



More Rope Tricks Reveal Why More Task Variants Will Never Lead to Strong Inferences About Higher-Order Causal Reasoning in Chimpanzees

Daniel J. Povinelli* and Ty Henley

Department of Biology, University of Louisiana

*Corresponding author (Email: povinelli@louisiana.edu)

Citation – Povinelli, D. J., & Henley, T. (2020). More rope tricks reveal why more task variants will never lead to strong inferences about higher-order causal reasoning in chimpanzees. *Animal Behavior and Cognition*, 7(3), 392-418. doi: <https://doi.org/10.26451/abc.07.03.08.2020>

Abstract – When chimpanzees (and other animals) use tools to pound, crack open, retrieve, soak up, pry apart, probe into, and/or dig up other objects (just to name a few of the operations of which they are capable), are these actions modulated by higher-order, structural, role-based representations of <weight>, <force>, <shape>, <connection> and so forth? We report a study that was designed to shed light on what chimpanzees understand about <intrinsic connection> and <transfer of force> between objects. We presented expert tool-using chimpanzees with very familiar objects (ropes) in a very familiar testing context (hooking and retrieving an object containing food). We investigated whether they could transfer the known operation (hooking and pulling) to two distinct setups involving ropes (a looped rope that could effectuate the transfer of force, and a draped length of rope that could not—unless both ends were grasped at the same time). We show that first-order, perceptually-based relational reasoning is both necessary and sufficient to explain not only the impressive pattern of results we did obtain, but any possible pattern of results we could have obtained. If correct, any claim that higher-order reasoning is necessary to explain the results of such a test would, by definition, be false. More importantly, we show that this is not an idiosyncratic limitation of this experiment. Instead, we show that all experimental protocols claiming to assay higher-order reasoning in animals rest upon an extremely suspect, and ultimately unprincipled, titration. Specifically, researchers implicitly assume that the tasks are perceptually *similar enough* to what their subjects have previously encountered to allow them make sense of the problem, but perceptually *different enough* that the subjects must (somehow) rely upon higher-order reasoning to navigate their way through it. We show why such assumptions are false.

Keywords – Causal cognition, Chimpanzees, Tool use, Physical cognition

For roughly two thousand years, humans have known (or at least strongly suspected) that other species make and use tools (Newmyer, 2005), a fact imperfectly mirrored in animal folk tales from around the world (Povinelli & Barker, 2019). One of the earliest systematic efforts to collate data on animal tool use in both natural and captive settings was conducted by Darwin (1871), whose efforts were quickly followed by several experimental investigations of animal tool use and manufacture (e.g., Hobhouse, 1915; Klüver, 1933; Köhler, 1925). Köhler's (1925) studies documenting the ability of chimpanzees to use boxes, sticks, strings, and other objects as tools, in particular, became widely known. In the century that has followed, the interest in animal tool use burgeoned, catapulted forward by Jane Goodall's detailed reports of wild chimpanzees using (and making) several distinct kinds of tools (e.g., twigs as termite fishing wands, chewed up leaves as sponges, rocks as missiles; van Lawick-Goodall,

1968). An impressive list of species, from a wide array of taxonomic groups, are now documented tool users and toolmakers, including insects, fish, birds, and mammals, even marine invertebrates (see reviews and catalogues by Beck, 1980; Bentley-Condit & Smith, 2010; Mann & Patterson, 2013; Shumaker et al., 2011). Claims of new species purportedly using tools, as well as new uses of tools by species already known for the ability, are regularly reported (e.g., nuthatches: Rutz & Deans, 2018; chimpanzees: Hicks et al., 2019; van Leeuwen et al., 2017; puffins: Fayet et al., 2020). Nonetheless, tool use remains rare, possibly occurring in less than 1% of all known species (see Hunt et al., 2013).

Many reports of animal tool use in the wild contain analyses restricted to a functional level of description (what the animals do with tools along with analyses of their ecological and evolutionary utility). Such reports often emphasize the flexibility, novelty, or context-dependency of tool use (e.g., Boesch & Boesch-Acherman, 2000; Patterson & Mann, 2011; Sanz & Morgan, 2009). An important historical shift, however, emerged with Visalberghi's landmark investigations of tool use in captive capuchin monkeys (e.g., Visalberghi & Limongelli, 1994; Visalberghi & Trinca, 1989). These investigations were aimed at distinguishing the kinds of causal information animals rely upon when using tools. Although interest in such representational-level questions had existed since Köhler (1925), and probably earlier, Visalberghi's work spawned renewed interest in how tests involving tools could be used to probe the causal reasoning of animals. Our own initial work on this topic, *Folk Physics for Apes*, was largely inspired by her work (see Povinelli, 2000). A quarter century later, a large (and growing) group of investigators have deployed a large (and growing) tool kit of tool-using tasks with the putative aim of distinguishing between the construct of "associative learning" and what researchers variously describe as "complex cognition", "the ability to engage in causal reasoning", and/or "the ability to process or recognize causal information/regularities/cues", just to mention a few of the phrases used (e.g., Alem et al., 2016; Bird & Emery, 2009b; Herrmann et al., 2008; Holzhaider et al., 2008; Horner & Whiten, 2007; Jelbert et al., 2014; Martin-Ordas et al., 2008; Mulcahy et al., 2005; Nissani, 2006; Riemer et al., 2014; Santos et al., 2005; Seed et al., 2009; Taylor et al., 2009, 2010; Visalberghi et al., 2009; Weir & Kacelnik, 2006; Whitt et al., 2009).

In an effort to rectify the conflation between functional- and representational-level questions surrounding tool use (and other intelligent behaviors) in animals, Penn et al. (2008a) offered their relational reinterpretation hypothesis (RRH).¹ The RRH provides a representational-level account of the functional differences between human and animal cognition by distinguishing between the cognitive requirements for first-order, perceptually-based relational reasoning versus higher-order, structural, role-governed relational reasoning. The model begins with the uncontroversial claim that animals possess rich, temporally stable, mental representations of those aspects of the physical (and social) world that are evolutionarily and developmentally important to them (i.e., first-order, perceptually-based mental representations). Abundant empirical evidence supports the further claim that these representations are part of a cognitive architecture that also includes a (weak) form of compositionality (that is, the representations can be recombined in innumerable ways that endow organisms with the ability to engage in intelligent, goal-directed behavior). New research is regularly published reinforcing these (and related) ecumenical premises (e.g., Gruber et al., 2019). By definition, however, first-order, perceptually-based, relational reasoning can be analytically distinguished from higher-order, structural, role-based, aspects of cognition that humans, at least, manifestly wield. Penn et al. (2008a) identified key aspects of this higher-order system, including the ability to do all of the following in an explicit, structural manner: (1) to represent and keep track of the relations between types (e.g., kinds, classes, roles, variables) and tokens (e.g., individuals, instances, fillers, values), (2) to represent structural relations (e.g., hierarchical relations), and (3) to represent that certain relations necessarily imply others, independent of domain or learning history (for details, see Penn et al., 2008a, p. 123-128). Importantly, these higher-order features of human cognition are what allow humans to redescribe perceptually-based categories and relations such

¹ An earlier formulation of this idea (simply, the reinterpretation hypothesis), was described in numerous publications from our group. The primary concern of the original reinterpretation hypothesis was to marshal resources from evolutionary theory and philosophy of mind to ground the possibility that higher-order thinking is unique to humans (e.g., Povinelli & Giambone, 1999).

that they can serve as the fodder for non-perceptually-based categories and relations involving constructs such as <weight>, <mass>, <shape>, <color>, <time>, <force>, <gravity>, <mental states>, etc.²

To be sure, animals successfully cope (in varying ways) with the detectable consequences of worldly relations that humans, at least, can redescribe via higher-order thinking. But the fact that humans can conceive of a variety of observable events under the structural, role-based relation of <gravity>, for example, and the fact that animals keep track of (and act upon) the spatial trajectories of unsupported objects, in no way implies that animals also represent <gravity>. In summarizing Alan Turing's greatest contribution to the cognitive sciences, Dennett (2009) made this latter point forcefully: "In order to be a perfect and beautiful computing machine, it is not requisite to know what arithmetic is" (p. 10061). In contrast, and by necessity, higher-order, structural, role-based representations do entail the presence of first-order, perceptually-based representations. This, in turn, entails an asymmetric dependency relationship between the evidence for two systems: strong evidence for higher-order representations in a cognitive system is strong evidence for first-order, perceptual-based representations, but not vice versa. The challenge, then, is to determine what kind of evidence warrants a strong inference for the presence/operation of the higher-order system. In the case of reasoning about physical objects, it may seem tempting to think that certain forms of tool use and tool-making require higher-order reasoning or, to put it more directly, that higher-order reasoning is necessary (but not sufficient) for a certain class of behaviors involving tools. Our central objective is to explore whether such an inference is warranted within the current experimental genre of comparative psychology.

In this paper, we report a previously unpublished study with our chimpanzees that built upon their previous experiences using ropes as tools (see below). It was conducted eight years after completion of the studies reported in *Folk Physics for Apes* (Povinelli, 2000).³ The study is a variant of what Jacobs and Osvath (2015) have described as "[o]ne of the most widely used and well-known experimental paradigms in comparative psychology" (p. 89), the so-called string pulling problem. In the most basic version of this test, animals are required to pull a string, rope or similar object in order to obtain a reward that is otherwise too far away. In an exhaustive review, Jacobs and Osvath trace the task's origins from ancient, sensationalistic interest in watching birds (and other animals) pull strings to earn food and water, to its use in comparative psychology, first as a measure of the speed of learning in different species, and later as a means of exploring sensorimotor intelligence and causal reasoning. They document over 200 studies, involving over 160 species, that have deployed "countless methodological variations" of the basic task (Jacobs & Osvath, 2015, p. 89).

Using a novel variant⁴ of the rope-pulling problem, we investigated whether our expert tool-using chimpanzees could apply knowledge from several known tool operations (including using a hook tool and pulling on ropes) to two new operations: (1) using a looped rope to effectuate the transfer of force, and (2) using a draped length of rope that could not transfer force, unless both ends were grasped at the same time. The second problem, in particular, was of great interest to us. We discuss the results in the context of whether this task (or any other existing task) can allow for strong inferences about the presence of higher-order, role-based relational reasoning in animals.

² For example, Povinelli (2012) offers an elaborate case-study of how perceptually-based relations involving object size, the effort while lifting an object, the deformation that results when one object is set upon another, the sound two objects make when they collide, etc., are used as the inputs for a higher-order, structural, role-based reasoning system that keeps track of, and reasons over, the higher-order, structural, role-based construct of <weight>.

³ One reviewer was worried that some readers might suspect that we delayed publication of this manuscript for perverse reasons. To any such reader: rest assured we have dozens of unpublished studies that have remained unpublished for a complex combination of personal idiosyncrasies (e.g., boredom at repeating oneself) and theoretical concerns. The theoretical concerns are fleshed out in Appendix 1.

⁴ As a part of a procedure designed to engineer cooperation between pairs of chimpanzees, Melis et al., (2006) fed a rope through rings on opposite ends of a long platform so that if an animal pulled one end, the other end would move out of reach of the partner, unless the partner pulled at the same time. Although the authors do not state this explicitly (it was not the purpose of their study), some of the apes appear to have been individually trained to pull both ends of the rope simultaneously. Again, because it was not the focus of their study, no data were reported on how many apes learned this feat, or how rapidly they did so.

Method

Subjects, Housing, Previous Experience

The participants of this study were seven adult chimpanzees (*Pan troglodytes*), one male, six females. All of the apes were born in captivity within a year of each other at the University of Louisiana and were raised together in a nursery with other chimpanzees. We selected them to participate in a cognitive and behavioral research program (Project Megan) when they were approximately two years of age. After they reached the age of four, they were moved to a specialized housing and testing compound (see details below) where they lived together in a capacious indoor-outdoor enclosure. The compound was enriched with perches, swings, burlap sacks, hay, and a wide variety of other objects and toys. Social interactions with each other and their human caretakers were nearly constant throughout the day. At the time of the current study, the chimpanzees were between the ages of 19 years, 3 months (19;3) and 20;4. Additional details about the rearing and testing histories of these special and beloved chimpanzees can be found elsewhere (Povinelli, 2000, 2012).

For seventeen years prior to this study, these apes had participated in hundreds of cognitive and behavioral studies, both published and unpublished. These studies were typically conducted in a secondary facility connected to their living compound, consisting of an outdoor waiting area and indoor testing unit, which, in turn, was divided into a transparent Lexan enclosure for the chimpanzees and a human workspace (see Povinelli, 2000, p. 18). Each chimpanzee had been trained (on request) to leave the main compound and enter the waiting area for individual testing. The apes typically remained in the outdoor waiting area as experimenters set up trials indoors. An opaque shuttle door connected the outdoor waiting area to the indoor testing unit. The shuttle door could be opened remotely, allowing the apes to enter and participate from inside their Lexan enclosure. The Lexan divider contained multiple openings (hereafter, windows) that allowed the animals to reach through and perform behavioral tasks.

These chimpanzees had grown highly proficient in using a wide range of tools to accomplish tasks involving hooking, poking, pushing, pulling, raking, lifting, opening, sliding, inserting, rolling, colliding, crushing, cracking, flipping, hanging, swinging, and other operations. Their expertise spanned a broad variety of tool types, the most relevant to the current study being hooks, rakes, and ropes (for examples with hooks and rakes, see Povinelli, 2000, Exp. 3-10, 15-16, 19-20, 27; Povinelli & Frey, 2016, Exp. 1-2; for examples with ropes see Povinelli, 2000, Exp. 14; Povinelli, 2012, Exp. 15-17; Povinelli & O'Neill, 2000, Exp. 1). Given the importance of ropes to the current study, we note that in addition to the published experiences detailed above, we also conducted several other unpublished studies involving ropes, strings, and hook or hook-like tools. The chimpanzees' indoor-outdoor compound was also permanently fitted with both loops and untied lengths of rope for enrichment purposes. These were used daily by the chimpanzees.

A familiar "request to respond" (or RTR) procedure was used in this study. This procedure allowed the apes to enter the test unit, approach a closed window/opening in the Lexan, and touch a small target directly above it. In response, an out-of-sight experimenter (who monitored the ape on a video screen) remotely opened the window. The purpose of the RTR procedure was two-fold: 1) it gave the animals control over when a trial started, and 2) it gave the experimenter an unambiguous signal for when to let the apes respond to experimental stimuli.

Video cameras in the testing unit were fed into a remote control room. Here, all setup and testing procedures were monitored live by other experimenters and recorded. All trial configurations and methods were double checked and verified via remote intercom before they were implemented. The video recordings were used for coding purposes as well as to ensure standard operating procedures were followed. Two manuals governed the conduct, recording, data disposition, archiving, and video coding of all experiments (Cognitive Evolution Group, 2010a, b). Following our laboratory practices, a full-time study director and study review coordinator were assigned to carry out and monitor the conduct of the experiment.

The animals had 24 hr access to water and indoor-outdoor housing and were never food deprived for the purpose of the studies. Round the clock veterinary care was available if needed. The study was approved by the Animal Care and Use Committee of the University of Louisiana and all federal and state guidelines for the care and use of animals were far exceeded.

Materials and Procedure

General Design

The study was designed to compare the performance of the apes in two main conditions using rope tools: a looped rope and an untied rope (hereafter, loop and untied rope, respectively; see Figure 1b-c). The chimpanzees had mastered numerous tasks (see above) that, on the surface, were both perceptually and conceptually similar to the loop and untied rope conditions (see details below). The materials and configurations used in the current study, however, had never been presented to them. We also administered a highly familiar hook tool condition (see below and Figure 1a), to (1) gauge the apes' motivation to participate, and (2) to verify their relative expertise on a condition conceptually (and perceptually) similar to the rope conditions. Given the apes' prior experience with hook tools (see above), we elected to offer no pre-training or re-familiarization with them.

The tool conditions were administered as separate trials, delivered across 12 sessions. The apes received 1-2 sessions per day over a 10 day period. The first six sessions consisted of four trials each. The first trial was always a hook trial. The remaining three trials included one of each of the following: hook, loop and untied rope. We presented these three trial types in a randomized order (using an across-session, 4-trial counterbalancing constraint). Because the apes made no errors on the hook trials during the first six sessions, we eliminated them during the final six sessions, which therefore consisted of two (randomized) loop and untied rope trials. Thus, the apes participated in 12 hook trials (two per session across the first six sessions), and 12 loop and 12 untied rope trials (one of each per session, across the 12 sessions).

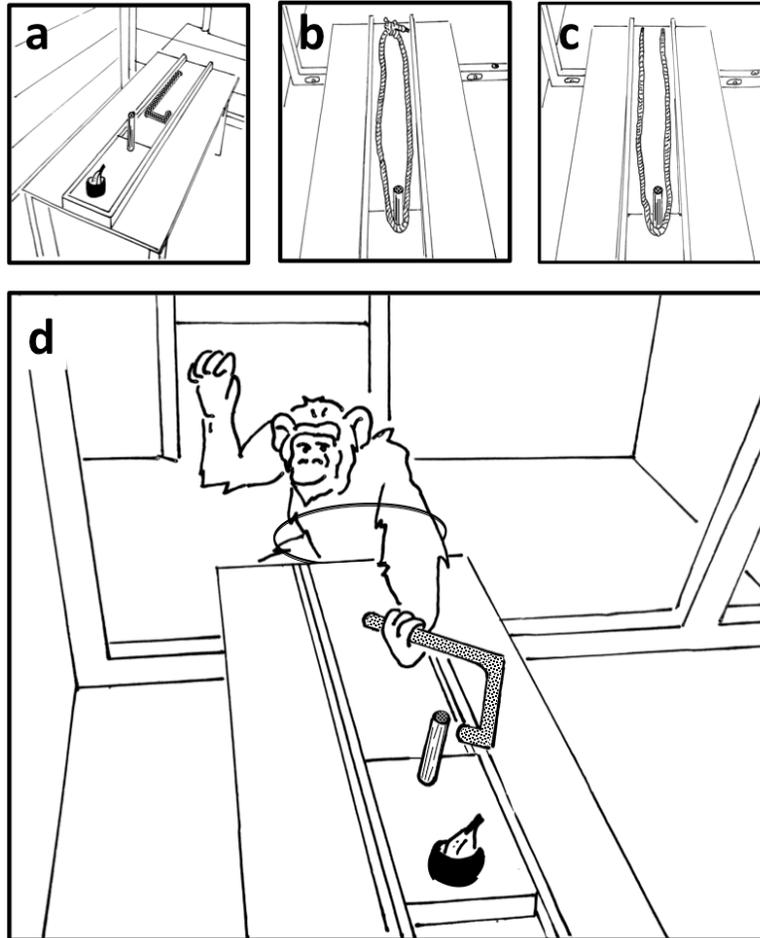
Trial Setup and General Procedure

Before each trial began, an experimenter positioned an elongated large table (72 x 180 x 60 cm) against the Lexan partition (see Figure 1). The table contained a channel extending away from the apes' enclosure. A small platform (45 x 25 x 3.5 cm) made from a rectangular block of wood could slide through the channel toward the ape. This sliding block contained a vertical peg on the end closest to the ape. A food reward (a slice of fruit or its equivalent) was placed into a food cup anchored on the far side of the block (see Figure 1a, d). The block was then placed into the channel at the far end of the table, out of the apes' reach. Using the tools provided, the apes could drag the block toward them. The apes had mastered a very similar procedure in the past, including a study using the same table apparatus and hook-like rakes (see Povinelli & Frey, 2016, Exp. 1-2).

The apes had access to two response windows. A main response window (which could be remotely opened or closed) allowed the apes to reach onto the table and manipulate the tools that were provided. A secondary response window near the far wall (away from the table) remained open and was used only during the loop and untied rope trials. This window allowed the apes to pass the rope tools (which were initially located in their enclosure) to the experimenter when he requested them (see details below).

Figure 1

Testing Configurations for the (a) Hook, (b) Loop and (c) Untied Conditions, and (d) Megan Using the Hook Tool to Retrieve the Sliding Block by Placing the Crooked End around the Post



Note. The visible food cup in (a) and (d) was also attached to the sliding block in (b) and (c). Figures redrawn from photographs (a-c) and still video (d) by DJP.

Procedures for the Testing Conditions

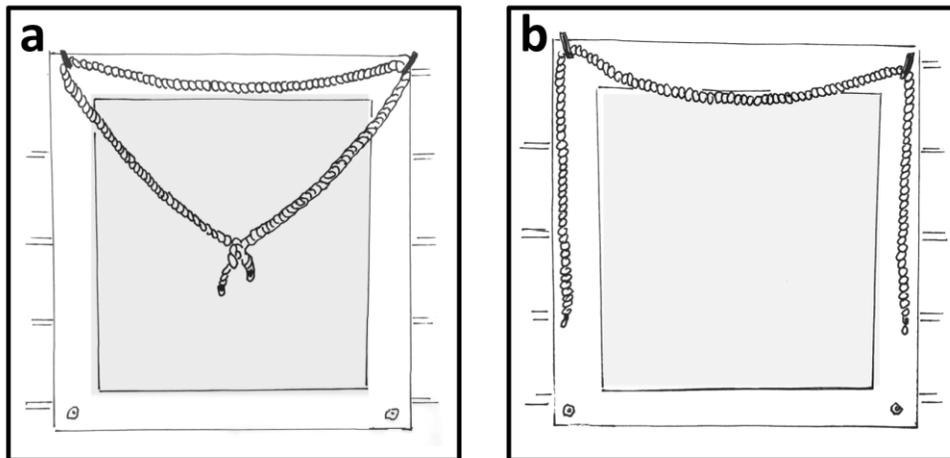
Hook tool. As the apes waited outdoors, the experimenter baited the food cup and placed a 60 cm long hook tool in the channel with the handle end within easy reach (see Figure 1a). The trial began as the experimenter concealed himself in a small booth at the rear of the test unit and opened the shuttle door. As soon as the apes entered, the shuttle door closed behind them. For these trials, the main window was already open, allowing immediate access to the hook and apparatus. When the ape first touched the hook tool, secondary experimenters (located in a remote control room) started a 30 s timer. During this time period, the apes were allowed to manipulate the tool in any manner they chose. As soon as the ape obtained the reward, or after 30 s had elapsed, the main experimenter began to lower the response window, signaling the end of the trial. The shuttle door was then opened and the ape returned to the outdoor waiting area.

Looped rope. For loop trials, the experimenter set up the table as described above. Instead of placing a hook tool on the table, however, a 175 cm length of rope that had been tied into a loop was

draped on the inside wall of the apes' test unit (see Figure 2a). After the ape entered the room, the experimenter emerged from the booth and, using gestures and speech, requested the rope. The apes readily removed the rope from the wall and passed it through the secondary window to the main experimenter who collected it, draped it around the peg of the block, and oriented the rope so that it was in a standardized position within reach of the apes (Figure 1b). The experimenter was careful to maintain the apes' attention throughout this procedure. Once finished, the experimenter returned to the booth. As soon as the ape touched the RTR symbol, the experimenter remotely opened the response window, and the trial began. The apes were able to manipulate the loop in any manner they chose until the trial ended (following the 30-second decision rule described above).

Figure 2

The Presentation of the (a) Loop and (b) Untied Rope Tools



Note. The ropes were inside the apes' portion of the testing unit when they entered and were handed to the experimenter by the apes as he requested them. This allowed the apes additional visual and tactile experiences of their properties before they watched him lay them into position as shown in Figure 1b and c. Figures redrawn from photographs by DJP.

Untied rope. The procedure for the untied rope trials was the same as loop trials, except that instead of the rope being tied into a loop, an untied rope of the same overall length was draped on the wall (see Figure 2b). Once the experimenter received the untied rope from the ape, he draped it around the peg and placed the two ends on the table nearest the apes in a standardized position. The ends were not touching and were equidistant from the peg on the sliding block.

The difference in how the tools were presented in the hook case (window open and tool already in place) versus the two rope conditions (ropes draped inside the apes' enclosure) was due to the use of the hook condition as a general procedure to ensure the apes were motivated to perform a retrieval task on which they had repeatedly demonstrated mastery. We were not interested in using the results of the hook condition to assist in licensing specific inferences about the apes' cognitive systems. Any such inferences were to be generated on performance comparisons between the tied and untied rope conditions.

Video Coding and Reliability

All trials were recorded on video for later coding of the apes' behavior by trained observers. Cameras were positioned on the ceiling and wall of the testing unit and were controlled by the experimenters in the remote control room. The cameras provided a full visual of the table, apparatus and tools used, and the apes' behavior (Figure 1c).

A primary rater coded all trials using a set of standardized written instructions, and a secondary rater independently coded 25% of the sample (the first three sessions for all 7 animals).⁵ For hook tool trials, the raters recorded if the apes successfully used the hook to drag the sliding block to within reach and obtained the reward. For the main conditions (loop and untied rope), the raters again recorded if the ape was successful. In addition, they recorded which of two actions the apes deployed to try to get the reward: (1) grasping one length of the rope or (2) grasping both lengths of the rope. For this measure, the raters recorded both the apes' initial and final actions on the ropes. Separate actions by the apes were defined as completely releasing the rope, or by changing his or her behavior (e.g., pulling one length, and then, without releasing it, grabbing the second length). If the apes were successful using their first action, then the raters recorded their first action as their final action as well.

The raters displayed perfect agreement on their judgments of success for hook trials (40/40 = 100% agreement, Cohen's κ = undefined), as well as loop and untied rope trials (40/40 cases of agreement, Cohen's κ = undefined). For the category of initial action, there was one disagreement (39/40, Cohen's κ = .918). For final action, there were two disagreements (38/40, Cohen's κ = .90). The study review coordinator resolved the three cases of disagreement by observing the trial with the primary rater. The resolved judgments were used in the data analysis.

Results

Our description of the results proceeds in three steps. First, we examine the apes' level of success (obtaining the reward or not) on the hook, loop, and untied rope trials. Next, we examine how the apes adjusted their specific actions on the loop and untied rope tools (pulling one or both lengths of rope) both within and across trials. Finally, we compare and contrast the performances of individual apes.

Overall Success

Figure 3 depicts the apes' averaged success in the hook, loop and untied rope trials in blocks of two sessions. This equates to blocks of 4 trials for the hook condition, and 2 trials for the loop and untied rope conditions. (Trial-by-trial data for the individual apes is provided in Table 1 and is discussed below.)

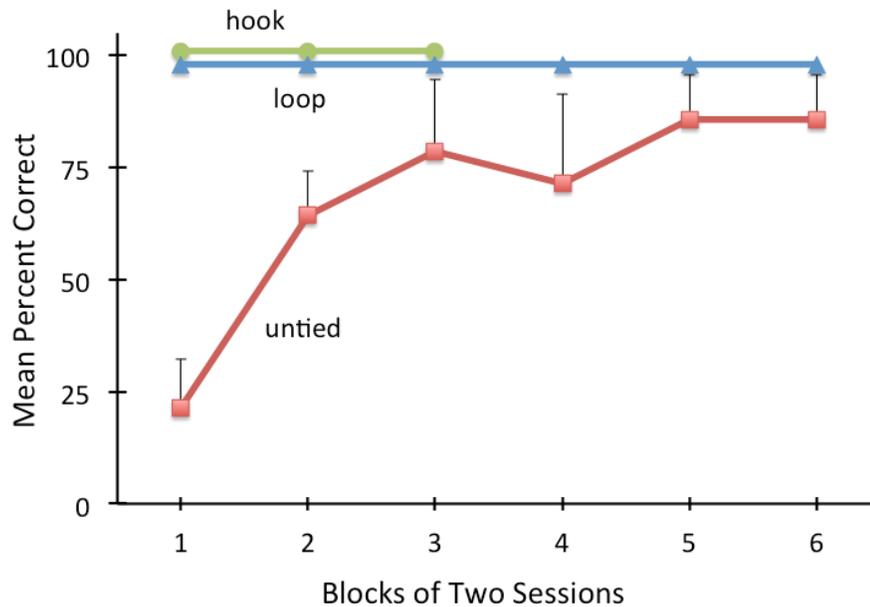
On hook trials, the apes were 100% successful in using the tool to obtain the reward (Figure 3). The apes typically entered the testing unit, proceeded directly to the testing table (recall that the window was already open and the hook tool was resting on the table), picked up the hook tool, extended it toward the baited sliding block, and oriented the crooked end around the post. From there, they easily dragged the block toward them and retrieved the reward. There were, however, idiosyncratic differences among the apes in the manner in which they maneuvered the crooked end of the tool to the peg (see discussion in Povinelli, 2000, p. 220-221).

Similar to the hook trials, the apes were successful on every loop trial (Figure 3). They achieved success by grasping either one or both lengths of the rope and pulling the platform within reach. In marked contrast, the apes' averaged success level on the untied rope trials was low and variable across the first two blocks of trials/sessions. Two of the seven apes (28.6%) were successful on trial one (Jadine, Mindy), and one of the seven (14.3%) was successful on trial two (Kara). Beginning with the second block of trials/sessions, however, the apes' success improved quickly, stabilizing at 75% success or higher (see Figure 3). A more fine-grained examination of the trial-by-trial data of the individual apes revealed that the majority of them (4/7, 57.1%) improved most between trials 3 and 4 (see also Individual Performances, below). The trial-by-trial data also reveal that most of the variability was due to the relatively lower success levels achieved by Megan and Mindy.

⁵ We planned the reliability analysis to be based upon 42 hook trials (3 sessions x 2 trials x 7 animals) and 42 loop and untied trials, combined (3 sessions x 2 trials x 7 animals). Due to a video recording error, however, raters were unable to extract data from trials 2-4 of Jadine's third session. The final reliability analysis was therefore based on 40 hook trials and 40 loop and untied rope trials.

Figure 3

Mean Percent of Trials (\pm SEM) in which the Apes were Successful in Retrieving the Food Reward from the Sliding Platform in Blocks of Two Sessions for Hook, Loop and Untied Conditions



Note. The apes received one trial of the loop and untied rope conditions per session, so the individual data points for these conditions are the average of two trials. Because the apes received two trials per session of the hook condition, individual data points for this condition represent the average of four trials per block of two sessions. Hook trials were discontinued after session 6 (see text for details). Data points for hook and loop conditions have been offset for visibility.

Behavioral Adjustments Within and Across Trials

Figure 4a depicts the average percent of trials in which the apes' initial actions were to grasp both lengths of rope. These data reveal a similar pattern in the loop and untied rope conditions. In the untied rope condition, there were only two instances in the first block of trials/sessions (one on the first trial and one on the second) in which an ape's initial action was to grasp both lengths of rope. Thus, for the majority of cases in the first two sessions, when confronted with the untied rope condition, the apes initially grasped one end of the rope. They proceeded to pull it, watching as the other end slowly traveled out of reach and snaked around the peg, thus making it impossible to obtain the reward. In the first few sessions, the apes also preferred to initially grasp one length of rope in the loop trials.

Figure 4b presents the apes' final actions in the two conditions across trials/sessions, and reveals how, as the experiment proceeded, their final actions changed from grasping one end to grasping both lengths/ends. This occurred in both conditions although this shift may have occurred slightly faster in the untied condition. Importantly, the trial-by-trial data (see Table 1) revealed there was an inflection point for the majority of the apes after trial 3 that characterized the shift from initially grasping one length of rope, to grasping both. As can be seen by comparing Figure 4a and 4b, this occurred in parallel in both conditions (see Individual Patterns, below).

Table 1

Trial-By-Trial Data for the Number of Lengths of Rope That Were Initially and Finally Grasped by the Chimpanzees in the Loop and Untied Rope Conditions

		Trial number											
		1	2	3	4	5	6	7	8	9	10	11	12
Apollo	loop	1,2 ^a	1,2	1,2	1,2,	2,1	1,2	2,2	2,2	2,2	2,2	2,2	2,2
	untied	1,1	1,1	2,2	2,2	1,1	2,2	2,2	1,2	2,2	1,2	2,2	1,2
Kara	loop	2,2	1,1	2,2	2,2	2,1	2,2	2,1	2,2	2,2	2,2	2,1	2,2
	untied	1,1	2,2	1,1	2,2	2,2	2,2	2,2	2,2	2,2	2,1	2,2	2,2
Candy	loop	1,2	1,2	1,2	2,2	1,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2
	untied	1,1	1,1	1,2	1,1	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2
Jadine	loop	1,2	1,1	-- ^b	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2
	untied	1,2	1,1	-- ^b	2,2	1,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2
Brandy	loop	1,1	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2
	untied	1,1	1,1	1,1	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2
Megan	loop	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1
	untied	1,1	1,1	1,1	2,2	2,2	2,2	1,1	2,1	1,1	2,2	1,2	2,1
Mindy	loop	1,1	1,1	1,1	1,1	1,2	1,1	1,1	1,1	2,2	2,2	2,2	2,2
	untied	2,2	1,1	1,2	1,1	1,1	1,1	1,1	1,1	2,2	2,2	2,2	1,1

Note. ^a 1=ape grasped one length of rope, 2=ape grasped both lengths of rope. The first value in each series indicates the ape's initial choice, the second value indicates the ape's final choice.

^b This trial is missing due to a video recording error (see footnote 3, main text).

Figure 4

The Mean Percent of Trials (± SEM) in which the Apes' (a) Initial and (b) Final Actions on the Ropes Involved Grasping Both Lengths of Rope, Depicted by Condition

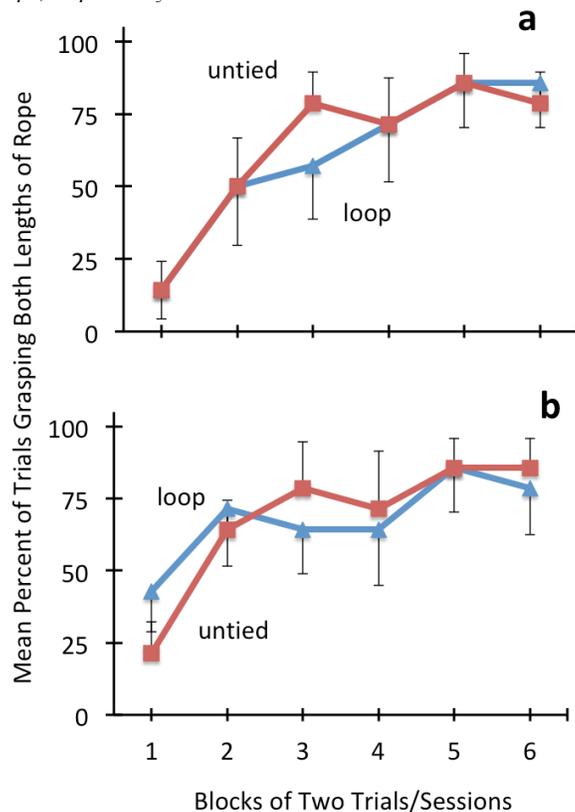


Figure 5a illustrates the apes' initial versus final actions in the loop trials. It is apparent that, in the first 2-3 blocks of trials/sessions, the apes executed considerable within-trials adjustments in the direction of initially grasping one length of rope to grasping both. Figure 5a also reveals that, across trials, most of the apes also shifted their initial actions to prefer immediately grasping both lengths of rope. The fact that it was not necessary to do so is evident from the fact that, even in the final loop trials, some apes occasionally grasped only one length of rope, yet they were still always successful. These within- and across-trial adjustments in the loop condition likely reflect the fact that, depending on the exact trajectory of their pulling motion, the initial action of pulling one length of rope did not always move the block as the loop circled around the peg. This was followed by many occasions in which the apes leaned or stepped backwards (while still holding only one length of the rope), until the loop became taut and the food reward moved toward them. This involved extra effort, and there were a number of occasions in which the apes began switching to two hands within a trial, before adopting the easier solution of simply initially grasping both lengths of rope (see Individual Patterns below).

Figure 5

The Mean Percent of Trials (\pm SEM) in Which the Apes' Initial and Final Actions Involved Grasping Both Lengths of Rope, in the (a) Loop and (b) Untied Condition.

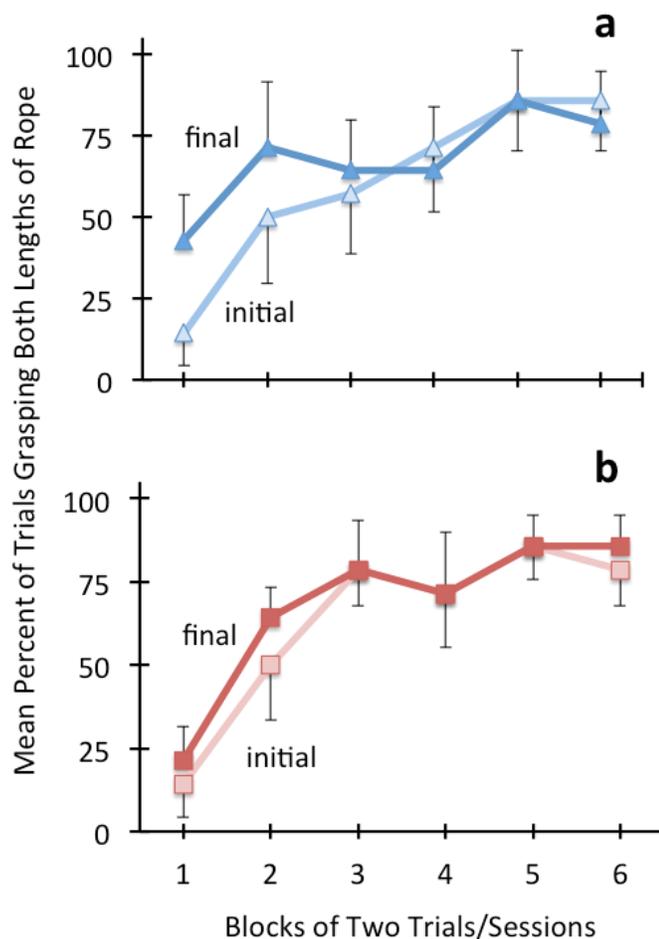


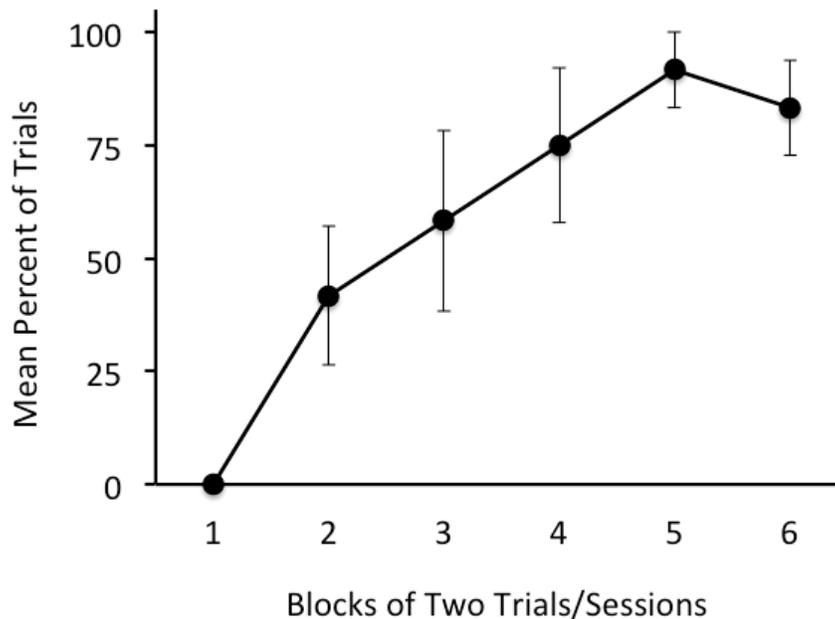
Figure 5b illustrates the apes' initial versus final actions during the untied rope trials. The within- and across-trial pattern reflects the much tighter causal constraint that the task imposed upon the apes' initial actions: pulling a single end of the untied rope generally made the task unsolvable. As described above, the apes began by almost never initially grasping both ends of the rope, to almost always doing so,

and as can be seen in Figure 5b, this change was most marked during the first 2-3 blocks of trials/sessions (the first 3-4 trials). The majority of cases, but not all of them, were in the direction of shifting from grasping one length to two (see below). Notably, there was only a single instance in the first two sessions (14 trials) of this condition (Jadine, trial 1), where an ape adjusted his or her behavior within a trial in the direction of grasping one to both ends of the rope, before it was too late.

Next, we examine the co-development of the initial action of grasping both lengths of rope in the two conditions across sessions. Figure 6 is a plot of the average percentage of sessions (in blocks of two sessions) in which the apes' initial actions in both the loop and the untied rope within a session involved grasping both lengths of rope (in other words, performing the same initial action for both trial types in a session). These data exclude Megan, who never grasped both lengths of rope in the loop condition (see Table 1). The apes shifted from never grasping both lengths of rope as their initial action on both trial types within a session, to doing so around 80% of the time.

Figure 6

The Mean Percent of Trials (\pm SEM) in which the Apes' Initial Actions (Within Each Sessions) for Both the Loop and Untied Rope Trials were to Grasp Both Lengths of Rope



Note. These data show how, over time, as a group, the apes converged on the same initial action of grasping both lengths of rope, regardless of condition. (NB: Megan's data are not included in this graph because in the loop condition her initial action never involved grasping both lengths of rope.)

Individual Patterns

An exploration of the individual behavioral patterns of our apes (see Table 1) provides important context to the findings described above. Although the apes were 100% successful in the loop condition, their routes to success varied. For example, the apes changed how many lengths of rope they grasped during 16.8% (14/83) of the loop trials, and 71.4% (10/14) of these adjustments were in the direction of grasping one length of the rope to both. Apollo, Kara and Candy, in particular, continued to adjust how many lengths of rope they pulled in the loop condition throughout the course of the experiment, whereas the remaining apes exhibited very few within-trial adjustments (see Table 1). As described above, it was our impression that the within-session adjustments occurred because the sliding block frequently did not initially move when the apes pulled only one length of the loop.

In the untied condition, the apes adjusted how many lengths of rope they grasped in 13.2% (11/83) of the trials. As was the case for the loop trials, where the majority (72.7%, 8/11) of the adjustments were in the direction of grasping one end of the rope to grasping both. There were three notable trials, however, where Megan (twice) and Kara (once) switched from grasping both ends of the rope, to then grasping only one, resulting in failure. Table 1 reveals how, in the untied condition, four of the apes (Kara, Candy, Jadine, Brandy) developed a successful initial approach to the untied rope condition within the first several trials. The remaining three apes (Apollo, Mindy, Megan) never appeared to do so reliably, each initially grasping both ends of the rope only 50% of time across the second half of the trials (trials 7-12). Of these three, only Apollo was able to learn to successfully adjust his behavior within trials, changing from grasping one end of the rope to grasping both ends. This occurred on three of his final six trials and allowed him to achieve 100% success in the second half of the experiment.

Discussion

We were impressed by the performance of our apes. First, they demonstrated a long-term retention of a skilled form of tool-use involving hooking an out-of-reach object. Second, they performed flawlessly when confronted with a new tool (the looped rope). Third, and most importantly, most of them quickly mastered a new problem involving the untied rope. Admittedly, the rate of skill development with the untied rope tool varied. Nonetheless, five of the seven apes (71.4%) cemented their skill on this tricky task within the first three trials. Thus, by trial 4, these five apes had learned to either immediately grasp both ends of the rope, or (in the case of Apollo) when they did not, learned to appropriately adjust their behavior before pulling the other end out of reach. Two of the apes (Megan and Mindy) "performed poorer" and "learned slower." Do these results necessitate the operation of higher-order, structural, role-based relational reasoning?

To begin, given the pre-existing competence of this particular group of chimpanzees, it is not surprising that their ability to use a hook tool would persist over time. The hard-won expertise of tool-using chimpanzees living in the wild is instructive. After young chimpanzees master a task like termite-fishing or cracking nuts, it is inconceivable they would lose this expertise over a matter of weeks, months, or even years of not having the opportunity to exercise it (see Boesch & Boesch-Achermann, 2000; Goodall, 1986). To wit, there are numerous reports of tool-related skills persisting in captive chimpanzees (and other animals) over long periods of time (Beck, 1980). Nonetheless, we suspect few psychologists would accept the mastery and long-term retention of using a hook tool (or a termite wand) as evidence that the animal in question possesses higher-order, role-based relational representations involving constructs like *<rigidity>*, *<connection>* and/or *<force>*. Why not? A typical objection would be that the animal could have learned the relations in question (how to use the tools) without the mediated influence of such higher-order representations, and like other perceptually-based representations, these learned skills could be encoded in long-term memory to be deployed as needed.

To rephrase all of this in the language of the broader argument we are building, this objection can be recast more formally: evidence of a particular first-order, perceptually-based relational ability (whether acquired via learning or evolutionarily endowed), is necessary but not sufficient evidence of higher-order, structural, role-based relational reasoning. Indeed, in the case of mastering the use of a hook to retrieve an out-of-reach banana (or fashioning twig-like probes to extract termites from a mound), scholars can always point to the extensive learning histories that precede competence (see, for example, data on the ontogeny of nut-cracking skills of chimpanzees in Boesch & Boesch-Acherman, 2000). Ultimately, such arguments depend on the assumption that such learning histories cripple our ability to discern if the skill is mediated by first-order, perceptually-based, relational reasoning alone, or whether higher-order relational reasoning also modulates the behavior. (We write the words "also modulates" with great care because, as we noted earlier, *a mental system that wields higher-order representations must necessarily wield first-order, perceptually-based ones as well*. Even human reasoning, no matter how high its highest-order relations reach, are ultimately dependent upon perceptually-based relations; see Penn et al., 2008a; Povinelli, 2012.)

The phrase "associative learning" is frequently invoked to capture this alternative explanation, both in the context of tool-use and other domains of cognition (e.g., Seed et al., 2009). Unfortunately, terms such as "associative learning" or "causal understanding" obscure crucial questions concerning one's commitments regarding an animal's cognitive architecture, and are therefore too broad to assess a principled theory of the kinds of representations that are necessary and sufficient for learning how to use and manufacture tools, or how the long-term retention of these skills tools relates to alternative models of mental structures. Such terminology also frequently obscures the sources of learning in tasks periodically heralded as breakthroughs in the understanding of animal cognition (e.g., Ghirlanda & Lind, 2017; Hennefield et al., 2018).

In response to these intuitively-based concerns, experimentally-minded psychologists and philosophers have therefore reasoned that more diagnostic tests should be invented and administered. We include our former selves in this category (see the appendix; see also Barker & Povinelli, 2019). The hope was (or is) that empirical results from new task variants would (or will) have the power to distinguish between organisms who solely rely upon intelligent first-order, perceptually-based relational reasoning, and those that also possess higher-order, structural, role-based relational reasoning (see Penn et al., 2008a). In the case of the rakes and hook tools, putative examples of such probative variants abound (e.g., chimpanzees: Povinelli, 2000, Exp. 3-10, 15-16, 19-20, 27; Povinelli & Frey 2016, Exp. 1-2; great apes: Martin-Ordas et al., 2008; cotton-top tamarins: Santos et al., 2005; crows: Bird & Emery, 2009a; Hunt & Gray, 2004; St Clair & Rutz, 2013; Weir et al., 2002).

As we explore further below, however, every task must possess its own perceptual grounding. Hence, any concern that an organism's learning history can fatally contaminate experimental tasks—and thus block (or, at the very least, undermine) strong inferences about the operation of higher-order relational reasoning—must, by extension, apply to any new task (or its future variants) as well. In response to this concern, the rate of learning exhibited by an animal is frequently invoked as somehow relevant, with special emphasis placed on trial one, in particular (we, too, once invoked such emphasis: see Barker & Povinelli, 2019; Povinelli, 1988). However, using an animal's excellent performance on trial one (or two or three) as adjudicating evidence in favor of the operation of higher-order reasoning rests upon an extremely suspect, and ultimately unprincipled, titration: namely, that the problem instantiated by the task is perceptually *similar enough* to things that the animals have previously encountered to allow them to make sense of it, but perceptually *different enough* that they must rely upon higher-order reasoning to navigate their way through it.⁶ It should be noted that this already daunting problem of perceptual grounding is dramatically magnified when one considers the evolutionary histories that have sculpted a given species' unique constellation of sensitivities to innumerable perceptually-based relations (see Clark & Thornton, 1997; Povinelli & Penn, 2011).

Consider the untied rope condition, which was originally conceived of as a possible method for assaying our apes' level of causal reasoning. In particular, let us contrast the animals who performed "poorly" or "learned slowly" (we paranomastically applied these epithets to Megan and Mindy earlier) versus those that performed "excellently" or "learned rapidly" (more laudatory terms we applied to the performances of our other apes). In the case of Mindy and Megan, one can readily enumerate many reasons why their performances do not seem impressive from the point of view of higher-order reasoning.⁷ What is less often considered is why the performances of the more accomplished animals

⁶ If it turns out that humans are alone in having evolved higher-order, structural, role-based relational reasoning, and this system's primary initial function was to reinterpret the already highly intelligent outputs of first-order, perceptually-based systems, then attempting such titrations might prove to be not only methodologically unprincipled, but also a fool's quest (see Bering & Povinelli, 2003).

⁷ Only one of our animals (Mindy) performed perfectly on *trial one* of the untied rope condition and she did not continue to exhibit this behavior. Does this warrant the inference that our apes did not engage in higher-order causal diagnostic reasoning as they handed the rope to the experimenter and watched him drape it around the peg? We have discussed the scientific folklore surrounding the idea of *trial one* elsewhere (e.g., Barker & Povinelli, 2019; Povinelli, 2012). Here, we simply emphasize that this question is distinct from whether animals possess, but in certain circumstances do not exercise, any higher-order capacities they may have. In various guises, this latter problem has appeared and reappeared throughout the history of comparative psychology, including in the case of the string-pulling problem. In assessing the results of more than a century of research on this topic,

cannot be attributed solely to the operation of first-order, relational reasoning (for an outstanding dissection of this problem in the case of temporal cognition and planning in animals, see Hoerl & McCormack, 2019).

Nonetheless, at a functional level of description, it is certainly true that the majority of our apes rapidly developed from failing to compensate for the objective (causal) fact that pulling on end of the untied rope would not transfer sufficient force to the peg (and thus produce no useful work), to doing so reliably. Could the rapidity of this learning be used to construct an argument in favor of the presence of higher-order representations? One such (inductive) argument (with two premises and a conclusion) proceeds as follows:

- P1 In order to master the skill of using the untied rope, the apes needed to keep track of how their actions on one end of the rope, caused the other end to move away from them.⁸
- P2 Learning to grasp both ends of the rope after only two or three opportunities of experiencing the mechanics of the system is evidence that they understood that the rope needed to tighten around the peg in order for the sliding block to move toward them.
- C To fully explain the observed pattern of results, we must appeal to higher-order relational reasoning involving constructs like *<transfer of force>* (or, more technically, higher-order reasoning is necessary, though not sufficient, to fully explain the observed pattern of results.)

Clearly, however, P2 must be wedded to some other, background theory about why the two systems under consideration differ in the rates of learning they can achieve, and how any such differences articulate to a specific, situated task. We know of no such theory. Indeed, as we elaborate below, the asymmetric dependency relation between first- and higher-order representations leads us to conclude that no such theory can be derived from the current representationalist theories of cognition in which these tasks are situated. In the absence of such a background theory, keeping track of the relevant first-order perceptual relations would seem necessary and sufficient to produce the actions described: *<pull one length of rope, no movement of block; pull two lengths of rope, block moves>*. By obvious extension, if first-order relational reasoning is both necessary *and* sufficient, then higher-order relational reasoning has no role to play in explaining the results.

A related (but typically overlooked) point concerns how many relevant learning units an animal experiences prior to mastery. Following the folk scientific practices of comparative psychology, we experimentally divided our experiment into "trials." But what we define as a single trial may not accurately map onto the learning machinery of our chimpanzees. As Apollo pulls on one end of the untied rope, what is the relevant unit of learning? The other end of the rope traveling just out of reach? The end snaking off the peg? The experimenter lowering the response window as time expires? If learning occurs continuously throughout rope movement, however, then our idea of a "trial" becomes highly misleading—at least with respect to using it as an objective measure of a real unit of learning. By consequence, invoking the number of trials to mastery as evidence in favor of higher-order relational reasoning is substantially weakened. Elsewhere, we have empirically and analytically exposed the

Jacobs and Osvath (2015) note, for example: "Older animals often perform better because they are more cognitively developed and less playful . . . although juveniles might be more successful because they can be more persistent" (p. 92).

⁸ For now, we set aside the problem of whether the shift to grasping both lengths of rope became an inevitable embodied act once it became apparent that grasping one length did not result in the block moving. We suspect, however, that in certain quarters of cognitive science this explanation would have many adherents (for example, see Barrett, 2011; Frigaszy & Mangalam, 2020.) After all, these data show how the apes converged on a simple solution to this task, regardless of condition: *<grasp both lengths of rope, pull>*. Properly understood, this would apply even to Apollo's actions, described earlier. (An even more deflationary idea was suggested by one our reviewers who suggested that perhaps the apes simply grabbed the second length or rope as it was moving away simply because they wanted "more rope", and thus perhaps never understood the causal necessity of doing so. We neither endorse nor reject this possibility.)

damaging implications of ignoring this problem (Barker & Povinelli, 2019; Hennefield et al. 2018). Note that this difficulty is distinct from both the perceptual grounding and rate of learning challenges described earlier.

Certain scholars have attempted to counter these arguments deploying one or more of three (related) arguments, each of which is fatally flawed:

- (1) Some scholars argue that our explanations are *ad hoc* (for a seminal statement of this view, see Tomasello & Call, 2006). Alas, such objections directly fail to address the full import of the asymmetric dependency problem. And so, by extension, they fail to address what we have (re)stressed here: the claim that first-order representations are necessary and sufficient to account for the results of such tests systematically applies in all cases of this type, including new cases yet to be articulated.
- (2) Scholars also argue that the explanation above is "less parsimonious" than the higher-order alternative. Once again, this argument is defeated by the asymmetric dependency relationship: a higher-order system entails the operation of its lower-order components, but not vice versa (for a detailed explication of this asymmetry in the context of social reasoning, see Povinelli & Vonk, 2004). And if the first-order components are necessary and sufficient to obtain the results, it cannot be less parsimonious (in the straightforward sense of positing more variables to account for the same observations). Indeed, the reverse would seem to hold. (For more detailed discussions and other interpretations of the parsimony question and how it relates to the case of higher-order reasoning about mental states, see Andrews, 2016; Clatterbuck, 2018; Dacey, 2016; Sober, 2016a.)
- (3) Finally, scholars also argue that it is "more plausible" to assume the presence of the higher-order system, especially given what are often described as the "growing bodies of evidence" in support of higher-order interpretations. But this reasoning is fatally circular. Unless one can first produce some independent evidence for either (a) the limits of the causal power of the first-order, perceptually-based reasoning, and/or (b) the unique causal power of the additional, higher-order system, then there is no way to assess the relative plausibility of the two alternatives.⁹ By "independent evidence" we directly exclude evidence derived from the experimenter's intuitions or folk ideas about how their own (or other) minds reason about the world (see Povinelli & Giambrone, 1999). Thus, the "growing bodies of evidence" should be restated as "growing bodies of utterly indeterminate evidence." In the appendix, we suggest how (a) and (b) could become a focal point for future research.

The above considerations suggest a different way of relating to our subjective feelings of being "impressed" by our apes. Our analysis suggests that uncontroversial tenets of representational theories of cognition urge us to be impressed by the causal power of first-order, perceptually-based, relational reasoning. Until such a time as we develop a principled account of the additional causal work that higher-order, role-based relational reasoning may contribute to such results, we have no business invoking it to explain them. This leads us to an additional question: Is there any (hypothetical) pattern of results in this experiment that could ground a strong inference in favor of the operation of the higher-order system? The answer is now obvious: no. Even flawless performance from trial one forward would suffer from the perceptual grounding issue described above.

For those who remain unsettled and/or unconvinced, and are thus understandably tempted to think that additional, cleverly-designed task variants can overcome this conceptual problem, it is worth noting

⁹ One might be tempted to think that evolutionary considerations (such as the degree of phylogenetic relatedness of two species) might help to adjudicate which alternative is more or less parsimonious or plausible (see Sober, 2012). However, these arguments quickly become problematic because the answers depend on unproven assumptions about (1) when the higher-order systems evolved in relation to the first-order systems, and/or (2) the precise role they play in modulating targeted aspects of human behavior (Povinelli & Giambrone, 1999).

that it would be conceptually impossible to conjure a way of presenting an animal with some stimuli (such as ropes) in a manner that did not fall under some pre-existing set of perceptual generalization gradients, each derived from either its evolutionary or developmental history or, more likely, the complex interaction of both. In the present case, our apes possessed nearly twenty years of structured and unstructured experiences not just with ropes and strings, but with a host of other materials as well (sticks, rakes, hooks, plastic chains, burlap sacks, sheets of cardboard, pull toys, etc.), all falling under robustly overlapping perceptual gradients. Given such developmental facts, how could a new, rope-like stimuli not connect to these existing representations?¹⁰ But even if one could conjure such stimuli, on what basis would one then expect the animal to perform well on trial one?

One narrow interpretation of the foregoing analysis is that tasks involving ropes and strings are simply bad choices for experimental attempts aimed at untangling the unique causal signatures of higher-order role-based relational reasoning. We believe this interpretation radically underestimates the problem facing comparative psychology. Properly understood, the asymmetric dependency relationship between first-order, perceptually-based reasoning system and the kinds of higher-order, role-based reasoning of which humans are manifestly capable, renders an entire genre of historical experimental strategies largely unhelpful. An example may suffice to illustrate the broader argument. Schrauf and Call (2011) allowed great apes to pull vertical ropes from which opaque containers were hanging (see also, Hanus & Call, 2011, for a similar approach not involving ropes or strings). Some of the containers were baited with food (and were therefore harder to pull) whereas others were empty (and therefore easier to pull). Across trials, the apes tended to favor completely pulling in the containers that contained food (that is, the ropes that were harder to pull). They did not perform as well when the experimenters used arbitrarily chosen color cues to signify which cups were baited. The authors concluded that apes are more sensitive to causal cues (i.e., the effort to pull cups loaded with food), than arbitrary ones (a randomly chosen color). We agree. In addition to their species' evolutionary histories, these apes had innumerable experiences lifting food and other objects (including objects containing food). The effort to lift or pull objects, then, is a prioritized causal relation involving a finely-tuned perceptual sensitivity (for some examples with chimpanzees, see Povinelli, 2012, Exp. 1-8). Thus, the work of Schrauf and Call should not be interpreted to suggest that an organism's evolved perceptual sensitivity to some causal fact about the world is evidence of higher-order relational reasoning. Instead, it should be understood as another demonstration of why "associative learning" is too broad a concept to be helpful in the debate over whether animals are capable of higher-order reasoning. No pattern of results from their experimental approach, no matter how many task variants are invented, could ever implicate the operation of higher-order reasoning about *<weight>*. Of course, not all perceptual relations deriving from the acceleration of mass in the Earth's gravitational field will be equally prioritized by all species. But as Clark and Thornton (1997) have demonstrated more generally, organisms must come into the world pre-prepared to learn a host of relations, and thus some relations will be easier to learn than others (see Povinelli & Penn, 2011).

We conclude that no experimental approach that presents physical (or virtual) objects to animals (as tools or otherwise), and then measures whether or how quickly they demonstrate mastery of objective causal facts about them, will ever allow for strong inferences about the presence of higher-order relational reasoning. Our analysis shows that first-order, perceptually-based, relational reasoning is both necessary and sufficient for such mastery. If correct, any claims for the necessity of higher-order reasoning to explain the results of such tasks are, by definition, false. Whether tasks that fall outside this genre can do so remains an open question.

Acknowledgments

We wish to thank James Reaux, Molly Bates, Leo Lawson and the rest of the staff of the former Cognitive Evolution Group who assisted in the implementation of this study. We are also deeply grateful

¹⁰ Povinelli and Vonk (2012) have provided an in-depth case study of this issue in the context of the ability of animals to reason about floppy versus rigid tools.

to Caroline Arruda for feedback on an earlier version of this article. The study was supported in part by a Centennial Fellowship to D.J.P. from the James S. McDonnell Foundation. We have no conflicts of interest to report. We dedicate this effort to the memory of Anthony ("Pop") Rideaux whose loving care of the chimpanzees who participated in this study will remain stitched in the hearts of those who knew him.

References

- Alem, S., Perry, C. J., Zhu, X., Loukola, O. J., Ingraham, T., Søvik, E., & Chittka, L. (2016). Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect. *PLoS Biology*, *14*(10), e1002564.
- Andrews, K. (2016). Snipping or editing? Parsimony in the chimpanzee mind reading debate. *Metascience*, *25*(3), 377–386.
- Arruda, C. T., & Povinelli, D. J. (2016). Chimps as secret agents. *Synthese*, *193*(7), 2129–2158.
- Barker, K. B. & Povinelli, D. J. (2019). Anthropomorphomania and the rise of the animal mind: A conversation. *Journal of Folklore Research*, *56*(1–2), 71–90.
- Barrett, L. (2011). *Beyond the brain: How body and environment shape animal and human minds*. Princeton University Press.
- Beck, B. B. (1980). *Animal tool behaviour: The use and manufacture of tools by animals*. Garland STPM Publishing.
- Bentley-Condit, V. K. & Smith, E. O. (2010). Animal tool use: Current definitions and an updated comprehensive catalog. *Behaviour*, *147*(2), 185–221, A1–A32
- Bering, J. M. & Povinelli, D. J. (2003). Comparing cognitive development. In D. Maestriperi (Ed.), *Primate psychology: Bridging the gap between the mind and behavior of human and nonhuman primates* (pp. 205–233). Harvard University Press.
- Bird, C. D., & Emery, N. J. (2009a). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences*, *106*(25), 10370–10375.
- Bird, C. D., & Emery, N. J. (2009b). Rooks use stones to raise the water level to reach a floating worm. *Current Biology*, *19*(16), 1410–1414
- Boesch, C., & Boesch-Achermann, H. (2000). *The chimpanzees of the Tai Forest: Behavioural ecology and evolution*. Oxford University Press.
- Brach, T. (2004). *Radical acceptance: Embracing your life with the heart of a Buddha*. Bantam Books.
- Butovskaya, M. L., & Kozintsev, A. G. (1996). A neglected form of quasi-aggression in apes: Possible relevance for the origins of humor. *Current Anthropology*, *37*(4), 716–717.
- Channel 4 News (2011, June 9). *Urinating chimp is 'scientifically interesting'* [Video]. YouTube. <https://www.youtube.com/watch?v=vTgihteq1MU&t=2s>
- Chemero, A. (2011). *Radical embodied cognitive science*. MIT press.
- Clark, A., & Thornton, C. (1997). Trading spaces: Computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences*, *20*, 57–90.
- Clatterbuck, H. (2018). The logical problem and the theoretician's dilemma. *Philosophy and Phenomenological Research*, *97*(2), 322–350.
- Cognitive Evolution Group (2010a). *Study review coordinator manual*. rev. 4.0. University of Louisiana.
- Cognitive Evolution Group (2010b). *Manual of standard operating procedures for chimpanzee cognition laboratory*, rev. 7.2. University of Louisiana.
- Dacey, M. (2016). The varieties of parsimony in psychology. *Mind & Language*, *31*(4), 414–437.
- Darwin (1871). *The descent of man*. [Reprinted Modern Library, New York, 1982]
- Dennett, D. (2009). Darwin's "strange inversion of reasoning". *Proceedings of the National Academy of Sciences*, *106* (Supplement 1), 10061–10065.
- Ebel, S. J., Schmelz, M., Herrmann, E., & Call, J. (2019). Innovative problem solving in great apes: The role of visual feedback in the floating peanut task. *Animal Cognition*, *22*(5), 791–805.
- Fayet, A. L., Hansen, E. S., & Biro, D. (2020). Evidence of tool use in a seabird. *Proceedings of the National Academy of Sciences*, *117*(3), 1277–1279.
- Fragaszy, D. M., & Mangalam, M. (2020). Folks physics in the twenty-first century: Understanding tooling as embodied. *Animal Behavior and Cognition*, *7*(3), 457–473.
- Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

- Ghirlanda, S., & Lind, J. (2017). 'Aesop's fable' experiments demonstrate trial-and-error learning in birds, but no causal understanding. *Animal Behaviour*, *123*, 239–247.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Harvard.
- Gruber, R., Schiestl, M., Boeckle, M., Frohnwieser, A., Miller, R., Gray, R. D., Clayton, N. S., & Taylor, A. H. (2019). New Caledonian crows use mental representations to solve metatool problems. *Current Biology*, *29*(4), 686–692.
- Hanus, D., & Call, J. (2011). Chimpanzee problem-solving: Contrasting the use of causal and arbitrary cues. *Animal Cognition*, *14*(6), 871–878.
- Hanus, D., Mendes, N., Tennie, C., & Call, J. (2011). Comparing the performances of apes (*Gorilla gorilla*, *Pan troglodytes*, *Pongo pygmaeus*) and human children (*Homo sapiens*) in the floating peanut task. *PLoS One*, *6*(6), e19555.
- Hennefield, L., Hwang, H. G., Weston, S. J. & Povinelli, D. J. (2018). Meta-analytic techniques reveal that corvid causal reasoning in the Aesop's Fable paradigm is driven by trial-and-error learning. *Animal Cognition*, *21*(6), 735–748.
- Herrmann, E., Wobber, V., & Call, J. (2008). Great apes' (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, *Pongo pygmaeus*) understanding of tool functional properties after limited experience. *Journal of Comparative Psychology*, *122*(2), 220–230.
- Hicks, T. C., Köhl, H. S., Boesch, C., Dieguez, P., Ayimisin, A. E., Fernandez, R. M., Zungawa, D. B., Kambere, M., Swinkels, J., Menken, S. B. J., Hart, J., Mundry, R., & Roessingh, P. (2019). Bili-Uéré: A chimpanzee behavioural realm in northern democratic republic of Congo. *Folia Primatologica*, *90*, 3–64.
- Hobhouse, L. T. (1915). *Mind in evolution*. Macmillan.
- Hoerl, C., & McCormack, T. (2019). Thinking in and about time: A dual systems perspective on temporal cognition. *Behavioral and Brain Sciences*, *42*, e244: 1–69.
- Holzhaider, J. C., Hunt, G. R., Campbell, V. M., & Gray, R. D. (2008). Do wild New Caledonian crows (*Corvus moneduloides*) attend to the functional properties of their tools? *Animal Cognition*, *11*(2), 243–254.
- Hopkins, W. D., Bard, K. A., Jones, A., & Bales, S. L. (1993). Chimpanzee hand preference for throwing and infant cradling: Implications for the origins of human handedness. *Current Anthropology*, *34*, 786–790.
- Horner, V., & Whiten, A. (2007). Learning from others' mistakes? Limits on understanding a trap-tube task by young chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Journal of Comparative Psychology*, *121*(1), 12–21.
- Hunt, G. R., & Gray, R. D. (2004). The crafting of hook tools by wild New Caledonian crows. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl 3), S88–S90.
- Hunt, G. R., Gray, R. D., & Taylor, A. H. (2013). Why is tool use rare in animals. In C. M. Sanz, J. Call, & C. Boesch (Eds.) *Tool use in animals: Cognition and ecology*, (p. 89-118). Cambridge University Press.
- Jacobs, I. F., & Osvath, M. (2015). The string-pulling paradigm in comparative psychology. *Journal of Comparative Psychology*, *129*(2), 89–120.
- Jelbert, S. A., Taylor, A. H., Cheke, L. G., Clayton, N. S., & Gray, R. D. (2014). Using the Aesop's fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. *PLoS one*, *9*(3), e92895.
- Klüver, H. (1933). *Behavior mechanisms in monkeys*. University of Chicago Press.
- Köhler, W. (1925). *Intelligenzprüfungen an Anthropoiden* [The Mentality of Apes]. Kegan Paul. (E. Winter, Trans., 2nd revised ed.). (Original work published 1917)
- Mann, J., & Patterson, E. M. (2013). Tool use by aquatic animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1630), 20120424.
- Martin-Ordas, G., Call, J., & Colmenares, F. (2008). Tubes, tables and traps: Great apes solve two functionally equivalent trap tasks but show no evidence of transfer across tasks. *Animal Cognition*, *11*(3), 423–430.
- Melis, A. P., Hare, B., & Tomasello, M. (2006). Engineering cooperation in chimpanzees: Tolerance constraints on cooperation. *Animal Behaviour*, *72*(2), 275–286.
- Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: Orangutans use water as a tool. *Biology Letters*, *3*(5), 453–455.
- Mulcahy, N. J., Call, J., & Dunbar, R. I. (2005). Gorillas (*Gorilla gorilla*) and orangutans (*Pongo pygmaeus*) encode relevant problem features in a tool-using task. *Journal of Comparative Psychology*, *119*(1), 23–32.
- Newen, A., & Starzak, T. (in press). How to ascribe beliefs to animals. *Mind & Language*.
- Newmyer, S. T. (2005). Tool use in animals: Ancient and modern insights and moral consequences. *Scholias: Studies in Classical Antiquity*, *14*(1), 3–17.

- Nissani, M. (2006). Do Asian elephants (*Elephas maximus*) apply causal reasoning to tool-use tasks? *Journal of Experimental Psychology: Animal Behavior Processes*, 32(1), 91–96.
- Patterson, E. M., & Mann, J. (2011). The ecological conditions that favor tool use and innovation in wild bottlenose dolphins (*Tursiops* sp.). *PLoS One*, 6(7), e22243.
- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008a). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31, 109–130.
- Penn, D. C., Holyoak, K. J. & Povinelli, D. J. (2008b). Darwin's triumph: Explaining the uniqueness of the human mind without a *deus ex machina*. *Behavioral and Brain Sciences*, 31, 153–178.
- Penn, D. C., & Povinelli, D. J. (2013). The comparative delusion: Beyond behavioristic and mentalistic explanations for nonhuman social cognition. In H. S. Terrace & J. Metcalfe, (Eds.), *Agency and joint attention* (p. 62–81). Oxford University Press.
- Povinelli, D. J. (1988). Intention and attribution in primates: A new quest for mind. *Yale Graduate Journal of Anthropology*, 1, 37–48.
- Povinelli, D. J. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works*. Oxford University Press. [Reprinted, 2003 with revisions and a new Forward by Nicholas Humphrey]
- Povinelli, D. J. (2012). *World without weight: Perspectives on an alien mind*. Oxford University Press.
- Povinelli, D. J., & Barker, K. B. (2019). Doctor Fomomindo's preliminary notes for a future index of anthropomorphized animal behaviors. *Journal of Folklore Research*, 56(1-2), 125–291.
- Povinelli, D. J., & Eddy, T. J. (1996). *What young chimpanzees know about seeing*. Monographs of the Society for Research in Child Development, 61(3), v–vi, 1–152.
- Povinelli, D. J., & Frey, S. (2016). Constraints on the exploitation of the functional properties of objects in expert tool-using chimpanzees (*Pan troglodytes*). *Cortex*, 82, 11–23.
- Povinelli, D. J., & Giambone, S. (1999). Inferring other minds: Failure of the argument by analogy. *Philosophical Topics*, 27, 167–201.
- Povinelli, D. J., & O'Neill, D. K. (2000). Do chimpanzees use their gestures to instruct each other? In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.). *Understanding other minds: Perspectives from autism*, 2nd ed. (pp. 459–487). Oxford University Press.
- Povinelli, D. J., & Penn, D. C. (2011). Through a floppy tool darkly: Toward a conceptual overthrow of animal alchemy. In T. McCormack, C. Hoerl, & S. Butterfill (Eds.), *Tool use and causal cognition* (pp. 69–88). Oxford University Press.
- Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, 19, 1–28.
- Povinelli, D. J., & Vonk, J. (2012). Case study: Levels of understanding floppy and rigid tools. In D. J. Povinelli (Ed.), *World without weight: Perspectives on an alien mind* (pp. 308–318). Oxford University Press.
- Riemer, S., Müller, C., Range, F., & Huber, L. (2014). Dogs (*Canis familiaris*) can learn to attend to connectivity in string pulling tasks. *Journal of Comparative Psychology*, 128(1), 31–39.
- Rutz, C., & Deans, S. (2018). Nuthatch uses tool in London park. *Ethology*, 124(2), 135–138.
- Santos, L. R., Rosati, A., Sproul, C., Spaulding, B., & Hauser, M. D. (2005). Means-means-end tool choice in cotton-top tamarins (*Saguinus oedipus*): Finding the limits on primates' knowledge of tools. *Animal Cognition*, 8(4), 236–246.
- Sanz, C. M., & Morgan, D. B. (2009). Flexible and persistent tool-using strategies in honey-gathering by wild chimpanzees. *International Journal of Primatology*, 30(3), 411–427.
- Schrauf, C., & Call, J. (2011). Great apes use weight as a cue to find hidden food. *American Journal of Primatology*, 73(4), 323–334.
- Seed, A. M., Call, J., Emery, N. J., & Clayton, N. S. (2009). Chimpanzees solve the trap problem when the confound of tool-use is removed. *Journal of Experimental Psychology: Animal Behavior Processes*, 35(1), 23–34.
- Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior: The use and manufacture of tools by animals*. Johns Hopkins University Press.
- Sober, E. (2012). Anthropomorphism, parsimony, and common ancestry. *Mind & Language*, 27(3), 229–238.
- Sober, E. (2016a). *Ockham's razors: A user's manual*. Cambridge University Press.
- Sober, E. (2016b). Replies to Kristin Andrews's, Gordon Belot's, and Patrick Forber's reviews. *Metascience*, 25, 393–403
- St Clair, J. J., & Rutz, C. (2013). New Caledonian crows attend to multiple functional properties of complex tools. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1630), 20120415.
- Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2009). Do New Caledonian crows solve physical problems through causal reasoning?. *Proceedings of the Royal Society B: Biological Sciences*, 276(1655), 247–254.

- Taylor, A. H., Medina, F. S., Holzhaider, J. C., Hearne, L. J., Hunt, G. R., & Gray, R. D. (2010). An investigation into the cognition behind spontaneous string pulling in New Caledonian crows. *PloS one*, 5(2).
- Tomasello, M., & Call, J. (2006). Do chimpanzees know what others see—or only what they are looking at. In S. E. Hurley & M. E. Nudds (Eds.), *Rational animals?* (pp. 371–384). Oxford University Press.
- van Lawick-Goodall, J. (1968). The behaviour of free-living chimpanzees in the Gombe Stream reserve. *Animal Behaviour Monographs*, 1(3), 161–311.
- van Leeuwen, E. J., Cronin, K. A., & Haun, D. B. (2017). Tool use for corpse cleaning in chimpanzees. *Scientific Reports*, 7, 44091.
- Visalberghi, E., Addessi, E., Truppa, V., Spagnoletti, N., Ottoni, E., Izar, P., & Frigaszy, D. (2009). Selection of effective stone tools by wild bearded capuchin monkeys. *Current Biology*, 19(3), 213–217.
- Visalberghi, E., & Limongelli, L. (1994). Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 108, 15–22.
- Visalberghi, E., & Trinca, L. (1989). Tool use in capuchin monkeys: Distinguishing between performing and understanding. *Primates*, 30, 511–521.
- Weir, A. A., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science*, 297(5583), 981–981.
- Weir, A. A., & Kacelnik, A. (2006). A New Caledonian crow (*Corvus moneduloides*) creatively re-designs tools by bending or unbending aluminium strips. *Animal Cognition*, 9(4), 317–334.
- Whitt, E., Douglas, M., Osthaus, B., & Hocking, I. (2009). Domestic cats (*Felis catus*) do not show causal understanding in a string-pulling task. *Animal Cognition*, 12(5), 739–743.
- Zhang, C. Y., Mandel, J. E., Pruett, J. Jr. & Povinelli, D. J. (2016, October 20). *Theory of mind priming pilot: Rating the ambiguity of revised line of sight probe task stimuli* [Poster presentation]. Second-Year Medical Student Research Symposium, Washington University Medical School.

Appendix 1

Toward the Reinvention of the Study of Higher-Order Reasoning in Animals

One of the reviewers of our article, after expressing sympathy with our main thesis, did not wish to let go of the rope completely. Surely somewhere out in the vast, existing genre of comparative psychology there must be an experimental protocol that could allow for a strong inference about the operation of higher-order causal reasoning in animals. Thinking aloud, the reviewer gestured to a familiar study involving a vertical transparent tube with a peanut at the bottom that can be retrieved only by filling the tube with water (see Mendes et al., 2007; see also Ebel et al., 2019; Hanus et al., 2011). The reviewer wondered if, with strict parameters, a cleaned-up variant of such a test could suffice: (a) the peanut would be out of reach at the bottom of the tube, (b) the tube would have no water inside (presumably to rule out some arcane associative learning explanation), and (c) the animal must never before have been presented with this set-up “or anything similar” (a familiar refrain when setting up such tests).¹¹ In the reviewer’s imagination, the animal, thus constrained, would have only one option to retrieve the peanut: they would need to “float it out” by using water from the cage’s drinking nozzle. If they did so, the reviewer wondered, wouldn’t this constitute a valid investigation of higher-order reasoning? After all, there would be no previous “learning history” or “clues within the task” to explain the performance. Finally, if this task does not get at higher-order reasoning, the reviewer asked, could we suggest one that would? Below, we address both of these questions. We end with a protocol for how comparative psychology can move forward.

We confess that after almost two decades of articulating why each extant claim for higher-order, role-based, relational reasoning in animals (e.g., theory of mind, abstract same/different judgments, analogical and hierarchical reasoning, logical rules, etc.) is subject to the flaw elucidated in our main report, we do sometimes tire of engaging in what some see as a parlor trick. Still, there are several reasons why in this case we are actually happy the reviewer floated one (final?) task variant to disrobe. First, the reviewer’s request offers a chance to show that we are not cherry-picking tasks especially vulnerable to our arguments. Second, it offers another case study of why pitting ill-specified models of “learning history” against ill-specified “higher-order” alternatives have impeded progress in this area. Finally, it offers one more opportunity for the sympathetic reader to assess our claim that the asymmetric dependency relationship between first-order, perceptually-based relational reasoning and higher order-order, structural, role-based relational reasoning *systematically applies to all tasks that fall under the scope of the genre we have identified* (and thus, with luck, allow any such reader to reject the incoherent, but frequent, complaint that our explanations are *ad hoc*; see footnote 9).

With that introduction, let us now formally lay out the best version of the reviewer’s inductive argument about the spitting-in-a-tube task’s relationship to higher-order reasoning we can muster. (NB: The text in bold refers to what we think are suppressed premises in the reviewer’s thinking that we believe need to be made explicit in order to construct the best version of their argument.) The imagined subject of the study will be our chimpanzee, Apollo:

¹¹ In suggesting this constraint, the reviewer did not engage with our detailed analysis as to why it is probably impossible to present an organism with stimuli that bears no perceptual similarity to anything they have encountered in their developmental or evolutionary history (see main text). As we point out, the genre of comparative psychology protocols that are designed to investigate higher-order reasoning (including this reviewer’s proposal) all rest upon scientific folklore: the idea that the experimenter can intuitively create stimuli that are *similar enough* to things the subjects already know about/represent, but *different enough* to allow just enough daylight between the competing explanations to squeeze out an inference that a certain pattern of results (e.g., spitting in a tube) implicates the necessity of higher-order thinking. The reviewer is not alone. Indeed, Mendes et al. (2007) believed they had done precisely this in their original version of the spitting-in-a-tube task: “One alternative to an insightful solution is that subjects previously solved this problem and here they simply remembered the solution. Although we cannot rule this out, we think that it is unlikely. Subjects had never received this task in the past and water was never required as a solution for a problem” (p. 454). (NB: “Insightful solutions” and higher-order causal reasoning are not synonymous, nor do Mendes et al. [2007] claim they are.)

- P1 Apollo possesses the requisite knowledge and skill to suck water out of a nozzle **as well as to spit it out;**
- P2 **Specifically, Apollo possesses the requisite knowledge and skill to spit water into a tube (or, if one prefers, “for Apollo, the tube affords water spitting” or even, “the water in Apollo’s mouth affords tube-spitting);**
- P3 Apollo has never been presented with this setup “or anything similar” before;
- C If Apollo spits water into the tube and the peanut rises to within reach, his cognitive system must have deployed higher-order reasoning related to constructs such as *<weight>* or *<buoyancy>* or *<floating>*.

Beyond the obvious tension between P2 and P3, it can be quickly shown that this argument is simply a variant of the weak (flawed) argument we detailed in our main report. Even if one granted (which we do not, see below) that the only possible goal-directed description of Apollo’s spitting behavior is: *<spit water in tube to make peanut rise>* (or, put more richly, “Apollo imagines spitting water into the tube, the tube filling with water, and the peanut rising”), C would still be just as weak as the inferences illustrated in our main report. The perceptual representations (sucking water out of tube, filling mouth with water, spitting water into tube, water rising in tube, etc.) are both necessary and sufficient to execute the alleged goal-directed action. Consider the issue starkly: if Apollo did not possess such stable perceptual representations, what causal role could ungrounded (dare we say “free-floating”), higher-order constructs play in his cognitive economy? This bespeaks the central problem we point out in the main text: once we accept the existence of first-order perceptual representations (an uncontroversial construct in current representational theories of cognition; see Penn et al., 2008a), it is incumbent upon us to take their causal role seriously.¹²

In addition to the fatal analysis above, it is important to note that the claim that when Apollo spits into the tube, he is intending to make the peanut rise is assumed, not proven. Thus, any inferences that rely on this premise are based upon circular reasoning. The assumption is that if Apollo is forced into the carefully orchestrated, tightly constrained situation dreamt up by the reviewer (a situation previously dreamt up by Mendes et al., 2007), and then, at some point, he performs the target behavior (spitting water in the tube), we are secure in concluding that he intended to produce the specific perceptual causal fact about the world of interest to the experimenter (the peanut rising). But such an assumption is unwarranted. Apollo’s behavioral propensities were straightjacketed from the beginning: he cannot use his fingers to work the peanut from the tube (the tube has been designed to prevent his apishly large fingers from fitting inside); he cannot lift the tube and shake the peanut onto the floor (the tube is anchored with chimp-proof nuts and bolts to the wall); he cannot use a stick to fish it out (all stick-like objects have been carefully removed). Thus, after some time, Apollo executes the only goal-directed action related to the food that remains in his repertoire: *<spit water into the tube>*.¹³ The constrained

¹² Of course, one could reject the standard representationalist framework and adopt a radical version of embodied or enactivist view of “cognition” that denies the role of mental representations altogether (*cf.* Chemero, 2011; Gallagher, 2017). Under such a view, however, the causal role of higher-order constructs would have even less of a place in explaining Apollo’s behavior. On the other extreme, one might call for a view of cognition that denies the distinction between first-order, perceptually-based relations and higher-order, structural, role-based constructs and relations, or claims that in practice these distinctions cannot be made. Before one charges headfirst in that direction, however, it is worth noting even if one were to accept such a view as coherent (which we do not), there would be little reason to conduct experiments on animals in the first place. We could simply ask human subjects to describe particular physical or social phenomenon to which animals are naturally responsive. If the human-derived descriptions contained traces of higher-order thinking, we would, on this view, be licensed to assume the animals themselves also possess such higher-order thinking. To illustrate, when humans observe a video of Apollo using a stick to rake in a banana, and they describe his actions using language about space and time, we would be able to conclude that Apollo thinks about *<space>* and *<time>*. We advocate resisting such an approach. Instead, we advocate heeding a repurposed version of Dennett’s (2009) earlier admonition: In order to behave in space and time it is not requisite to know a thing about *<space>* or *<time>*.

¹³ Strictly speaking, this is untrue. In a follow up study to Mendes et al.’s (2007) study with orangutans, Hanus et al. (2011) tested chimpanzees and gorillas on slight variants of the spitting-in-a-tube task. At least one chimpanzee in their study developed the solution of urinating into the tube to get the peanut (for video, see Channel 4 News, 2011). To the uninitiated reader, a chimp

situation now dictates that the peanut will now move toward him. Once this causal fact is discovered, Apollo repeats his action.¹⁴ Although we cannot prove this, we are even willing to grant the very likely possibility that Apollo's spitting behavior is goal-directed with respect to the peanut. But why must we assume that his goal is to make the peanut rise as the waters crest inside the tube (let alone to make the peanut <float>). His goal may be to use the water to make contact with the peanut in much the same manner as if the experimenters had not deprived him of his sticks. Given that we are writing this article to celebrate the twentieth anniversary of the publication of our book, *Folk Physics for Apes*, it is probably worth noting that one of the central conclusions of that project was that an irreducible element of the chimpanzee's first-order representation of tools is the ability of one object (the tool) to make *perceptible contact* with another object (e.g., a food reward) [see Povinelli, 2000, Chapter 12, pp. 303-308; see also Hennefield et al., 2018].) (NB: In the context of the asymmetric dependency problem, any higher-order claim that Apollo's spitting behavior is a goal-directed act aimed at making the peanut <float> to within reach, entails that the perceptual claim that he knows how to make the water contact the food. Whether he initially integrates this with the perceptual fact that such contact will lead to the upward motion of the peanut is beside the general point we are making. In either case, the first-order, perceptual-based components of his reasoning are both necessary *and sufficient* to explain his initial and subsequent spits. Thus, we do not need to appeal to any additional explanatory devices.¹⁵

At this juncture, we are used to intelligent minds (even our own) still resisting: *But wait, why would Apollo spit water into the tube unless he knew that it would make the peanut float? And <floating> and <sinking> are higher-order constructs related to variables like <weight> and <buoyancy>. So, by spitting into the tube, Apollo must be reasoning about such things.* As we have just seen, however, we ought to calm such folk thinking with a rational response: *Those thoughts are circular. They assume the presence of the very thing for which we were claiming to test.* The naked fact that humans possess an explanatory drive that asks *why* observable events occur, says nothing about whether Apollo must also possess a reason for his repeated spitting beyond the fact that it produces the effects he intends. Clearly, then, our answer to the reviewer's first question is no, the spitting-in-a-tube task does not offer any hope of assaying higher-order thinking. First-order, perceptually-based, relational reasoning is both necessary and sufficient to explain Apollo's intelligent actions in this task.

We thus arrive at the reviewer's final question: Can we suggest a task that *would* assay higher-order reasoning? This is a tricky question (at least as tricky as the history of the rope tricks that motivated our original study). On the one hand, we are tempted to say, *yes, we can*, and direct the interested reader to the list of seventeen experimental protocols we outlined in the appendix of Penn et al. (2008b, pp. 165-169). It should be noted, however, that one of the architects of those protocols (DJP) is even less sanguine now than he was then about the ability to overcome the asymmetric dependency problem within the standard experimental paradigm of comparative psychology. Most (if not all) of those protocols clearly fall prey to the analyses herein. And some are scaled-up variants of protocols already known to be of great difficulty for animals. So the simple answer is, *no, we cannot outline a standard protocol in this genre that directly overcomes the asymmetric dependency problem.* Indeed, our ropes report was, until now, just one among dozens of lonely residents in an orphanage of abandoned experiments (involving

urinating in the middle of an experimental protocol might seem to muddy the waters. To the contrary, we believe it highlights the ubiquity of the first-order, perceptual representations that organisms must carry inside them.

¹⁴ This matches the results with the orangutans reported by Mendes et al. (2007) (see also Ebel et al., 2019). On average, the orangutans first spat in the tube 540 s (9 min) after it was first presented to them. The time to subsequent spits decreased exponentially.

¹⁵ Those who have spent considerable time around apes in captivity will, from time to time, find themselves recounting tales of how these animals cleverly fill their mouths with water and wait for an unsuspecting dupe to pass before their enclosure. The backwards leap of such hapless humans, coupled with their ripe, expletive-filled exclamations, is a causal fact about the world that captive chimpanzees know a great deal. Indeed, spitting water (or flinging feces) to create the movement of distal humans is a favorite goal-directed activity of many captive apes (see Hopkins et al., 1993). For those who wish to hold out for a higher-order, structural, role-based interpretation on such behavior, we suggest consulting Butovskaya and Kozintsev's (1996) analysis of both of these actions as a form of proto mockery. (To be fair, tales of being doused by spat water (or flung feces) are told by many a casual zoo visitor, as well).

both chimpanzees and young children) that have remained unpublished because we long ago realized they suffer from a common flaw: the asymmetric dependency problem. We are well aware this answer will not satisfy experimentalists whose livelihood and reputations depend upon, well, *conducting experiments*. Indeed, lectures on the asymmetric dependency problem are frequently met with an indignant reaction: *Well, then what are we supposed to do?*

This frustration has led to a reverse parlor trick in which philosophers and experimentalists propose new experiments they believe can implicate the necessity of higher order reasoning in animals. Elsewhere, we have reviewed the fate of these proposals. Rather than spelling out our critiques of each of these again, and risk this publication becoming even more recursive than it already is, it may be more productive to examine the lessons learned from one of these scholars, Elliot Sober, who elected to enter the game. After acknowledging the serious implications that our analysis poses for designing experiments that can provide good evidence that chimpanzees possess a theory of mind—implications we have hopefully generalized and expanded upon in this article—Sober (2016a) proposed a “two-winged” experiment involving chimpanzees navigating their way through opaque tunnels and trap doors affixed with buzzers and bells. For present purposes, the details do not matter. What is instructive is Sober’s (2016a) partial concession to Andrews’ (2016) objection that his proposal would not actually overcome the asymmetric dependency problem (the so-called “logical problem”):

In designing the two-winged experiment, I wanted to find two tasks that would “look different” to chimpanzees if they were behavior readers, but might “look the same” if they were mind readers. It may be that the tasks I chose were not different enough. Opaque and transparent tunnels seem rather different from silent and noisy trapdoors, but both tasks involve obtaining food by reaching into a chamber where a human being is present. This may mean that the details of the experiment I described need to be tweaked. I nonetheless hope that the two-task *n*-trial format will be useful in cracking what is in fact a very hard nut (p. 399).

Although we do not have the space to trace out the full implications of this concession, we will at least note that it highlights two core reasons why any future variants of his experimental design are destined to fail him. First, the design rests upon the unprincipled titration we have repeatedly stressed: designing a task (or two paired tasks) that is (or are) robustly accessible to the subjects’ perceptual representations, and yet are not *so accessible* that perceptual-based reasoning can fully explain the results. Second, and related, his approach pits a higher-order account (in this case, theory of mind) against a lower-order, perceptually-based account, without explicitly considering the full implication of the fact that *any purported higher-order thinking on the part of the apes necessarily requires robust, paired perceptual representations from the lower-order system as input* (i.e., the asymmetric dependency problem).¹⁶

¹⁶ For those who wish to puzzle over the relationship between the details of Sober’s (2016a, b) proposed experiment and the asymmetric dependency problem, we can think of no better place to begin than the rhetorical question posed by Sober (2016a): “If chimpanzees are not mind readers, why should their thinking ‘I will get food if I use the opaque tunnel, but not if I use the transparent tunnel’ raise the probability of their thinking ‘I will get food if I use the silent trapdoor, but not if I use the noisy trapdoor’?” (p. 398). That this is an erotetic move and not a serious question, is suggested by the simplicity of the answer: *Because in both cases the experimenter turns toward Apollo and snatches away his food!* And, once again, before rushing off to try to patch up this fatal flaw (such as tweaking the exact manner in which the human <notices> the chimpanzee), it should be noted that this would merely recapitulate the individual histories of the innumerable task variants strewn throughout this genre of comparative psychology—a process that always ends in erotesis: *Well, how else could they do it?* Perhaps the greatest irony of deploying such rhetorical devices occurs when the user then immediately proceeds to admonish us not to “concoct” our representational claims “out of thin air” and to be sure they “have some antecedent and independent plausibility in cognitive science” (Sober, 2016a, p. 398). For what it’s worth, we are quite comfortable letting the reader decide whether our ecumenical assumptions about cognition are rooted in mainstream cognitive science (see Penn et al., 2008). To be clear, we are not faulting Sober (2016a,b) and others for wanting to play the parlor game. It’s fun. We know because our own investigations of higher-order reasoning in chimpanzees followed the same line of reasoning. First, we designed tasks that we (as adult humans) could interpret in a higher-order manner. Next, we let our chimpanzees solve the task. After that, we presented them with “novel” (read: “perceptually different”) variants of the task that we could also interpret in the same higher-order manner. Our reasoning? If the chimpanzees could cognize a higher-order relation common to the two tasks, despite the perceptual differences, they ought to exhibit rapid transfer—after all, *How else could they do it?* (see Povinelli, 2012, for an exhausting case study of this approach in

The good news is that we will now end on a positive, forward-looking note. Despite not having a secret formula that can rescue us from our own analysis, we do possess a nuanced, forward-looking, multi-phase protocol for how comparative psychologists who are interested in higher-order reasoning could spend their time. In phase one, psychologists would practice radical acceptance (see Brach, 2004), deeply absorbing the truism that recognizing a problem is distinct from having a solution. The experiments of comparative psychologists are situated within a representationalist tradition whose bearers seek to discern the nature of the mental states that animals possess. And although the precise views of individual practitioners may differ, the tradition carries infrequently articulated assumptions about the necessary and sufficient conditions for higher-order thinking. Once these suppressed assumptions are made explicit (see Penn et al., 2008a), difficult choices arise. As we explained in footnote 13, we are not suggesting that there are no alternatives to a representationalist stance toward animal cognition. Fragaszy and Mangalam's (2020) essay in this special issue explores one such view (see also Barrett, 2011). But we are suggesting that, properly understood, these alternatives are even less friendly toward the current *experimental project* of trying to demonstrate higher-order thinking in animals than ours. To summarize, phase one requires making one's commitments about cognition explicit, beginning and ending with an account of how the sensory apparatuses of organisms under study interact with the rest of the body and brain to generate a temporally stable basis for behavior (see Penn et al. 2008a; Povinelli, 2012).

All comparative psychologists who pass phase one would then advance to phase two. In this phase, practitioners would politely agree to eschew the false dichotomy between "behaviorism", "associative learning", "arbitrary rule learning" on the one hand, and "causal learning" (or some such), on the other. Beginning with *Folk Physics for Apes*, we joined the voices of many others in arguing that such contrasts are too colloquial to make progress in the broader debates (for our contributions to this discussion see Penn et al., 2008a; Penn & Povinelli, 2013; Povinelli, 2000, 2012; Povinelli & Penn, 2011). Ironically, in the context of the structurally similar debate over higher-order social reasoning (i.e., theory of mind), discussed earlier, Andrews (2016) has complained that our first-order representational hypothesis (what has come to be known as the "behavior-reading" hypothesis in the literature) is "a bit of shifting sand that does not manage to stay still for very long" (p. 385), a captivating and poetic restatement of the unfounded and illogical criticism that such explanations are *ad hoc* (see above). Curiously, as he labored over what he described as the "hard nut" of designing an experiment that can isolate the unique causal work of the higher-order system, Sober (2016a) endorsed the shifting sand metaphor. He asserted that while our account explicitly states what is *not* involved in explaining chimpanzee social behavior (i.e., representations of <mental states>, it fails to specify "the causal connections between belief states¹⁷ that behavior-reading hypotheses are permitted to postulate" (p. 399). Alas, in addition to misunderstanding the full scope of the asymmetric dependency problem, this assertion has the additional disadvantage of being demonstrably false. Not only have we *precisely specified* the representational commitments of our hypothesis about human cognition (see Penn et al. 2008a), we are still patiently waiting for a reciprocal gesture on the part of those scholars who are committed to the idea that higher-order representations are necessary to explain the results of the genre of experiments in question. To date, however, instead of laying their cognitive cards on the table, our most vocal critics have instead attempted to saddle our ideas with monikers that bear no relationship to them (e.g., "associative learning", "inflexible rule learning", "modified behaviorism", etc.; see Povinelli, 2012, p. 20-

the context of the higher-order construct of <weight>). Indeed, one of us (DJP) was once so convinced of the merits of this strategy, that for many years he asserted that if a chimpanzee showed evidence that they cognized the contrast between, and the commonality within, two simultaneously-presented, infinitely large sets of perceptually distinct tasks, one set requiring them to discover contingencies involving the color of stimuli and the other set involving the shape of stimuli, this would provide evidence for the higher-order concepts of <color> and <shape>. Careful reflection on the implications of the asymmetric dependency problem has disabused him of this idea.

¹⁷ Nothing in the current debate hinges on whether one chooses to describe these internal representations as "belief states" (as Sober, 2016a does here) or "first-order, perceptual-based representations" (see Arruda & Povinelli, 2016, for one view on why we ought to consider such first-order mental states as functioning like belief and desire states; for a similar conclusion, see Newen & Starzak, 2020).

23, for a dissection of these polemics along with additional references). Sadly, we know first-hand how hard it can be for experimentalists to abandon long-standing, culturally embedded rhetorical devices from their “work products” (see Povinelli, 2000, pp. ix-xiii; Barker & Povinelli, 2019). However, it seems to us that success in phase one (spelling out one’s commitments about the *human* cognitive system) would naturally pave the way for success in phase two. Obviously, those comparative psychologists failing phase two, cannot advance to phase three. Instead, they will be forever fated to chase Andrews’ (2016) shifting piles of sand—but for the opposite reason they imagined.¹⁸

Finally, in phase three, comparative psychologists would, for the time being, end all attempts to test for higher-order reasoning in animals. Instead, they would devote their unique intellectual resources toward understanding how higher-order reasoning modulates adult human behavior in ecologically relevant tasks. Specifically, they would work to identify the causal fingerprints of higher-order thinking in everyday human behaviors, behaviors that can be (and often are) managed by first-order perceptual-based reasoning alone. If techniques to prime human higher-order reasoning could be refined, researchers could then search for law-like impacts on both the speed and accuracy of simple physical or social judgments, as well as other, more temporally extended behavioral topographies. This could be achieved using remarkably simple tasks (for preliminary work along these lines, see Zhang et al., 2016). One exciting possibility is that doing so might upend some of our worst folk intuitions about the impact of higher-order reasoning on everyday behavior (for example, the intuition that it uniformly “improves” performances). Regardless, once identified, such relationships might allow psychologists to identify “causal workspaces” in which higher-order thinking affects the kinds of behaviors we share in common with primates and other taxa (see Povinelli, 2012, Chapter 11, for a brief discussion of this idea).

Although this phase of our protocol may seem unduly aimed at understanding human cognition, it might just create a new genre of experimental procedures in comparative psychology that do not suffer from the asymmetric dependency problem. The hope is that such investigations would be grounded in something more substantial than the folk intuitions of psychologists themselves, intuitions on full display in our sympathetic reviewer’s musings about the spitting-in-a tube task. To be sure, the success of this final phase presupposes an interest in shouldering up against the much harder task administered in phase one: acknowledging that comparative psychology has a serious problem. Widespread interest in doing so would be encouraging. After all, acknowledging one has a problem is often a first step toward recovery.

¹⁸ *Note from DJP:* I have a folk theory about why the criticism that our hypothesis is *ad hoc* appeals to so many thinkers, despite the fact it is so blatantly false. One afternoon a couple of years ago, an enterprising graduate student approached me after I delivered a lecture on this topic in Leipzig, Germany. She explained that she was testing chimpanzees for some facet of higher-order thinking at the Max Planck Institute for Evolutionary Anthropology. She further explained that she was very familiar with the asymmetric dependency problem (the so-called “logical problem”). “Honestly, it keeps me up at night,” she said. “Why is that?” I asked. She replied: “Because all day long I come up with ideas that I am sure overcome the very hard problem you have isolated in our field. But then at night, as I am about to fall asleep, I suddenly realize that, no, my latest protocol suffers from the same problem. It’s like your hypothesis is a kind of magic wand that makes my intricate web of procedures and control conditions just vanish!” I sighed: “I understand. I had the same experience—for twenty years.” My folk theory is this: As one’s mind deploys the same structural system over and over again to work around a problem that is not avoidable using that system, the instability caused in one’s own thinking is attributed to the problem itself, not to one’s own (unexamined) assumptions. I would not be surprised to learn that this phenomenon has a well-known psychological name.