*(3),142–148.

Crofts, H., Muggleton, N., Bowditch, A., Pearce, P., Nutt, D., & Scott, E. (1999). Home cage presentation of complex discrimination tasks to marmosets and rhesus monkeys. *Laboratory Animals, 33*(3), 207–214.

Cronin, K. A., Jacobson, S. L., Bonnie, K. E., & Hopper, L. M. (2017). Studying primate cognition in a social setting to improve validity and welfare: A literature review highlighting successful approaches. *PeerJ, 5* (3649).

Cunningham, C., Anderson, J., & Mootnick, A. (2006). Object manipulation to obtain a food reward in hoolock gibbons, *Bunopithecus hoolock*. *Animal Behaviour, 71*(3), 621–629.

Cunningham, C., Anderson, J., & Mootnick, A. (2011). A sex difference in effect of prior experience on object-mediated problem-solving in gibbons. *Animal Cognition, 14*(4), 599–605.

Cunningham C. L., Anderson J. R., Mootnick A. R. (2016). The evolution of technical intelligence: Perspectives from the Hylobatidae. In U. H. Reichard, C. Barelli, H. Hirai, & M. G. Nowak (Eds.), *Evolution of gibbons and siamag* (pp. 291-311). Springer.

Damerius, L. A., Burkart, J. M., van Noordwijk, M. A., Haun, D. B., Kosonen, Z. K., Galdikas, B. M., Saraswati, Y., Kurniawan, D., & van Schaik, C. P. (2019). General cognitive abilities in orangutans (*Pongo abelii* and *Pongo pygmaeus*). *Intelligence*, *74*, 3-11.

DeLuca, C. R., Wood, S. J., Anderson, V., Buchanan, J., Profitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the Cantab. I: Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology*, *25*(2), 242–254.

Dolins, F., Schweller, K., & Milne, S. (2017). Technology advancing the study of animal cognition: Using virtual reality to present virtually simulated environments to investigate nonhuman primate spatial cognition. *Current Zoology, 63*(1), 97–108.

Dunbar, R. I. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, *32*(1), 163-181.

Egelkamp, C., Hopper, L., Cronin, K., Jacobson, S., & Ross, S. (2016). Using touchscreens to explore the welfare and cognition of zoo-housed primates. *PeerJ Preprints*, *4*(2312), 2167–9843.

Egelkamp, C. L., & Ross, S. R. (2018). A review of zoo-based cognitive research using touchscreen interfaces. *Zoo Biology*, *38*(2), 220-235.

Fagot, J., Gullstrand, J., Kemp, C., Defilles, C., & Mekaouche, M. (2014). Effects of freely accessible computerized test systems on the spontaneous behaviors and stress level of Guinea baboons (*Papio papio*). *American Journal of Primatology, 76*(1), 56–64.

Fedor, A., Skollár, G., Szerencsy, N., & Ujhelyi, M. (2008). Object permanence tests on gibbons (*Hylobatidae*). *Journal of Comparative Psychology, 122*(4), 403–417.

Freeberg, T. M., Dunbar, R. I., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society of London. 367*(1597), 1785–1801.

Gronqvist, G., Kingston-Jones, M., May, A., & Lehmann, J. (2013). The effects of three types of environmental enrichment on the behaviour of captive Javan gibbons (*Hylobates moloch*). *Applied Animal Behaviour Science, 147*(1), 214–223.

Heschl, A. & Fuchsbichler, C. (2009). Siamangs (*Hylobates syndactylus*) recognize their mirror image. *International Journal of Comparative Psychology*, *22*(4), 221-233.

Hill, A., Collier-Baker, E., & Suddendorf, T. (2011). Inferential reasoning by exclusion in great apes, lesser apes, and spider monkeys. *Journal of Comparative Psychology, 125*(1), 91–103.

Hill, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*, *30*(1), 3-41.

Hood, K. L., Postle, B. R., & Corkin, S. (1999). An evaluation of the concurrent discrimination task as a measure of habit learning: Performance of amnesic subjects. *Neuropsychologia*, *37*(12), 1375–1386.

Hopper, L. M. (2017). Cognitive research in zoos. *Current Opinion in Behavioral Sciences*, *16*, 100–110.

Hopper, L. M., Allritz, M., Egelkamp, C. L., Huskisson, S. M., Jacobson, S. L., Leinwand, J. G., & Ross, S.R. (2021). A comparative perspective on three primate species' responses to a pictorial emotional stroop task. *Animals*, *11*(3), 1–22.

Hopper, L. M., Egelkamp, C. L., Fidino, M., & Ross, S. R. (2019). An assessment of touchscreens for testing primate food preferences and valuations. *Behavior Research Methods*, *51*(2), 639–650.

Inoue, Y., & Inoue, E. (2002). *The trap-table problem with a young white-handed gibbon (Hylobates lar)* [Paper presentation]. Abstract of the joint symposium of COE2/SAGA5; Evolution of apes and the origins of human beings. Inuyama, Japan.

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), R1004–R1005.

Jerison, H. (1973). *Evolution of the brain and intelligence*. Academic Press.

Jerison, H. (1982). Allometry, brain size, cortical surface, and convolutedness. In E. Armstrong & D. Falk (Eds.), *Primate brain evolution* (pp. 77-84). Springer.

Joly, M., Ammersdorfer, S., Schmidtke, D., & Zimmermann, E. (2014). Touchscreen-based cognitive tasks reveal age-related impairment in a primate aging model, the grey mouse lemur (*Microcebus murinus*). *PLoS One, 9*(10), 109–393.

Kim-McCormack, N., Smith, C., & Behie, A. (2016). Is interactive technology a relevant and effective enrichment for captive great apes? *Applied Animal Behaviour Science, 185*, 1–8.

Köhler, W. (1925). *The mentality of apes*. Routledge & Kegan Paul, Ltd.

Lacreuse, A., Raz, N., Schmidtke, D., Hopkins, W. D., & Herndon, J. G. (2020). Age-related decline in executive function as a hallmark of cognitive ageing in primates: An overview of cognitive and neurobiological studies. *Philosophical Transactions of the Royal Society B*, *375*(1811), 20190618.

Leighton, D. R. (1987). Gibbons: Territoriality and monogamy. In B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham, & T. T. Struhsaker (Eds.), *Primate societies* (pp. 135-145). University of Chicago Press.

Leighty, K., & Fragaszy, D. (2003). Primates in cyberspace: Using interactive computer tasks to study perception and action in nonhuman animals. *Animal Cognition, 6*(3), 137–139.

Leighty, K., Maloney, M., Kuhar, C., Phillips, R., Wild, J., Chaplin, M., & Betting, T. (2011). Use of a touchscreen-mediated testing system with mandrill monkeys. *International Journal of Comparative Psychology, 24*(1).

Liebal, K. (2016). Communication and cognition of small apes. In U. H. Reichard, H. Hirai, & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 313-347). Springer.

Martin-Ordas, G., Haun, D., Colmenares, F., & Call, J. (2010). Keeping track of time: Evidence for episodic-like memory in great apes. *Animal Cognition*, *13*(2), 331–340.

Markowitz, H. (1982). *Behavioral enrichment in the zoo*. Van Nostrand Reinhold.

Morimura, N., & Matsuzawa, T. (2001). Memory of movies by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *115*(2), 152–158.

Munger, E., Takemoto, A., Raghanti, M. A., & Nakamura, K. (2017). Visual discrimination and reversal learning in aged common marmosets (*Callithrix jacchus*). *Neuroscience Research, 124*, 5–62.

Nagahara, A., Bernot, T., & Tuszynski, M. (2010). Age-related cognitive deficits in rhesus monkeys mirror human deficits on an automated test battery. *Neurobiology of Aging, 31*(6), 1020-1031.

Osvath, M., & Martin-Ordas, G. (2014). The future of future-oriented cognition in non-humans: Theory and the empirical case of the great apes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130486.

Palombit, R. (1994). Extra-pair copulations in a monogamous ape. *Animal Behaviour, 47*(3), 721–723.

Pearce, J. (2013). *Animal learning and cognition: An introduction.* Psychology Press.

Perdue, B., Clay, A., Gaalema, D., Maple, T., & Stoinski, T. (2012). Technology at the zoo: The influence of a touchscreen computer on orangutans and zoo visitors. *Zoo Biology, 31*(1), 27–39.

Ramirez, K. (1999) *Animal training: Successful animal management through positive reinforcement*. Shedd Aquarium Society.

Reichard, U. H., & Barelli, C. (2008). Life history and reproductive strategies of Khao Yai *Hylobates lar*: Implications for social evolution in apes. *International Journal of Primatology*, *29*(4), 823–844.

Reichard, U. H., Barelli, C., Hirai, H., & Nowak, M. (1995). Extra-pair copulations in a monogamous gibbon (*Hylobates lar*). *Ethology, 100*(2), 99–112.

Reichard, U. H., Barelli, C., Hirai, H., & Nowak, M. G. (2016). The evolution of gibbons and siamang. In U. H. Reichard, H. Hirai, & C. Barelli (Eds.), *Evolution of gibbons and siamang* (pp. 3-41). Springer.

Rilling, J. K., & Insel, T. R. (1999). The primate neocortex in comparative perspective using magnetic resonance imaging. *Journal of Human Evolution*, *37*(2), 191–223.

Rilling, J., & Seligman, R. (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *Journal of Human Evolution, 42*(5), 505–533.

Rodriguez, J., Zürcher, N., Bartlett, T., Nathanielsz, P., & Nijland, M. (2011). CANTAB delayed matching to sample task performance in juvenile baboons. *Journal of Neuroscience Methods, 196*(2), 258–263.

Rumbaugh, D. (1970). Learning skills of anthropoids. *Primate Behavior: Developments in Field and Laboratory Research, 1*, 1–70. Rumbaugh, D., Bell, C., & Gill, T. (1972). Two discrimination test apparatuses for primates. *Behavior Research Methods & Instrumentation, 4*(1), 6–10.

Rumbaugh, D., & McCormack, C. (1967). The learning skills of primates: A comparative study of apes and monkeys. D. Starck, R. Schneider, and H. J. Kuhn (Eds.), *Progress in primatology* (pp. 289-306). Verlag.

Rumbaugh, D. M., & Rice, C. P. (1962). Learning-set formation in young great apes. *Journal of Comparative and Physiological Psychology, 55*(5), 866-868.

Ryan, A. M., Freeman, S. M., Murai, T., Lau, A. R., Palumbo, M. C., Hogrefe, C. E., Bales, K.L., & Bauman, M. D. (2019). Non-invasive eye tracking methods for New World and Old World monkeys. *Frontiers in Behavioral Neuroscience*, *13*(39), 1–10.

Semendeferi, K., Lu, A., Schenker, N., & Damásio, H. (2002). Humans and great apes share a large frontal cortex. *Nature Neuroscience, 5*(3), 272.

Smith, R., & Jungers, W. (1997). Body mass in comparative primatology. *Journal of Human Evolution*, *32*(6), 523–559.

Swaisgood, R., & Shepherdson, D. (2006). Environmental enrichment as a strategy for mitigating stereotypies in zoo animals: A literature review and meta-analysis. *Stereotypic Animal behaviour: Fundamentals and Applications to Welfare, 2*, 256-285.

Taffe, M., & Taffe, W. (2011). Rhesus monkeys employ a procedural strategy to reduce working memory load in a self-ordered spatial search task. *Brain Research, 1413*, 43–50.

Tarou, L., Kuhar, C., Adcock, D., Bloomsmith, M., & Maple, T. (2004). Computer-assisted enrichment for zoo-housed orangutans (*Pongo pygmaeus*). *Animal Welfare, 13*(4), 445–453.

Thorndike, E. (1911). *Animal intelligence: Experimental studies*. The Macmillan Company.

Tobias, P. (1968). Cranial capacity in anthropoid apes, *Australopithecus* and *Homo habilis*, with comments on skewed samples. *South African Journal of Science*, *64*(2), 81–91.

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University Press.

Ujhelyi, M., Merker, B., Buk, P., & Geissmann, T. (2000). Observations on the behavior of gibbons (*Hylobates leucogenys, H. gabriellae, and H. lar*) in the presence of mirrors. *Journal of Comparative Psychology, 114*(3), 253–262.

Vogel, E. R., Haag, L., Mitra-Setia, T., van Schaik, C. P., & Dominy, N. J. (2009). Foraging and ranging behavior during a fallback episode: *Hylobates albibarbis* and *Pongo pygmaeus wurmbii* compared. *American Journal of Physical Anthropology, 140*(4), 716–726.

Vonk, J. (2013). Matching based on biological categories in orangutans (*Pongo abelii*) and a gorilla (*Gorilla gorilla gorilla*). *PeerJ, 1*(e158), 1–25.

Vonk, J., Jett, S. E., Mosteller, K. W., & Galvan, M. (2013). Natural category discrimination in chimpanzees (*Pan troglodytes*) at three levels of abstraction. *Learning & Behavior*, *41*(3), 271–284.

Vonk, J., & MacDonald, S. E. (2004). Levels of abstraction in orangutan (*Pongo abelii*) categorization. *Journal of Comparative Psychology*, *118*(1), 3–13.

Vonk, J., & Rastogi, G. (2019). The "sh-ape bias" in non-linguistic categorization: Comparisons between children and other apes. *Journal of Cognition and Development*, *20*(3), 380–398.

Wagner, K. E., Hopper, L. M., & Ross, S. R. (2016). Asymmetries in the production of self-directed behavior by chimpanzees and gorillas during a computerized cognitive test. *Animal Cognition*, *19*(2), 343–350.

Weed, M., Bryant, R., & Perry, S. (2008). Cognitive development in macaques: Attentional set-shifting in juvenile and adult rhesus monkeys. *Neuroscience, 157*(1), 22–28.

Wobber, V., Herrmann, E., Hare, B., Wrangham, R., & Tomasello, M. (2014). Differences in the early cognitive development of children and great apes. *Developmental psychobiology*, *56*(3), 547–573.

Yerkes, R., & Yerkes, A. (1929). *The great apes: A study of anthropoid life.* Yale University Press.

Zilles, K., & Rehkämper, G. (1988). The initial brain concept: A work in progress. *Behavioral and Brain Sciences*, *11*(1), 105–106.

Zuberbühler, K., & Janmaat, K. (2010). Foraging cognition in non-human primates. In M. Platt & A. Ghazanfar (Eds.), *Primate neuroethology* (pp. 64-83). Oxford Academic Press.

Zurcher, N., Rodriguez, J., Jenkins, S., Keenan, K., Bartlett, T., McDonald, T., & Nijland, M. (2010). Performance of juvenile baboons on neuropsychological tests assessing associative learning, motivation and attention. *Journal Neuroscience Methods, 188*(2), 219–225.