



---

## The ‘Shoulds’ and ‘Coulds’ of Meaningful Failures: Introduction to the Special Issue

David A. Washburn<sup>1\*</sup> and Will Whitham<sup>1</sup>

<sup>1</sup>Georgia State University, Atlanta, United States

\*Corresponding author (Email: [dwashburn@gsu.edu](mailto:dwashburn@gsu.edu))

**Citation** – Washburn, D. A., & Whitham, W. (2018). The ‘shoulds’ and ‘coulds’ of meaningful failures: Introduction to the special issue. *Animal Behavior and Cognition*, 5(1), 1-8. <https://dx.doi.org/10.26451/abc.05.01.01.2018>

**Keywords:** Meaningful failures, Animal cognition, Animal behavior, Comparative psychology, Learning

---

“Failure is not an option.” This phrase, typically associated with Gene Kranz and the National Aeronautics and Space Administration (NASA), can also be an important principle in the interpretation of the behavior of humans and other animals. Patterns of behavior might be interpreted as an organism’s failure to perceive or to respond in ways that are maximally adaptive, efficient, or effective; nevertheless, these same behavioral patterns may themselves be adaptations that have been learned through experience. For example, an animal may respond to a very difficult discrimination problem by giving up, and this learned helplessness could accurately (but incompletely) be characterized as a failure to master the primary discrimination; however, it is only a failure relative to what *could have been learned*. The animal did not fail to learn or to adapt to the situation, but it may have failed to learn the most adaptive or effective response that was available in the situation. If such a failure were to be replicated across conditions or animals, we might conclude that there is something meaningful and interpretable about this particular failure, and make inferences about the animals’ lack of responsiveness to the primary discrimination in question.

Psychological research has benefitted greatly by the examination and interpretation of failures, of course. Visual illusions and other perceptual errors are a type of meaningful failure that reveals fundamental truths about how organisms learn to perceive size, color, distance, object boundaries, and so forth (e.g., Agrillo, Parrish & Beran, 2014; Purves & Lotto, 2011). False memories and related memory failures have provided important insights into how and why organisms remember—and misremember (e.g., Loftus, 2005). Examples of meaningful failures in judgment and decision making, such as those documented by Tversky and Kahneman (e.g., 1974), elucidate the heuristics and biases that underlie choice behavior and that are generally adaptive and effective, but that lead to irrational choices when they are overgeneralized.

Meaningful failures of learning have also been studied and interpreted. In a classic example, Breland and Breland (1961) documented instances in which animals failed to establish or to maintain conditioned responses, despite consistent reinforcement contingencies that favored the desired responses. For example, they attempted to train a raccoon to drop coins into a box. The animal was easily shaped to pick up a coin and, despite some hesitation, to drop the coin in a box; however, when the Brelands rewarded the animal for gathering two coins and dropping them both in the box, the raccoon failed to learn, resisting the reinforcement contingencies and preferring rather to hold the coins and to rub them

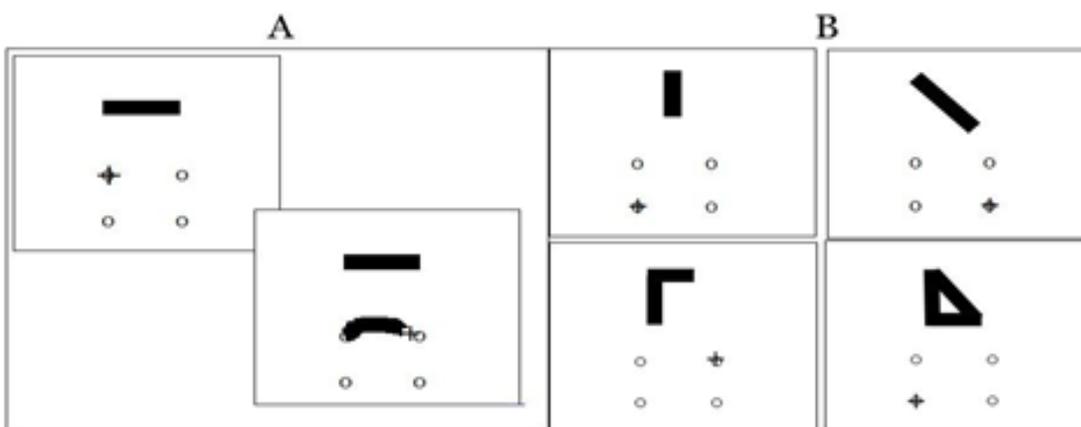
together, despite efforts to extinguish this behavioral pattern. In summarizing a series of similar anecdotes, Breland and Breland inferred a meaningful pattern in the failures, which they termed “instinctive drift,” or the tendency of associative learning to fail, despite clear and consistent contingencies, in the face of competition from strong, instinctual responses. Animals may defy the laws of conditioning and drift back into instinctive behavior patterns.

Extending Breland and Breland’s (1961) important insights about “the misbehavior of organisms” (to quote the title of their classic paper), Rumbaugh and Washburn (2003) proposed “associative drift” as an analogous description for a different group of meaningful failures. Just as organisms may default to instinctive behaviors and fail to learn through operant or respondent conditioning, humans and other animals may at times default to associative learning and fail to learn rule-like, generalized relations that are available to be perceived. That is, animals that are capable of learning relationally sometimes fail to do so, becoming stuck instead in relatively simple stimulus-response-reward learning that is a “failure” only in the sense that a more efficient and effective rule-learning response might have been acquired instead. In some instances, associative drift manifests in positional biases or other stimulus-specific biases that are rewarded with sufficient frequency to sustain the behavior, even when an experimenter attempts to discourage such associative responses (e.g., by ensuring that position, color, and other potential confounds are randomized). In other instances, the associative chain might be more complex and more difficult to control for, because it covaries entirely with the rule-like relations through the training, until generalization tests.

### “Could-learn” Failures

To illustrate this latter point, consider our attempt to teach four adult male rhesus monkeys (*Macaca mulatta*) to draw. The monkeys had been previously trained to respond to computer-generated stimuli by manipulating a joystick, which in turn controlled the movements of a computer-graphic cursor (“+”) on the screen. The monkeys engaged the task ad libitum, working or resting whenever they chose. They were not deprived of food or fluids for purposes of testing, but rather had continuous access to food (primarily, but not exclusively, from the pellets that could be earned through the computer tasks) and water.

For the DRAW task, cursor movements left a trail on the screen, so that one could draw marks on the screen by manipulating the joystick (see Figure 1). The cursor was restricted to moving around the bottom half of the screen, where a four-dot grid was presented for connect-the-dots responding. At the top of the screen, a sample stimulus was presented on each trial. The goal for the task was to train the monkeys to reproduce a likeness of this sample stimulus.

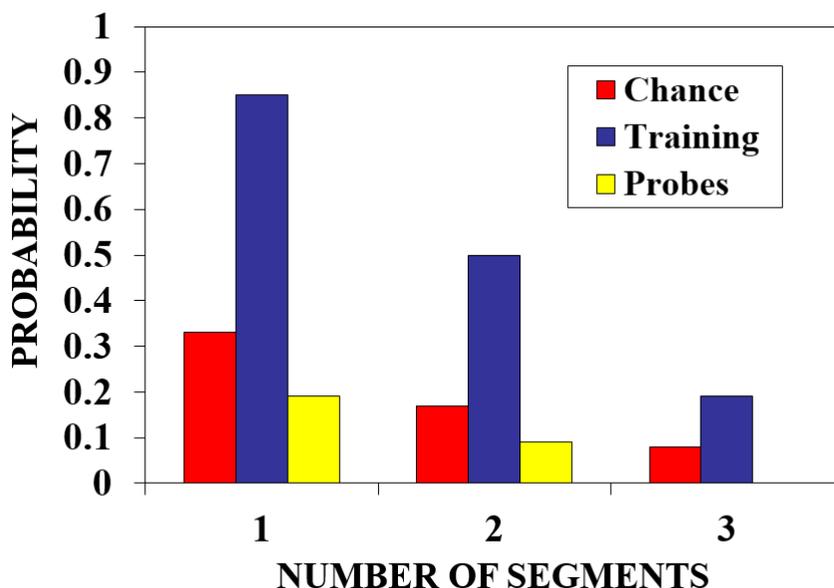


**Figure 1.** Panel A: Schematic of a trial, showing the sample and a cursor start position (top), and a simulated correct response (bottom). Panel B: Four other sample stimuli and cursor start positions, including diagonal, two-line, and three-line probe stimuli.

For training, the sample stimulus was either a horizontal or a vertical line. Each trial began with presentation of one of these two sample stimuli, randomly selected. The cursor began each trial positioned atop one of the four dots in the response grid, randomly selected. To earn a reward, the monkey had to direct the joystick into the dot on the grid that would reproduce the sample line; for example, if the sample was the horizontal line and the cursor began on the top-left grid-dot, the monkey would be reinforced for bringing the cursor into contact with top-right grid-dot, and not rewarded if the cursor contacted any other option. If the sample was a vertical line and the cursor began on the top-right grid-dot, only a response to bottom-right grid-dot was reinforced. As the cursor moved, it left a trail or line on the screen, such that a direct path to the correct grid-dot resulted in a reproduction of the sample; however, the monkeys could (and often did) take a more circuitous route—potentially moving all around the bottom half of the screen, in fact—before touching one of the grid-dots. Across training trials, the monkeys quickly learned which dot(s) to touch with the cursor as a function of sample image and starting location, and indeed with practice the animals got increasingly efficient at connecting the dots with straight lines, so that increasingly accurate versions of the sample image were drawn.

To determine whether the monkeys were using stimulus-response associations or a “draw the sample” rule for responding, probe trials were introduced after the animals were at least 80% accurate on both sample images (chance = 33%). The probe trials were diagonal samples (/ or \). With two probe stimuli and four potential starting locations, each monkey thus received eight novel probe trials, interspersed randomly within the horizontal and vertical trial types. After testing, training continued until the monkeys exceeded 80% accuracy on all four stimulus types, and then novel two-line probe trials were interspersed randomly within the horizontal, vertical and diagonal trials. The probability of getting one of these two-line probes correct by chance alone is only 0.17. After the eight novel probe trials per monkey had been obtained, training continued until performance exceeded 50% accuracy on the two-line problems, whereupon three-line probes were introduced randomly within the block of horizontal, vertical, diagonal, and two-line stimuli.

Note in this paradigm that the monkeys could learn how to respond accurately through stimulus-response associations (e.g., “if sample = X and cursor = location Y, then move to dot Z”), although these associations would provide no information for accurate responding on the novel probe trials. Alternatively, the animals might have perceived the relational solution to the task (e.g., “draw the sample”), which would provide a basis for generalizing to new trials. The results of this training and testing are summarized in Figure 2.



**Figure 2.** Mean probability of a correct response (and chance levels) on training and novel probe trials, as a function of the number of line segments in the sample stimulus).

The results of this study clearly show that (a) the monkeys learned to respond correctly to the various stimuli at levels significantly in excess of chance, and (b) that they did so by stimulus-response associative learning rather than by mastering the “draw the sample” generalized concept. Across trials, the animals produced many “drawings” that were reasonable approximations of the various sample stimuli; however, when presented with a novel image to reproduce, the monkeys did not generalize, but tended to select one of the grid-dots that had been associated with reward. On the diagonal-line probes, for example, the four monkeys connected the correct dots on only 6 of the 32 total novel trials, for instance, and no individual monkey produced more than two of these correct responses.

It is certainly possible that the study would have turned out differently with more training, more samples, different conditions, or different monkeys. It would be unwise to claim that rhesus monkeys cannot learn to draw, based on these findings alone. However, these animals received iterative training-and-testing cycles (albeit with increasingly difficult stimuli), and showed no evidence of incremental improvement or insight in their probe-trial responses. To the contrary, additional experience only seemed to have strengthened the tendency to treat this task associatively, given the relatively small number of trial types and the high number of training trials. This seems to qualify as a “could-learn” type of meaningful failure: Drawing, as operationalized here, is a competency that the rhesus monkeys might reasonably have demonstrated, but it is also possible that the cognitive demands of drawing, which has been demonstrated by apes (e.g., Boysen, Bernston, & Prentice, 1987; Iverson & Matsuzawa, 1996; Smith, 1973) are beyond the abilities of monkeys. What can be concluded at present is that rhesus monkeys have failed to demonstrate the capacity so far to draw in this way. As such, this conclusion joins other tentative qualitative species differences that have been posited in cognitive domains such as language, theory of mind, self-recognition, and others.

### “Should-learn” failures

Other meaningful failures occur on tests for which there is evidence that an animal or species should be able to pass. Perhaps this evidence comes from neighboring species on the phylogenic scale, leading to predictions like, “If a smaller-brained species can demonstrate some cognitive competency, then larger-brained species should also show the cognitive ability.” In other cases, the evidence comes from animals of the same species, or even from the exact same individual animals, and demonstrations of some cognitive competency using complementary or converging tests. Washburn and Gullledge (1995) reported that rhesus monkeys could learn a memory task that is modeled after the popular Milton Bradley toy called Simon™, in which people are required to reproduce increasingly long sequences of color+tone+position stimuli. In the monkeys’ version (which we called ALVIN), a computer-generated cursor appeared in the middle of four differently-colored borders. To start a problem, a randomly selected border would brighten (e.g., the red wall turned pink) and a tone would sound. After a delay, the monkey was required to move the cursor into contact with that same wall. A successful response would result in a two-stimulus sequence (e.g., red wall turns pink, then blue wall turns cyan), and the monkey was rewarded for reproducing the sequence. This continued, with the to-be-remembered color+tone+position sequences getting longer and longer, until an error was made in recall. (Note that this is different from the serial learning or chaining paradigm made popular by Terrace and collaborators; e.g., Swartz, Chen & Terrace, 1991). The next problem began with a single randomly selected stimulus, and the process continued.

After 10,000 training trials per animal, three of the five rhesus monkeys were able to perform the task at levels significantly above chance (Washburn & Gullledge, 1995). These three monkeys showed significant, generalized relational learning that permitted them to recall novel stimulus sequences, although the monkeys were able to reproduce lists of only about 3.5 stimuli on average, whereas human adults could remember sequences of about twice that length. The macaques performed at levels above chance with retention intervals as long as 12 to 15 s, and at list-lengths comparable to those reported for 5- to 7-year-old children on a similar task (Moore & Staum, 1987).

Whereas attention to this work has understandably been on the impressive performance by these three animals, the two monkeys that failed to learn the task are also noteworthy. Despite the extensive training, these macaques never advanced beyond chance performance, fixating instead on stimulus-response associations and failing to generalize to new and longer sequences.

Individual differences such as this are common in comparative cognition, where some animals may “succeed” and others “fail” on the same learning tests. Researchers seem increasingly interested in this variability within, as well as between, groups of animals. Some of these individual differences may simply reflect measurement error or other random factors, or even differences in the rate of learning; however, others might be considered “meaningful failures” by those animals if the pattern of results can be shown to be consistent across tests and conditions—that is, if they satisfy the same criteria for meaningfulness that allow researchers to interpret the results from animals that passed the test to be considered “meaningful successes.” For each animal, such individual differences reflect a potential “should-learn” type of meaningful failure; that is, it reflects a failure to learn in the way that other animals succeeded in learning, a way it is reasonable to expect that the entire sample should have learned. For example, if one accepts the conclusion that performance by the three monkeys in the Washburn and Gulledge (1995) study defies a simple, stimulus-response associative explanation, then the other two animals should have been similarly capable of showing generalized, relational learning as well.

On the strength of the 1995 results, we recently undertook to train five new rhesus monkeys on the ALVIN task, with the goal of determining whether the monkeys might show statistical learning if a nonrandom grammar were to be embedded in the sequences. The five macaques were previously trained to respond to computer-generated stimuli by manipulating a joystick to control a computer-graphic cursor, and had been tested on many tasks (including delayed matching-to-sample tests of memory; see Rumbaugh, Richardson, Washburn, Hopkins & Savage-Rumbaugh, 1989). The monkeys were not deprived of food or otherwise reduced in body weight for purposes of testing, but rather engaged in the task or rested ad libitum, with continuous access to the computer-task apparatus.

In the course of 11 months of training, the five monkeys performed many thousands of trials, but not one of the animals learned the ALVIN task. Rather, each monkey remained at chance levels, responding positionally or only to one color, and thus earning a reward on 25% of the trials. Efforts to eliminate the positional or stimulus biases (e.g., with correction trials) simply caused the monkeys to switch to a different favored stimulus or position. Attempts to facilitate learning by titrating task difficulty (e.g., by gradually increasing the number of target walls surrounding the cursor) were ineffective, and across time the monkeys became increasingly unwilling to engage the task at all—a reticence that was resistant to increases in reward magnitude. Despite the fact that we had succeeded in training the three monkeys of the Washburn and Gulledge (1995) study following similar procedures, and despite extensive efforts to facilitate learning by these five macaques that had shown excellent learning on a wide range of other tasks, this study ended in a “should-learn” meaningful failure of relational learning, and in an example of associative drift, in which animals that have shown themselves capable of relational learning nonetheless got stuck in stimulus-response conditioning that could not lead to accurate performance on sequences that changed from problem to problem, and never acquired the “reproduce the sequence” rule for the game-like task.

It is not always clear how failures to learn should be interpreted. What makes a “meaningful failure” meaningful is not necessarily knowing what it means, but rather the confidence that one has that it is a real failure, versus one that is an artifact of inadequate training procedures or durations, or of some confound that prevented the participants from learning in the way that the task was designed to be learned. The ALVIN performance underscores the difficulty of teaching animals to make ordered recall (versus to recognize the most recent to-be-remembered stimulus), particularly with a stimulus set limited to four colors/positions. These findings do not cast doubt on whether rhesus monkeys *can* perform a sequence-memory task like ALVIN, but they do make the performance by the three animals in the Washburn and Gulledge (1995) study even more remarkable, and challenge the field to continue to try to replicate those findings so as to provide further support for that “meaningful success,” as well as to

identify the reasons for the individual differences. That is, further research is needed to determine why some animals succeeded whereas others failed.

### Introduction to the Special Issue

This issue of *Animal Behavior and Cognition* is devoted to “could learn” and “should learn” meaningful failures such as the ones described above and elsewhere (Rumbaugh & Washburn, 2003). The challenge for behavioral scientists is always to determine what is meaningful in the pattern of results they obtain. There is, in principle, as much to learn from animals’ so-called failures as from their successes. That said, our discipline has by convention set a stringent threshold for drawing conclusions from null results or for concluding that an individual or species does not or cannot demonstrate some ability. Thus, the goal for this special issue has been to go beyond simple “null results” and rather to highlight some of the surprising failures that remain after animals have been given fair and comprehensive opportunities to demonstrate a learning phenomenon of interest. As the meaningful failures described in this introduction illustrate rhesus monkeys’ susceptibility to associative drift—their tendency for associative conditioning, even though they have shown themselves capable of relational learning—the meaningful failures in this special issue suggest new ways that we should understand and study the animals that have failed in these interesting ways.

Some of the empirical reports contained in this issue fall into the “could learn” type of meaningful failure. They are reasonable explorations of the potential limits of cognitive competency for the participating animals. **Gazes, Chee and Hampton (2018)** reported that rhesus monkeys failed to learn by exclusion, even though they were able to identify which stimuli are new (i.e., to choose by exclusion). **Lewis, Berntsen and Call (2018)** examined long-term memory for events that occurred weeks prior, and failed to find clear evidence of binding of the social and object information in the memory of the chimpanzees, bonobos, and orangutans. The capuchin monkeys in the study by **Prétôt, Weiss and Brosnan (2018)** failed to show strong evidence of being able to perceive a conspecific’s volition in a social learning paradigm. **Smith, Smith and Beran (2018)** also examined the cognitive competencies of capuchin monkeys, and described those primates’ surprisingly consistent failure to show clear evidence of metacognitive responding. **Vonk and Jett (2018)** concluded that bears may be limited with respect to the capacity for abstract concept learning, based on the animals’ failure to demonstrate generalized rule learning on several tasks and conditions.

Other studies in this special issue illustrate “should learn” types of meaningful failures—absences of learning effects that are strongly predicted by theory or by empirical results from other animals or other tests. **Bowers and Timberlake (2018)** tested whether rats’ behavioral interventions would yield the effects on learning that are predicted by current theories, but failed to find evidence for intervention effects on goal-directed behavior. **McGuire and Vonk (2018)** described the surprising failure of gorillas to learn the conditional discriminations that would result in more, versus less, food, and suggested that gorillas may struggle particularly with associating two-dimensional stimuli with real-life outcomes. **Parrish et al. (2018)** reported evidence that capuchin monkeys succeed on multiple tests of self-control, albeit with individual differences; however, the monkeys failed to show evidence that would support the ego-depletion model of cognitive control. Similarly, **Rick, Donaire, Papini, Torres, and Pellón (2018)** tested the predictions of a popular theoretical model, but found a failure of frustration to play a strong role in the alterations of rats’ behavior resulting from surprising changes in reward magnitude or frequency. **Seki, Osmanski and Dooling (2018)** attempted to bring the vocal production of budgerigars under operant control; however, the birds generally failed to learn appropriate vocal responses, despite extensive operant conditioning and extensive evidence of vocal learning by budgerigars in nature. Working with pigeons and rats, **Zentall, Case, Berry and Luong (2018)** explored the role of impulsive choice in explaining the curious failure of primates, pigeons, and rats to maximize rewards in the ephemeral paradigm.

A recurring theme across these reports is the hope that the meaningful failures documented here will serve as catalysts for further research and discussion. We also hope that researchers will be

encouraged to document and explore the potential meaningful failures in their own studies—to avoid the “file drawer” problem (Rosenthal, 1979) of unpublishable null findings by systematically exploring what is interpretable and important in other failures to learn. Doing so will not only enhance the field’s understanding of cognition as it is manifest across species, and across individuals within each species, but it will also help to sharpen our interpretation of the “meaningful successes” that fill the comparative cognition literature.

## Acknowledgments

The data described here were collected with support from NICHD grants 060563 and 38051, and with support from Georgia State University. All animals were treated in compliance with federal regulations and in accordance with the ethical guidelines of the American Psychological Association and the University System of Georgia. The authors acknowledge the collaboration and assistance of Mike Beran, Chris Conway, John Gulledge, and the Language Research Center research and caretaking staff.

## References

- Agrillo, C., Parrish, A. E., & Beran, M. J. (2014). Do rhesus monkeys (*Macaca mulatta*) perceive the Zöllner illusion? *Psychonomic Bulletin & Review*, *21*, 986-994.
- Bowers, R. I., & Timberlake W. (2018). Intervening to learn about effects of one's actions. *Animal Behavior and Cognition*, *5*, 79-102.
- Boysen, S. T., Berntson, G. G., & Prentice, J. (1987). Simian scribbles: A reappraisal of drawing in the chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, *101*, 82-89.
- Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681-684.
- Gazes, R. P., Chee, N. W., & Hampton R. R. (2018). Monkeys choose, but do not learn, through exclusion. *Animal Behavior and Cognition*, *5*, 9-18.
- Iversen, I. H., & Matsuzawa, T. (1996). Visually guided drawing in the chimpanzee (*Pan troglodytes*). *Japanese Psychological Research*, *38*, 126-135.
- Lewis, A., Berntsen, D., & Call, J. (2018). Remembering past exchanges: Apes fail to use social cues. *Animal Behavior and Cognition*, *5*, 19-40.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361-366.
- McGuire, M. C., & Vonk, J. (2018). Gorillas (*Gorilla gorilla gorilla*) fail to learn abstract cues of differential outcomes in a novel cognitive bias test. *Animal Behavior and Cognition*, *5*, 103-117.
- Moore, R. S., & Staum, M. (1987). Effects of age and nationality on auditory/visual sequential memory of English and American children. *Bulletin of the Council for Research in Music Education*, *91*, 126-131.
- Parrish, A. E., James, B. T., Rossettie, M. S., Smith, T., Otolara-Garcia, A., & Beran, M. J. (2018). Investigating the depletion effect: Self-control does not waiver in capuchin monkeys. *Animal Behavior and Cognition*, *5*, 118-138.
- Prétôt, L., Weiss D. J., & Brosnan, S. F. (2018). Do non-human primates ascribe goals to the choices of conspecifics? *Animal Behavior and Cognition*, *5*, 41-54.
- Purves, D., & Lotto, R. B. (2011). *Why we see what we do redux: A wholly empirical theory of vision*. Sunderland, MA: Sinauer Associates.
- Rick, P., Donaire, R., Papini, M. R., Torres, C., & Pellón, R. (2018). Can surprising nonreward and adjunctive behavior influence each other? *Animal Behavior and Cognition*, *5*, 139-153.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.
- Rumbaugh, D. M., Richardson, W. K., Washburn, D. A., Hopkins, W. D., & Savage-Rumbaugh, E. S. (1989). Rhesus monkeys (*Macaca mulatta*), video tasks, and implications for stimulus- response spatial contiguity. *Journal of Comparative Psychology*, *103*, 32-38.
- Rumbaugh, D. M., & Washburn, D. A. (2003). *The intelligence of apes and other rational beings*. New Haven: Yale University Press.
- Seki, Y., Osmanski, M. S., & Dooling, R. J. (2018). Failure of operant control of vocal learning in budgerigars. *Animal Behavior and Cognition*, *5*, 154-168.

- Smith, D. A. (1973). Systematic study of chimpanzee drawing. *Journal of Comparative and Physiological Psychology*, 82, 406-414.
- Smith, T. R., Smith, J. D., & Beran, M. J. (2018). Not knowing what one knows: A meaningful failure of metacognition in capuchin monkeys. *Animal Behavior and Cognition*, 5, 55-67.
- Swartz, K. B., Chen, S., & Terrace, H. S. (1991). Serial learning by rhesus monkeys: I. Acquisition and retention of multiple four-item lists. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 396-410.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Vonk, J. & Jett, S. E. (2018). “Bear-ly” learning: Limits of abstraction in black bear cognition. *Animal Behavior and Cognition*, 5, 68-78.
- Washburn, D. A., & Gulledge, J. P. (1995). Game-like tasks for comparative research: Leveling the playing field. *Behavior Research Methods*, 27, 235-238.
- Zentall, T. R., Case, J. P., Berry, J. R., & Luong, J. (2018). The ephemeral reward task: Pigeons and rats fail to learn unless discouraged from impulsive choice. *Animal Behavior and Cognition*, 5, 169-180.