L

Learning to Learn

Roger K. Thomas
Department of Psychology, University of Georgia, Athens, GA, USA

Synonyms

Deutero-learning; Learning set formation; Transfer of learning

Definition

Learning to learn refers to the observation that prior learning often facilitates subsequent learning.

Introduction

> . . . in the late 1940s he [Harlow] achieved a major conceptual and methodological breakthrough with his discovery of learning sets. (Suomi and LeRoy 1982, p. 321)

> Harlow's 1949 article, clearly describing for the first time, the concept of learning set formation, is one of the most widely cited articles in the animal behavior literature. (Schrier and Thompson 1984, p. 109)

> Harry F. Harlow (1905–1981) is known for discovering the learning set (e.g., learning how to learn) phenomenon . . .. (Rumbaugh 1997, p. 197)

Despite Harlow's students' (e.g., Schrier and Suomi) and colleagues' enthusiasm for attributing the "discovery" of "learning to learn" to Harlow, it had been known in various guises, such as "formal discipline" or "transfer of training," at least, since the 1890s (Hall 1966, pp. 477–479). Nevertheless, an important event in the study of learning to learn, especially with nonhuman animals, was the publication of Harry Harlow's concept of "learning set formation" and, more importantly, his developing methods to investigate it and providing insightful measures to assess it. This was first presented in his presidential address on May 7, 1948, at the meeting of the Midwestern Psychological Association and then published in the *Psychological Review* (Harlow 1949).

Generally unrecognized, 7 years prior to Harlow (1949), Gregory Bateson (1942), using the term *deutero-learning,* published a highly similar conceptualization to Harlow's (1949) learning set formation. Originally, Bateson's was a 16-page commentary on a presentation by his wife, the renowned anthropologist, Margaret Mead. It was reprinted with the publisher's permission in Bateson's more accessible *Steps to an Ecology of Mind* (Bateson 1972a) where he added a significant, related essay (Bateson 1972b).

Although Bateson did not have empirical data as Harlow did, Bateson presented two

J. Vonk, T. K. Shackelford (eds.), *Encyclopedia of Animal Cognition and Behavior*,
https://doi.org/10.1007/978-3-319-47829-6_739-1

hypothetical data graphs (Bateson 1972a; please consult Bateson for “Fig. 1” and “Fig. 2”) that projected results that remarkably anticipated Harlow’s (1949) first two experimental results graphs (“Fig 2” and “Fig. 3”); Harlow’s “Fig. 1” (please consult Harlow for “Figs. 1, 2, and 3”) was the well-known Wisconsin General Testing Apparatus (WGTA) developed by Harlow at the University of Wisconsin. The WGTA’s most important function was to prevent an experimenter from inadvertently giving animals cues to make correct responses. Additionally, both Bateson (1942) and Harlow (1949) used the phrase “learning to learn” referring to the processes each was writing about.

Perhaps for understandable reasons, Harlow did not cite Bateson’s prior work, as it was in a relatively obscure source for animal learning investigators. However, priority per se is not an issue here, as neither Bateson nor Harlow claimed originality regarding “learning to learn.” Because Bateson published his contribution first, it will be addressed first. However, Harlow’s contribution had a greater impact on comparative cognition, and it will be addressed more extensively.

Gregory Bateson (1904–1980) and Deutero-Learning

Bateson (1972a) coined the terms “proto-learning” and “deutero-learning” to make the following distinction.

> Let us say that there are two sorts of gradient discernible in all continued learning. The gradient at any point on a simple learning curve (e.g., a curve of rote learning) we shall say chiefly represents rate of proto-learning. If, however, we inflict a series of *similar learning experiments* on the same subject, we shall find that in each successive experiment the subject has a somewhat steeper proto-learning gradient, that he learns somewhat more rapidly. This progressive change in rate of proto-learning, we will call deutero-learning. (p. 167)

Later, Bateson (1972b) renamed proto-learning and deutero-learning as Learning I and Learning II, and he anticipated an even higher order or learning, Learning III, which is mind-boggling in its implications. If someone can determine how to investigate it scientifically, it has the potential to have a far greater impact than deutero-learning or learning set formation. However, further consideration of that is not appropriate here, and this entry will continue using the term “deutero-learning.”

Bateson (1942) did not consider his concept of deutero-learning to be original. Before coining the terms proto-learning and deutero-learning, he had written:

> Now it so happens that in the psychological laboratories there is a common phenomenon of a somewhat higher degree of abstraction or generality than those which the experiments are planned to elucidate. It is a commonplace that the experimental subject – whether animal or man, becomes a better subject after repeated experiments. He not only learns to salivate at the appropriate moments, or to recite the appropriate nonsense syllables; he also, in some way*, learns to learn.* He not only solves the problem set him by the experimenter, where each solving is a piece of simple learning; but, more than this, he becomes more and more skilled in the solving of problems. (Bateson 1972a, p. 166; emphasis added)

As Bateson expressed it, he had merely coined the term “deutero-learning,” “. . . to avoid the labor of defining operationally all the other terms in the field (transfer of learning, generalization etc.)” (Bateson 1972a, p. 167).

Among his predecessors Bateson cited Maier (1940). Bateson wrote “ . . . the concept of deutero-learning can be seen as almost synonymous with Professor Maier’s concept of “direction” (Bateson 1972a, p. 167). Bateson was too generous. Maier’s concept of “direction” differed from Bateson’s “deutero-learning” and Harlow’s “leaning set formation,” because Maier referred vaguely to improvement over unrelated tasks and Maier did not show how “direction” might be quantified.

Harry Harlow (1905–1981) “The Formation of Learning Sets”

As mentioned above, Harlow’s (1949) article was his presidential address at the meeting of the Midwestern Psychological Association in May, 1948. In terms of the chronological development of the learning set formation concept, a footnote to the 1949 article shows that it was based on research that was " . . . supported in part by grants from the Special Research Fund of the University of Wisconsin for 1944–1948" (Harlow 1949, p. 51). The origin of Harlow's learning set formation may be traced to an article published in 1944 (Harlow 1944) whose manuscript was received on June 18, 1942. Thus, although Harlow did not use the phrase "learning set formation" until 1949, he may have been developing it empirically and contemporaneously with Bateson (1942). Meanwhile, it is the 1948 presidential address and the resulting 1949 article that are typically cited as the beginning of Harlow's learning set formation, and they will be considered first.

In Harlow's1948 presidential address, he began to discuss the broader role that learning plays in human and nonhuman animals' lives. He asserted that "Our emotional, personal, and intellectual characteristics are not the mere algebraic summation of a near infinity of stimulus-response bonds" (Harlow 1949, p. 51). He continued,

> The learning of primary importance to primates, at least, is the formation of learning sets; it is the *learning how to learn efficiently in* the situations the animal frequently encounters. (p. 51)

Harlow then described some of his empirical research, including an illustration of the WGTA and several graphs of illustrative data. As noted earlier, Harlow's empirically based Figs. 2 and 3 were remarkably well-predicted by Bateson's (1942/1972a) hypothetical data graphs, Figs. 1 and 2.

Perhaps because the article was the text for an oral presentation, Harlow cited no references. However, Suomi and Leroy (1982) listed Harlow's 323 publications in chronological order beginning in 1932 and ending in 1978. Near the end of Harlow's (1948) presidential address, he identified what he believed to be his unique contribution.

> The emphasis throughout this paper has been on the role of the historical or experience variable in learning behavior- the forgotten variable in current learning theory and research. Hull's Neo-behaviorists have constantly emphasized the necessity for an historical approach in learning, yet they have not exploited it fully. Their experimental manipulation of the experience variable has been largely limited to the development of isolated habits and their generalization . . .. Psychologists working with human subjects have long believed in the phenomenon of learning sets and have even used sets as explanatory principles . . .. These psychologists have not, however, investigated the nature of these learning sets . . . we have carried out studies that outline the development and operation of specific learning sets . . . it is our hope that our limited data will be extended by those brave souls who study real men and real women. (pp. 64–65)

The "Evolution" for Harlow of the Concept of Learning Set Formation

Harlow (1944) was his first experimental research article related to "learning set formation" (hereafter LSF). However, he did not use the phrase LSF; rather, he considered the interpretation that the discrimination learning was occurring *insightfully* based on the use of "hypotheses" (Harlow 1944, p. 7).

Later, in the Discussion section, Harlow described the LSF process, although, as already noted, he did not use the phrase "learning set formation."

> Indeed, once a monkey has solved a preliminary series of discriminations and has formed habits of responding to stimulus objects regardless of their position in space, later discriminations will be solved in a single trial or less, in a majority of cases. Thus, if the first response is by chance correct, no additional errors will be made. If the first response is by chance incorrect, the error will be corrected on the succeeding trial and no additional errors will be made. In gestalt terminology the discrimination learning is occurring 'insightfully'. (Harlow 1944, p. 10)

[Five paragraphs later]

> . . . once appropriate *reaction sets* have been formed in monkeys these sets may be transferred from one pair of discrimination objects to another making it possible for the subjects to meet a strict criterion for formation of a discrimination with a minimum amount of specific training. (p. 11)

Zable's and Harlow's (1946) interpretational emphasis was on forming "hypotheses," but their use of "hypotheses" was not that developed later by Harlow's student, Marvin Levine, whose "win-stay, lose-shift" hypotheses became so well associated with LSF (Levine 1959). Levine meant, if the animal chooses correctly on trial 1, that is a "win" and to obtain maximum reinforcement on succeeding trials, it should "stay" with its original choice. If the animal chooses incorrectly in trial 1, that is to "lose" and it should "shift" to the other discriminandum to obtain maximum reinforcement. As tests of LSF developed, new problems might be introduced for varying numbers of trials. However, the common practice that eventually emerged was to introduce a new problem every six trials. Discriminanda have typically been objects, and Harlow often described the task as "object quality learning set."

Learning to learn can be investigated in other ways as Harlow (1959) did with reversal learning. Harlow's emphasis here was on the development of learning, and he used rhesus monkeys ranging from neonates to adults. One example was to use a Y maze to train an infant monkey to learn to first go to the left arm of the Y for reinforcement until it met a predetermined number of trials-to-criterion (e.g., 90% correct responses in 20 trials). Then, the investigator reverses location of the reinforcers to the right arm of the Y maze. Typically, there is a persistence in going to the left arm but eventually, the monkey learns that reinforcement is now received with responses to the right arm. After a number of reversals and if the animal is learning to learn, it will detect the reversal more quickly. Often, one nonreinforced trial is sufficient to inform the animal that a reversal has occurred. Of course, reversal learning can be done in many ways. For example, rather than left versus right in a Y maze, the animal might learn reversals between, say, black versus white alleys with position of the alleys changed on 50% of the trials. Generally, Harlow's young monkeys also performed the alley reversal task well.

As noted, Harlow changed the left-right position of the "correct" alley in the black-white alley reversal task on 50% of the trials, and it is widely realized that with most tasks (including LSF), the positions of the discriminanda must be changed over trials randomly or quasi-randomly. Quasi-random changes have advantages as described by Fellows (1967) whose article includes quasi-random series that many investigators have used in their research.

The 6-trial, object quality LSF task has been used with many species, so it will be the remaining focus here. The experimenter determines which discriminandum will be reinforced (typically but not always with food) for the six trials. The animal can choose only by chance on trial 1, but if learns to learn, it will learn to use information about trial 1 to choose correctly (e.g., "win-stay, lose-shift") for the remaining five trials. Trial 2 performances have been the most used measure of LSF.

A Failed Application of LSF

Warren (1965) suggested that LSF formation might be a good way to compare species' learning abilities, and he published a graph where the ordinate was PER CENT CORRECT ON TRIAL 2 and the abscissa was PROBLEMS. Warren reported data from six species. The best performer was a rhesus monkey with 85% correct on trial 2 after 400 problems. A rat and a squirrel tied for worst performance with 60% correct after 1800 problems.

This application of LSF reached its zenith with a graph presented by Hodos (1970) on which he plotted data from 16 species including two children. The best performer was a child (age unspecified but IQ = 136) who had 100% correct on Trial 2 after 100 problems. The tree shrew was the worst performer at 50% correct on trial 2 in 1000 problems.

However, Warren (1974), who had initiated the use of LSF as a way to compare learning abilities of various species, also brought it to its end when he wrote:

> Primates differ from other mammals in their extraordinary development of the visual system. They surpass most other mammalian species in respect to their capacity for color vision, stereopsis, and visual acuity… it is apparent that we have no basis for guessing the degree to which the inferior performances of non-Primate species in visual learning set problems reflect an inferiority in visual

> sensitivity and perception instead of a defective capacity for learning. (Warren 1974), p. 448.

Warren and others (e.g., Thomas 1980) discussed other contextual variables that might provide some species with advantages and other species with disadvantages in performing learning tasks. Optimally, in species' learning or intelligence test comparisons, such contextual variables will be adapted to be most suitable for each species.

Warren (1974) cited a study by Slotnick and Katz (1974) who used olfactory learning set discriminanda (floral scents) to study LSF by rats. The scents could only be delivered one at a time via a complicated, expensive, and tedious-to-use apparatus. The authors did not report percentages correct on trial 2, but they did report a number errorless or one-error problems, and they "… suggested that a "win-stay, lose shift" hypothesis was in effect" (Slotnick and Katz 1974, p. 798).

Not cited by Warren (1974), Langworthy and Jennings (1972) used a much simpler and less expensive way to present olfactory discriminanda to study oddity concept learning by rats which can also reveal LSF as shown below. They used ping pong balls saturated with the odors of one of eight food flavorings. As a given odor might be odd on some problems and nonodd on other problems, no odor could be exclusively associated with either odd or nonodd.

Three balls, two of the same odors and the third of a different odor, were presented side by side on a platform holding an open-air chute in which the balls were inserted. Marks on the chute showed how far the rat had to nudge the odd ball aside to access the food cup beneath it. Langworthy and Jennings (1972) reported good results, but it was unclear whether the food reinforcer was only beneath the odd ball. If so, it is possible that the rats detected the correct ball by smelling the food beneath it.

Thomas and Noble (1988) and Bailey and Thomas (1998) followed Langworthy and Jennings in using odoriferous ping pong balls as discriminanda, but among other differences, they used 16 (Thomas and Noble) or 18 (Bailey and Thomas) food flavorings and 5-trial problems. As with Langworthy and Jennings, no odor was associated exclusively with odd or nonodd. They also baited all food wells, so food odor could not be a cue to the odd ball.

Bailey and Thomas (1998) made other changes, and for present purposes their study might be considered the more useful than the other. Both studies were designed to determine whether rats might learn the oddity concept, but the problems were presented in a learning set paradigm. Neither Thomas and Noble (1988) nor Bailey and Thomas (1998) found evidence of oddity concept learning, as performances on the first trials of new problems were at chance, but Bailey and Thomas who administered 60 problems found that their three rats averaged 87% correct on trial 2, on problems16–30 and 81% correct on trial 2 on problems 16–60.

LSF's Place Among Other Learning Processes

Harlow (1958) identified some problems' association with the study of' the evolution of learning including:

> Another difficulty lies in existing limitations to a precise classification of the forms of learning and learning problems into levels of difficulty… the problem is far from solved; and no one has even attempted to scale the various learning problems or classes of problems …. (p. 269).

Before we can return to the topic of this section, it is necessary first to show the progress made since Harlow's observation just quoted.

Gagné (e.g., 1970), who focused on human learning, proposed a hierarchy of eight types of learning, which he asserted to encompass any and all types of human learning. They were, generally deemed to be hierarchical on the assumption that lower levels, usually, were prerequisites for higher levels. From lowest to highest in Gagné's hierarchy were signal learning (i.e., classical or Pavlovian conditioning), stimulus-response learning (i.e., operant conditioning), chaining, verbal association, discrimination learning, concept learning, rule learning, and problem solving. Gagné's examples for verbal association and for his top three levels were taken from human learning literature, and most would not be feasible with nonhuman animals.

Thomas (1980) sought to develop a comprehensive hierarchy of learning types that could be used with all animals including humans. Thomas discovered that Bourne's (1970) approach to concept learning, which was to base it on tasks constructed to comply with formal logic (explained below), could replace Gagné's highest three levels and still account fully for any and all types of learning at Gagné's three highest levels. Additionally, this approach could be used with humans and nonhuman animals.

Gagné (1970) regarded verbal association to be parallel to chaining, so verbal association being less amenable to testing among most animals could be eliminated from Thomas's hierarchy (Thomas 1980). Thomas noted that Gagné had overlooked a lower form of learning than signal learning, namely, habituation and its complementary process, sensitization. Gagné clearly described discrimination learning as learning multiple discrimination problems concurrently, so Thomas renamed it concurrent discrimination learning.

Bourne (1970) based concept learning on tasks that were constructed to comply with the truth-function tables in formal logic. Bourne's lowest level involved only the logical operations affirmation and negation, which Thomas considered to be the foundations for class concepts. For Thomas (1980), this became level 6, class concepts. Bourne's middle level of concepts were based on the logical operations, conjunction, disjunction, or conditional and their complementary processes, all of which Bourne considered to be parallel processes. Agreeing with Bourne that they were parallel processes and assuming that class concepts were prerequisites for concepts based on them, Thomas's level 7 became relational concepts involving conjunction, disjunction, or conditional operations and their complementary processes. Finally, Bourne regarded the logical operation for the biconditional and its complementary process to be above the three at his middle level, because the conditional is a prerequisite to the biconditional. In Thomas's hierarchy, level 8 became relational concepts based on the biconditional and its complementary process. Most research has focused on the primary as opposed to the complementary processes, so the complementary processes will be ignored here henceforth. However, one must always assume they can be present and tasks may be constructed based on them. Thomas equated learning ability with intelligence, and he considers his hierarchy to be a learning-intelligence hierarchy.

Thomas's Learning-Intelligence Hierarchy

8. Relational Concepts based on the Biconditional
7. Relational Concepts based on the Conditional, Conjunctive, or Disjunctive
6. Absolute and Relative Class Concepts based on Affirmation and Negation
5. Concurrent Discrimination Learning
4. Chaining (Chains of S-R Learning Units)
3. Stimulus-Response (S-R) Learning or Operant Conditioning
2. Classical or Pavlovian Conditioning
1. Habituation and Sensitization

Note that both Gagné's (1972) and Thomas's (1980) hierarchies are on an *ordinal scale*. Thomas followed Gagné's order generally; however, Thomas considers that levels 2 and 3 (Classical and Operant Conditioning, respectively) might be parallel. Also, as it has been shown that representative species from all vertebrate classes (except amphibians appear not to have been tested) can succeed to some extent at level 5, Thomas believes that most mammalian and avian differentiations in learning abilities are likely to be found among Levels 6–8, class and relational concepts.

To show class concept learning at level 6, one must use trial-unique discriminanda, or if the same discriminanda are presented more than one time, evidence for concept learning must be limited to the data from the first trials of presentations of new discriminanda. Otherwise, the animal might have learned the task by rote based on trial and error. *With absolute class concepts, it is not necessary to compare discriminanda*. For example, if the concept being investigated is "tree" and the animal is shown trial-unique examples of trees, and if the animal knows or has acquired the concept of "tree," it will respond directly to the discriminandum of a tree to show that it recognizes new exemplars as being trees. Some exemplars may be a fuzzy fit, and those may cause errors (e.g., shrubs vs. trees).

With relative class concepts *it is necessary to compare discriminanda to affirm which discriminandum represents the concept.* Perhaps the most studied relative class concept with animals has been the "oddity concept." Typically, the animal is shown trial-unique examples of three discriminanda, two of which are identical and the other is different or "odd"; and to recognize which one is odd it must compare the discriminanda.

Since Thomas's first publication of the learning-intelligence hierarchy 1980, he has had occasion to make minor emendations (e.g., Bailey et al. 2007). In Bailey et al., Thomas addressed a matter that had puzzled him for decades, namely, where in the hierarchy of learning types does LSF fit. He decided it best fit as a parallel process subsumed under concurrent discrimination learning. For Gagné, concurrent discrimination learning referred to multiple discrimination problems learned in parallel. LSF refers to multiple discrimination problems learned serially. Thomas (2012) explained further why LSF should not be considered to be an instance of concept learning.

Concluding Remarks

"Learning to learn" has a long history under a variety of labels. However, it was Gregory Bateson and Harry Harlow who gave it a more substantial theoretical foundation, and Harlow provided clear means to investigate and measure it. Ironically, Bateson (1942, 1972a) had expressed doubt that it could ever be studied in the laboratory, although later (Bateson 1972b) acknowledged Harlow (1949).

Thomas (2012), who concluded that LSF is not at the level of concept learning, also wrote:

> Nevertheless, some may consider it to be an open question whether to agree with Thomas and colleagues, and many who study concept learning in animals regarding (a) what "conceptualization" means or (b) that the necessary evidence for conceptualization requires that the subject respond correctly to trial-unique [exemplars} or first trials with new exemplars of the concept. For example, one might argue that learning a strategy or "rule" such as that which humans might verbalize as "win-stay, lose-shift" involves conceptualization. However, it must also be recognized that with animals it is unlikely that they learn anything akin to such verbalizations of a rule or strategy, not to overlook that it is unlikely that an experimenter could provide unequivocal evidence that they did. (p. 1968.

Finally, there may be room to advance the status of LSF, and that may be a challenge for future investigators.

Cross-References

- ▶ Comparative Cognition
- ▶ Concept Formation
- ▶ Discrimination Leaning
- ▶ Harry Harlow
- ▶ Intelligence
- ▶ Learning
- ▶ Olfactory Discrimination
- ▶ Primate Sensory Systems
- ▶ Reversal Learning
- ▶ Transfer of Learning
- ▶ Win-Stay
- ▶ Wisconsin General Testing Apparatus

References

Bailey, A. M., & Thomas, R. K. (1998). In investigation of oddity concept learning by rats. *The Phycological Record*, *48*, 333–344.

Bailey, A. M., McDaniel, W. F., & Thomas, R. K. (2007). Approaches to the study of higher cognitive functions related to creativity in animals. *Methods, 42*, 3–11.

Bateson, G. (1942). Social planning and the concept of deutero-learning. A commentary on Margaret Mead's "The comparative study of culture and purposive cultivation of democratic values". In L. Bryson & L. Finkelstein (Eds.), *Science philosophy and religion: Second symposium* (pp. 81–97). New York: Conference on Science, Philosophy and Religion.

Bateson, G. (1972a). Social planning and the concept of deutero-learning. In *Steps to an ecology of mind: A revolutionary approach to man's understanding of himself* (pp. 159–176). New York: Ballantine Books.

Bateson, G. (1972b). The logical categories of learning and communication. In *Steps to an ecology of mind: A revolutionary approach to man's understanding of himself* (pp. 279–308). New York: Ballantine Books.

Bourne, L. E., Jr. (1970). Knowing and using concepts. *Psychological Review, 77*, 546–555.

Fellows, B. J. (1967). Chance stimulus sequences for discrimination tasks. *Psychological Bulletin, 67*, 87–92.

Gagné, R. M. (1970). *The conditions of learning* (2nd ed.). New York: Holt Rinehart and Winston, Inc.

Hall, J. F. (1966). *The psychology of learning*. New York: J. B. Lippincott.

Harlow, H. F. (1944). Studies in discrimination learning by monkeys: I. The learning of discrimination series. *Journal of Psychology, 30*, 3–12.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56*, 51–65.

Harlow, H. F. (1958). The evolution of learning. In A. Roe & G. G. Simpson (Eds.). Behavior and evolution (pp. 269–290). New Have, CT: Yale University Press.

Harlow, H. F. (1959). The development of learning in the rhesus monkey. *American Scientist, 47*, 459–479.

Hodos, W. (1970). Evolutionary interpretation of neural and behavioral studies of living vertebrates. In F. O. Schmitt (Ed.), *The neurosciences: Second study program* (pp. 26–39). New York: Rockefeller University Press.

Langworthy, R. A., & Jennings, J. J. (1972). Odd ball, abstract, olfactory earning in laboratory rats. *The Psychological Record, 22*, 487–490.

Levine, M. (1959). A model of hypothesis behavior in discrimination learning set. *Psychological Review, 66*, 353–366.

Maier, N. R. F. (1940). The behavior mechanisms concerned with problem solving. *Psychological Review, 47*, 43–58.

Rumbaugh, D. M. (1997). The psychology of Harry F. Harlow: A bridge from radical to rational behaviorism. *Philosophical Psychology, 10*, 197–210.

Schrier, A. M., & Thompson, C. R. (1984). Are learning sets learned? A reply. *Animal Learning & Behavior, 12*, 109–112.

Slotnick, B. M., & Katz, H. M. (1974). Olfactory learning set formation in rats. *Science, 185*, 796–798.

Suomi, S. J., & LeRoy, H. A. (1982). In memoriam: Harry F. Harlow (1905–1981). *American Journal of Primatology, 2*, 319–342.

Thomas, R. K. (1980). Evolution of intelligence: An approach to its assessment. *Brain, Behavior and Evolution, 17*, 454–472.

Thomas, R. K., & Noble, L. M. (1988). Visual and olfactory oddity learning in rats: What evidence is necessary to show conceptual behavior? *Animal Learning and Behavior, 16*, 157–163.

Thomas, R. K. (2012). Learning set formation and conceptualization. In N. Seel (Ed.), *Encyclopedia of the sciences of learning: Part 12* (pp. 1966–1968). New York: Springer.

Warren, J. A. (1965). Primate learning in comparative perspective. In H. F. Harlow, F. Stolnitz, & A. M. Schrier (Eds.), *Behavior of nonhuman primates: Vol. 1, Modern research trends* (pp. 249–281). New York: Academic.

Warren, J. M. (1974). Possibly unique characteristics of learning by primates. *Journal Human Evolution, 3*, 445–454.

Zable, M., & Harlow, H. F. (1946). The performances of rhesus monkeys on series of object-quality and positional discriminations and discrimination reversals. *Journal of Comparative psychology, 39*, 13–23.