Biol. Rev. (1996), 71, *pp.* 157–190
Printed in Great Britain

A REVIEW OF SIMULTANEOUS VISUAL DISCRIMINATION AS A METHOD OF TRAINING OCTOPUSES

BY J. G. BOAL

Marine Biomedical Institute, University of Texas Medical Branch, Galveston, Texas 77555-0863/(409) 772-2133

(*Received* 12 *October* 1994; *revised* 9 *March* 1995; *accepted* 24 *March* 1995)

CONTENTS

I. INTRODUCTION

Regions of intersection between disciplines provide particularly fertile ground for misunderstandings and inexact interpretations. The expanding region of animal behaviour in which methods from psychology are being used to address ethological questions is no exception. One example of these problems can be found in octopus learning research; experimental results have been both uncritically accepted and summarily rejected. These uncritical evaluations have significant impact on our discussions of the evolution of the comparative behaviour of invertebrates and vertebrates.

Learning in octopuses has been extensively studied. Topics have included appetitive and operant conditioning (Dews, 1959; Crancher *et al.*, 1972; Papini & Bitterman, 1991), spatial learning (Fiorito *et al.*, 1990; Mather, 1991*a*; and review in Wells, 1978), visual and tactile discrimination learning (see reviews in Boycott, 1954, 1965; Boycott & Young, 1950; Boyle, 1986*a*; Sanders, 1975; Schiller, 1948; Thorpe, 1956; Wells, 1962, 1966*a*, *b*, 1978; Young, 1961, 1977), and the learning of repeated reversals, a type of learning set formation (Mackintosh, 1962; Young, 1962; Mackintosh & Mackintosh, 1963, 1964*a*).

The recent demonstration of observational learning in *Octopus vulgaris* (Fiorito & Scotto, 1992) piqued renewed interest in octopus learning. Evidence of such complex social behaviour in a solitary invertebrate has widespread implications for our understanding of the evolution of complex behaviour in general and social behaviour in particular. (See Packard, 1972 for a review of remarkable convergences between cephalopods and fish.) It is not surprising that Fiorito & Scotto's paper has generated some controversy. While the debate in *Science* (Biederman & Davey, 1993; Fiorito, 1993; Suboski *et al.*, 1993) focused on whether the behaviour truly constituted observational learning, rather than something less complex such as stimulus enhancement, questions resurfaced amongst cephalopod researchers and students of invertebrate behaviour about the control and replicability of octopus learning experiments in general.

Bitterman published strong criticisms of octopus learning experiments (1966, 1975). Aquarium facilities were reported to be inadequate for the proper maintenance of healthy animals. Training techniques were considered crude and rough on the animals. More serious for the interpretation of published work, experimental techniques could have provided opportunities for subjectivity and variability in the presentations of stimuli and in the classifications of trial outcomes. Both human expectancy (Rosenthal & Fode, 1963) and interactions between experimenters and subjects (Davis & Balfour, 1992) are known to influence experimental outcomes with many taxa. Indeed, when I eliminated many potential sources of interaction between experimenters and subjects, I was unable to train octopuses in a simultaneous discrimination task (Boal, 1993). These unresolved issues provided the impetus for this review.

Many reviews of octopus learning research have been published (see above). This paper differs from these in its careful examination of training methods. My intent is to examine experimental methods critically rather than simply to review results. Although most discrimination experiments using octopuses as subjects have employed successive presentations, I chose to focus on experiments using simultaneous presentations. Simultaneous presentations offer greater versatility in the types of questions that can be addressed and better assurance of identical external conditions during presentations of the rewarded and unrewarded stimuli. I have included all published experiments that used simultaneous, two-stimulus, visual discrimination training with octopuses as subjects. These experiments were designed to answer different questions; as a result, not all data are equally complete. I also describe several previously unpublished experiments that specifically address discrimination training issues.

II. PUBLISHED EXPERIMENTS

I begin by briefly reviewing each of the simultaneous discrimination experiments found in the literature. Subjects in each experiment were *Octopus vulgaris*, unless otherwise noted. In typical simultaneous training, each octopus (about 600 g) was trained in its own home tank (about 60 × 100 × 50 cm). A brick den was provided at one end where the octopus sat between trials. The two stimuli were attached to clear plastic rods and were presented at the far end of the tank. Usually, the rods were hand-held

Table 1. *Simultaneous unrewarded choices*

Preferred stimulus	Unpreferred stimulus	Back-ground	Percent preference	Reference
White circle	White square	(Grey)	71	Sutherland & Muntz, 1959
Black square	Black circle	(Grey)	67	Sutherland & Muntz, 1959
White square	Black square	(Grey)	83	Muntz *et al.*, 1962
Vertical rectangle moved vertically	Vertical rectangle moved horizontally	(Grey)	95	Sutherland & Muntz, 1959
Horizontal rectangle moved horizontally	Horizontal rectangle moved vertically	(Grey)	86	Sutherland & Muntz, 1959
Red ball	White ball	window	53 on day 1, 78 on day 5	Fiorito & Scotto, 1992

and moved up and down to attract the attention of the octopus. The octopus swam out and 'attacked' one of the two objects. The researcher then reinforced the response with either a bit of food or electrical shock from the end of a probe and removed the stimuli. Departures from typical training methods will be noted.

Boycott & Young (1956), in a study on octopuses' reactions to shape, provided the first, exploratory attempt to teach octopuses using simultaneous presentations. Only one animal was trained and procedures varied as the authors explored different possibilities. The subject was rewarded for choosing the smaller of two white squares (8 and 4 cm). Previous experiments by the authors suggested that octopuses more readily attacked the smaller of two squares in this size range. Initially, because of position preference and reluctance to respond after receiving punishment for an incorrect choice, the authors resorted to presenting each stimulus on the same side of the tank on every trial. On the fourth day, the octopus was responding with consistent success, so presentations from then on were on alternating sides. The octopus was successful 85 % of the time in days 4–9 when the rewarded stimulus was presented on alternate sides.

Sutherland & Muntz (1959) performed a series of four experiments to explore the possibility of simultaneous discrimination training with octopuses. The authors provide no indication that they were aware of Boycott & Young's (1956) attempt. Three of these experiments included simultaneous discrimination training. No punishment was used, and animals were pretrained with presentations of the positive (rewarded) stimulus alone.

In Experiment 1, a single octopus attained 92 % success on a discrimination task between a vertical (rewarded, +) and a horizontal (unrewarded, −) rectangle, when both objects were moved vertically by hand. To determine if this performance was affected by a pre-existing preference, six subjects were then tested with moving rectangles and rewarded randomly. When the rectangles were moved vertically, subjects chose the vertical rectangle in 21 out of 22 trials; when the rectangles were moved horizontally, they chose the horizontal in 18 of 21 trials (Table 1). The authors concluded that the results of this first experiment could be explained by pre-existing preferences.

In Experiment 2, five animals were trained to discriminate between a rewarded 5 × 5 cm square and a circle of equivalent area. Two of the animals never attacked the square, even when presented alone. The other three completed 5, 14 and 60 trials each. This last subject was successful 28 % of the time, with no evidence of improvement. The authors concluded that octopuses must have a strong pre-existing preference for the circle. In confirmation of this, when they presented six experimentally naive animals with the white square and circle for four unrewarded trials, the white circle was chosen 71 % of the time. Interestingly, when they tried the same preference test with black figures, the black square was chosen 77 % of the time (not significant) (Table 1). (In Experiment 3, animals were trained sequentially.)

In Experiment 4, subjects were trained to discriminate between horizontal and vertical rectangles. For three subjects the vertical was rewarded and for three the horizontal was rewarded. No evidence for improvement was found; vertical positive animals succeeded 88 % of the time, whereas horizontal positive animals were successful 38 % of the time. Choices of the horizontal rectangle were significantly different between the two groups, indicating a weak effect of training on choices.

In 1962, Muntz *et al.* used simultaneous discrimination procedures in three experiments designed to study the effects of the function of the vertical lobe of the octopus brain. Here, I review only the performance of control and pre-operative animals.

In Experiment 1, animals were trained to discriminate between white and black squares. A permanent divider separated the two stimuli, which were presented by hand. In preliminary preference tests, four subjects chose the white square 83·4 % of the time (Table 1). Subjects were trained against this preference, therefore, with fish rewards for choosing the black square and shock as punishment for choosing the white square. Subjects attained a success of 91·6 %.

In Experiments 2 and 3, a more complex procedure was followed, as described in Sutherland *et al.* (1963) (see below). (Note: the authors cite Sutherland *et al.* (1963) in describing their methods, despite their own earlier date of publication.) In Experiment 2, animals were trained to discriminate between a square and a parallelogram, in a balanced design (half of the subjects rewarded for choosing the square and half for choosing the parallelogram). The eight subjects attained a success of 93 % by the end of a month of training. In Experiment 3, the stimuli were a 'V' and a sideways 'W' shape. The eight subjects attained a success of 75 % after an unspecified number of trials. In these two experiments, evidence for improvement was not provided.

Rhodes (1963) published the first and only attempt to mechanize simultaneous discrimination training for octopuses. The stimuli were attached to rods which were in turn attached to a motor that moved the stimuli up and down, eliminating hand motions. In one group of two subjects, stimuli were circles differing in diameter by 1 cm, and animals were rewarded for choosing the smaller. In the other group, the four subjects were trained to discriminate between circles whose diameters differed by 2 cm. For two of these subjects, the smaller of the circles was rewarded and for the other two, the larger of the circles was rewarded. Initial success for the first group was approximately 55 % and increased to 90 % after 70 trials; for the second group initial success was almost 70 % and increased to about 85 % after the same number of trials. Although statistics are not presented, improvement is significant if one assumes homogeneity between the individuals within each group ($\chi^2 = 9{\cdot}90$, $4{\cdot}29$, d.f. = 1, $P < 0{\cdot}05$).

Sutherland *et al.* (1963) devised a non-mechanized method that avoided hand presentations of objects in their study of simultaneous discrimination and transfer of learning. Stimuli were attached to rods fastened to a horizontal bar which rested across the far end of the octopus's tank. When the bar was rotated through 90 degrees, the objects moved through an arc into the water and rested in front of the far tank wall away from the octopus's den. A partition was attached to the horizontal bar so that the two stimuli were separated. An opaque partition with a transparent door directly in front of the octopus's den forced a delay between the introduction of the stimuli and the octopus's response. This ensured that the subject did not simply respond to the first stimulus seen. A further modification of procedure was the incremental introduction of shock as punishment for incorrect responses.

One group of subjects was trained with a square and with a parallelogram having angles of approximately 60 degrees; for the other group the parallelogram angles were closer to 80 degrees. The latter was an easier discrimination than the former. The group trained with the easier discrimination attained a success of 93 % in 250 trials. The group trained with the more difficult discrimination attained only 54 % success in 170 trials. Although 54 % was not significantly better than random, choices on unrewarded transfer trials with the easy parallelogram indicated learning had, in fact, occurred.

In Mackintosh & Mackintosh (1963), eight octopuses were trained to discriminate between black and white vertical rectangles in order to study reversal learning and the role of irrelevant cues. Only the initial task will be discussed here. Partitions similar to those used by Sutherland *et al.* (1963), were used to separate the two stimuli and to ensure a central start position. No mention was made of a transverse bar being rotated to introduce stimuli; I assume objects were supported by hand. Subjects were trained until they reached a criterion of 90 % success over 2 d of 10 trials each. Only average errors to criterion were published. These were so few (2·67) that animals must have reached criterion almost immediately.

In Mackintosh & Mackintosh (1964*a*), octopuses were trained with repeated reversals. Subjects were trained in one simultaneous discrimination task; once they attained a criterion the task was reversed and the rewarded object became unrewarded. Procedures followed those of Sutherland *et al.* (1963). Subjects completed from seven to 14 reversals. The criterion was eight correct out of 10 trials within 1 d. Evidence for significant learning across successive reversals was found in subjects' more quickly attaining a success of seven correct in 10 trials and their increasing levels of success on days 1 and 2 of each reversal.

In Mackintosh & Mackintosh (1964*b*), the effect of overtraining was explored. The same procedures were used as in Sutherland *et al.* (1963). Sixteen subjects were trained to discriminate between shapes resembling square brackets, one opening up and the other opening down. The criterion was 15 correct out of 20 trials (2 d). Half of the subjects were overtrained by presenting 100 further trials of this same task. Then all subjects were trained to discriminate between an upside down V and an M. Results were presented as mean trials and errors to criterion. On the first discrimination, subjects reached criterion in an average of 6 d. On the second discrimination the non-overtrained group reached criterion in 6 d, whereas the overtrained group required on average 8 d. This result indicated that overtraining on one task impeded the learning of another.

Sutherland *et al.* (1965) studied the effect of pretraining along different dimensions. They trained 16 octopuses on either size, with shape as an irrelevant cue, or shape, with size as an irrelevant cue. The same octopuses were then trained on a task where both shape and size were relevant. Finally, animals were tested with a task where the dimension that was relevant for them initially was now irrelevant, and *vice versa*. Procedures followed those in Sutherland *et al.* (1963). Cumulative success during training averaged 73 % for the first task and 77 % for the second task. Improvement was not indicated. Final testing indicated that pretraining affected later learning.

Messenger (1977) used simultaneous discrimination training in a colour vision experiment. Six subjects were presented with a total of four different discrimination tasks. Two tasks were presented in each session, in alternating trials. Subjects were trained with vertical (+) versus horizontal rectangles, and black (+) and white, yellow (+) and grey, and violet (+) and grey vertical rectangles, all presented against a 'creamy white' background. Success during the first two discriminations was 96 % and 88 % over 56 trials; for the hue discriminations, it was 50 % and 47 % over 24 trials. Performances were not evaluated for change through time; there appears to be evidence for improvement only for the black/white discrimination. The author noted that vertical is preferred to horizontal; black has also been reported as preferred to white on light backgrounds (Bradley & Messenger, 1977; Young, 1968). The authors concluded that the lack of success on the hue discriminations indicated that octopuses are colour blind.

Fiorito & Scotto (1992) studied observational learning in a simultaneous discrimination between red and white balls. The coloured balls were attached to the ends of rods and presented to demonstrator animals by hand. Two electrodes protruded from the back of each ball; the flow of current was controlled by buttons on the top of each rod. A piece of fish was skewered onto the back of the positive ball; an electric shock of 12 V was administered in the case of an incorrect response. Nine animals were trained to choose the white ball and 18 were trained to choose the red ball. Tests of untrained subjects indicated a preference for red (Table 1). All demonstrator animals attained a success of 100 % in the last 10 trials of the 40 training trials given.

Dawes *et al.* (1963) replicated Fiorito & Scotto's (1992) observational learning experiment. The procedures appear to have been identical to those of Fiorito and Scotto except that all animals were pretrained with a single black ball. Results were consistent with those of Fiorito and Scotto. Only the abstract was published; this experiment will not be discussed further because of the lack of procedural details.

In Boal (1993), I reported three-object simultaneous discrimination training; however, one experiment used just two objects. Subjects were four *Octopus bimaculoides*. White bivalve and gastropod shells were attached to rods which were inserted into a bracket which latched onto the top of the aquarium during presentations. As the experimenter, I stood behind barriers arranged in such a way that I could neither see nor be seen by the octopuses until they had chosen one of the stimuli. For two subjects, choosing the bivalve was rewarded, and for the two others, choosing the gastropod was rewarded. After 10 d, tasks were reversed. No evidence for learning was found. All animals preferred the bivalve to the gastropod, and further trials did not improve performances.

III. NEW EXPERIMENTS

In order to address questions about training procedures, I present several new, previously unpublished experiments that also used simultaneous visual discrimination training. In each case, subjects were *Octopus vulgaris*. Unless otherwise specified, procedures were the same as in the typical experiment described above.

(1) *Fiorito, Experiment* 1 (*personal communication*)

Method

Procedures followed those described in Fiorito & Scotto (1992) unless otherwise stated. Stimuli were red and white balls, attached to the ends of plastic rods. Stimuli were placed into the octopuses' tanks and not moved again until the end of the trial. During training, food was attached directly to the positive stimulus and an 8-volt AC electric current was delivered directly to the negative stimulus. Animals were trained to a criterion of 100 % in two consecutive sessions of five trials each. After a three to five day break in training, animals were trained again (overtrained) to the same criterion.

Results

Animals trained to choose the white ball (against preference) showed clear evidence of improvement. Cumulative success during training averaged 76 %. Cumulative success during overtraining averaged 87 %. The improvement in average success between the two bouts of training was significant (Wilcoxon signed ranks test, $T+ = 15$, $n = 5$, $P < 0.05$), providing evidence that this discrimination was learned.

Animals trained to choose the preferred red ball, while clearly successful at significantly better than chance (89·9 % cumulative success during training and 90·85 % during overtraining), showed no improvement across training trials. Three out of four animals were actually less successful during overtraining trials. No learning can be inferred for animals trained to choose the preferred red ball.

(2) *Boal, Experiment* 1

Subjects were 12 *Octopus vulgaris*. The same bivalve and gastropod shells were used as in my earlier experiment (Boal, 1993). Tank lids were hinged in two places, making it possible to open just one end at a time. At the start of each trial, the lid at the end closest to the experimenter was lifted. If the octopus was not in sight, the stimuli were introduced. If the octopus was in sight, the experimenter held the lid open until the octopus came all the way to the near end of the tank; the lid was then closed, the lid at the far end opened, and the stimuli introduced by hand.

Experiment 1 *a*

Method

The first group of subjects ($n = 2$) was trained with horizontal (+) and vertical rectangles, the second group ($n = 3$) with gastropod (+) and bivalve shells, and the third group ($n = 3$) with bivalve (+) and gastropod shells. Positive reinforcement was fish; no negative reinforcement was given. On days 1 and 2, a guest experimenter presented the stimuli to the first group only.

Results

In 80 trials, only one animal (bivalve (+) group) showed significant improvement, when performances on days 1–3 were compared with performances on days 8–10 (36–79% successful; $\chi^2 = 19$, d.f. = 1, $P < 0.05$). No other evidence for learning was found.

Responses of the two animals in first group were not significantly different from each other (Wilcoxon–Mann–Whitney $W = 37$, $m = n = 6$, $P > 0.40$). The performances of these two subjects were significantly better on the 2 d when the guest experimenter trained them than on any of the following 4 d with the usual experimenter (Kruskall–Wallis and Multiple Comparisons tests, KW = 7.7, $n = 12$, $P < 0.05$). Performances of the animals in the other two groups were compared over the same 6-d period to check that external conditions were not somehow responsible for this difference. No difference was found; subjects were slightly but insignificantly less successful on those same 2 d (Wilcoxon Signed Ranks Test, $T+ = 18$, $n = 9$, $P > 0.40$). It was not clear why results between experimenters differed.

In this experiment, up to five correction trials were provided. The number of consecutive errors during corrections did not change over the course of training (for three groups, $\chi^2 = 0.45$, 1.70, 0.68; $P > 0.05$ each), a result consistent with that of Sutherland & Muntz (1959).

Experiment 1 b
Method

All animals from Part 1 plus four new animals were trained with gastropod (+) and bivalve shells (against pre-existing preference, Boal, 1993). There were four groups of subjects, therefore: New task, Same task, Reverse of old task, and Naive. In an attempt to increase motivation, small live crabs were substituted for pieces of fish as reward for correct responses.

Results

On the first day of trials with shells, for all animals (Parts 1 and 2), the bivalve was chosen 71% of the time (Wilcoxon Signed Ranks Test, $T+ = 176$, $n = 12$, $P \sim 0.00$), a clear confirmation of bivalve preference.

Only one animal (group Naive) showed significant improvement with training (27–53% successful; $\chi^2 = 5.45$, d.f. = 1, $P < 0.05$). The combined shell choices of groups Same and Reverse differed significantly ($\chi^2 = 3.789$, d.f. = 1, $P < 0.05$), evidence that training did have some effect. Curiously, subjects were significantly more successful at the end of Experiment 1 *b* than they were at the start of Experiment 1 *a* (Wilcoxon Signed Ranks Test, $T+ = 97$, $n = 16$, $P < 0.05$), despite the mix of tasks involved. This result provides further evidence that learning did occur.

Experiment 1 c
Method

Eight animals from this same group of subjects were trained with horizontal and vertical rectangles in a balanced design. Reward was a piece of fish while punishment was an 8-volt shock. On days 7 and 9 a second and third experimenter performed training trials.

Results

Response rates deteriorated from 93 % on day 1 to 30 % on day 9, at which point the experiment was discontinued. No evidence for learning was found. In order to determine if reduced responses were a result of deteriorating health, I compared subjects' responses to free crabs with responses of experimentally naive animals. I found no differences in response times between the two groups (Wilcoxon–Mann–Whitney $z = -0.59$, $P \sim 0.28$).

The experimenter could be seen during trials only when stimuli were presented in the front of the tank. As a check against possible inadvertent experimenter cuing during trials, success was compared between presentations at the front and at the back of the tanks. No difference was found (Wilcoxon Signed Ranks Test, vertical (+) $T+ = 9$, horizontal (+) $T+ = 29$, $n = 9$ both, $P > 0.40$). No learning at all was found during this experiment.

No differences in success were found between experimenters 1, 2 and 3, or within days for experimenter 1 (Friedman two-way analyses of variance $F_r = 0.06$, 8.73; $P > 0.05$ both).

(3) *Boal, Experiment* 2

Method

Four animals were trained in larger tanks (60 × 80 cm) than in Experiment 1 (30 × 100 cm). Stimuli were grey and white plastic squares, with two animals rewarded for choosing grey and two for white. Correct choices were rewarded with pieces of fish attached to the back of the square. Incorrect responses resulted in an 8-volt shock from an auxiliary probe.

Results

Average success for all four animals was greater than 85 % at the end of the 80 possible training trials. The three animals that continued responding throughout training all demonstrated significant improvement over the course of training. The animals rewarded for choosing grey completed 25 and 49 trials ($\chi^2 = 17.1$ and 6.36, d.f. $= 1$, $P < 0.05$); the animals rewarded for choosing white completed 72 and 17 trials ($\chi^2 = 14.4$ and 2.33, d.f. $= 1$, $P < 0.05$ and $P > 0.20$).

(4) *Boal, Experiment* 3

In this experiment, training tanks were larger still (60 × 100 cm). Procedures were modelled after those of Fiorito & Scotto (1992), in consultation with Fiorito. Stimuli were black and white bivalve shells.

Experiment 3 *a*
Method

One animal was trained with white rewarded, the other with black rewarded. The correct shell had a piece of fish fastened onto the back. No punishment was used.

Results

No evidence for learning was found. The black shell was preferred. Initially, the animal rewarded for choosing black chose black 90 % of the time; after 95 trials the black was chosen 67 % of the time. The animal rewarded for choosing white chose black a steady 67 % of the time.

Experiment 3b
Method

Subjects were rewarded for choosing the white shell (not preferred) and incorrect choices resulted in an 8-volt shock from electrodes attached to the back of the black shell. One subject was experimentally naive while the other had been rewarded for choosing black in the previous part of the experiment (a reversal). After training, the non-reversed subject was tested for possible effects of extraneous cues. The experimenter remained out of sight and presentation order followed Fellows (1967) sequences. Further conditions were as follows: (i) food attached to both stimuli while the electrical cord was still plugged in (current leakage possible); (ii) food attached to both stimuli but the electrical cord unplugged; and (iii) no food and electrical current unplugged from the wall socket.

Results

Both animals demonstrated significant learning. The reversed animal responded in only 18 trials. In the first nine trials, the black (previously rewarded) shell was chosen eight times. In the last nine responses, seven were to the white shell. This difference is significant ($\chi^2 = 40.5$, $P \sim 0.00$). The naive animal completed 77 trials. Success in the first 10 choices was 56%; in the last 10 choices it was 100% ($\chi^2 = 9.0$, d.f. = 1, $P < 0.01$).

In the tests for extraneous cuing, in (i) success was 67 % (four white and two black chosen), in (ii) success was 60 % (three white and two black chosen), and in (iii) success was 89 % (eight white and one no response). Only in the last case was success significantly greater than chance 50 % ($\chi^2 = 0.67, 0.20$, $P > 0.80$), which suggests that cues from food and possibly shock may influence choices.

IV. ANALYSIS OF EXPERIMENTAL PROCEDURES

In order to address as many concerns about simultaneous training with octopuses as possible, I review all of the above experiments in terms of conditions and procedures. Experiments are summarized in Table 2. I discuss each variable in turn; the corresponding section and column in Table 2 is indicated within the text by the associated number in parentheses. Separate experiments within a single published report are numbered (1, 2, 3, ...); if the same subjects were used in more than one experiment then letters instead of numbers were assigned (e.g. 3*a*, 3*b*). Table entries in parentheses were inferred from indirect sources of information.

Non-parametric statistical procedures were used to determine the relationship between procedural variables and the experimental outcomes (30) (Siegel & Castellan, 1988; Wilkinson, 1990). It was not clear whether outcome should be considered as a dichotomous learn/not-learn variable or whether all four categories (yes, maybe, weak, no learning) should be included. Because we are interested in finding any possible procedural biases influencing experimental outcome, the statistic presented is the more sensitive of the two. Actual statistical values are provided any time the associated probability was less than 0.10, and are summarized in Table 3.

(1) *Subjects*

Subjects in all simultaneous discrimination experiments were *Octopus vulgaris*, with the exception of Boal (1993), who used *O. bimaculoides*. Learning experiments have been conducted with other species, including *Eledone cirrhosa* (Angermeier & Dassler, 1992), *E. moschata* (Milhailoff, 1920; von Uexkull, 1894), *O. bimaculoides* (Hanlon & Forsythe, 1984), *O. cyanea* (Crancher *et al.*, 1972; Papini & Bitterman, 1991), and *O. hummelincki* (Pack, 1979). In Boal (1993) no difference was found between results from *O. vulgaris* and *O. bimaculoides* in 3-object discriminations. I do not yet have enough data to compare learning abilities between species further, but it would not be surprising if differences were found that reflect differences in ecology and life history.

(2) *Maintenance*

Octopuses are notorious for escaping from their tanks, particularly soon after capture. The animals (250–500 g) used by Fiorito (1992 and unpublished) did not attempt to climb out of his 60 × 100 cm tanks, even when lids were left off for many hours. These tanks had clear lids and one side of the tank was glass. Boal's (1993) tanks also provided a window, having glass fronts and tops. In these exceptionally small tanks (30 × 45 cm), most animals (usually less than 200 g) did not try to climb out, but did position themselves so they could watch any activity in the lab through the front glass wall. In both the 30 × 100 completely opaque tanks and the 60 × 80 opaque-sided tanks with clear lids (J. G. Boal, unpublished), most animals did attempt to climb up and out when the lids were not fastened down. As far as I know, tanks used in the other experiments reviewed all had opaque sides and lids which provided only a small amount of light through ventilation holes and no opportunity for animals to see any movement or activity around them. It appears that windows, as well as space and perhaps lighting are all relevant factors in providing satisfactory housing for octopuses.

Octopuses eat a wide variety of prey, including fresh or frozen crabs, fish, bivalves, gastropods, and other cephalopods. Live crabs are preferred. Some crustaceans and/or molluscs in the diet are necessary for the production of haemocyanin (Wells, 1978). Many researchers who have used bits of fish as rewards during training have supplemented this fare with a live crab at the end of the day. Wells (*ibid.*) points out that food deprivation for an invertebrate does not necessarily increase motivation and can depress it; on the other hand, response rates of satiated octopuses are unpredictable.

Bitterman (1966) claimed that octopus experiments were compromised by poor aquarium facilities at the Stazione Zoologica in Naples, where virtually all of the octopus learning experiments have been performed. I do not have sufficient information to assess past aquarium conditions. All recent behavioural work at the Stazione Zoologica involves adequate awareness of animal behaviour, health and appropriate environmental conditions (Fiorito *et al.*, 1990).

Further information on octopus maintenance can be found in Anderson (1987), Van Heukelem (1977) and Walker *et al.* (1970). Nixon (1969) provides free-feeding levels for *Octopus vulgaris* of differing weights.

Table 2(*a–e*). *Summary of procedures used in simultaneous discrimination experiments on octopuses*

Table 2(*a*). *Apparatus*

1 Reference number	2 Experiment or group	3* Space (width × length)	4* Specific start loc.	5* Dividers between stimuli	6* Stimuli: (P) preferred, (+) rewarded, (−) punished or not rewarded	7 Background
1†	1	40 × 100	Yes	Simple	White squares 4 (+) or 8 (−) cm	(Grey)
2	1	100 × 200	Yes	Complex	White rectangles (10 × 2 cm), vert (+) (P)/horiz (−)	(Grey)
3	2	100 × 200	Yes	Simple	White square (+)/circle (P) (−)	(Grey)
4	4	60 × 100	Yes	Simple	White rectangles vert/horiz	(Grey)
5	1	60 × 100	Yes	Simple	Squares black (+)/white (P) (−)	(Grey)
6	2	60 × 100	Yes	Complex	White square (5 × 5 cm)/ parallelogram (angle 45 deg.)	(Grey)
7	3	60 × 100	Yes	Complex	White 'V'/sideways 'W'	(Grey)
8	1	40 × 100	(Yes)	None	White circles (4·5–10 cm)	(Grey)
9	1	60 × 100	Yes	Complex	White square (P)/ parallelogram (angle ~ 60 deg.)	(Grey)
10	2	60 × 100	Yes	Complex	White square (P)/ parallelogram (angle ~ 80 deg.)	(Grey)
11	1	—	Yes	Complex	Vertical rectangles, black/white	(Grey)
12	1	60 × 100	Yes	Complex	Rectangles (10 × 4 cm), black/white	(Grey)
13	1*a*	60 × 100	Yes	Complex	'U' and inverted 'U', both with squared corners	Black
14	1*b*	60 × 100	Yes	Complex	'M' and inverted 'V'	Black
15	1*a*	60 × 100	Yes	Complex	'X', cross, and box, each 96, 24 and 6 cm^2	(Grey)
16	1*b*	60 × 100	Yes	Complex	Inverted 'V' and sideways 'W', each 36 and 16 cm^2	(Grey)
17	1*a*	60 × 100	Yes	Simple	Vertical rectangles (10 × 2 cm), black (P) (+)/white (−)	Off-white
18	1*b*	60 × 100	Yes	Simple	Rectangles (10 × 2 cm), vertical (P) (+)/horizontal (−)	Off-white
19	1*c*	60 × 100	Yes	Simple	Rectangles (10 × 2 cm), yellow (+)/grey (−)	Off-white
20	1*d*	60 × 100	Yes	Simple	Rectangles (10 × 2 cm), violet (+)/grey (−)	Off-white
21	1	60 × 100	Yes	None	Balls (3 cm) white/red (P)	Window
22	6	30 × 45	No	None	White shells (6–7 cm), bivalve (P)/gastropod	Black
23	1*a*	60 × 100	Yes	None	Balls (3 cm), red (P) (+)/white (−)	Window
24	1*b*	60 × 100	Yes	None	Balls (3 cm), white (+)/red (P) (−)	Window
25	1*a*1	30 × 100	No	None	White bivalve (P)/ gastropod (5–7 cm)	Black/ brown
26	1*a*2	30 × 100	No	None	White rectangles (2 × 10 cm), horiz. (+)/vert. (P) (−)	Black/ brown
27	1*b*	30 × 100	No	None	White shells (5–7 cm), gastropod (+)/bivalve (P) (−)	Black/ brown
28	1*c*	30 × 100	No	None	White rectangles (2 × 10 cm), vertical (P)/horizontal	Black/ brown
29	2	60 × 80	Yes	None	Squares (5·6 cm) white/grey	Grey/brown
30	3*a*	60 × 100	Yes	None	Bivalve shells (4 cm), white/black (P)	Window
31	3*b*	60 × 100	Yes	None	Bivalve shells (4 cm), white (+)/black (P) (−)	Window

† References: 1, Boycott & Young (1956); 2–4, Sutherland & Muntz (1959); 5–7, Muntz *et al.* (1962); 8, Rhodes (1963); 9–10, Sutherland *et al.* (1963); 11, Mackintosh & Mackintosh (1963); 12, Mackintosh & Mackintosh (1964*a*); 13–14, Mackintosh & Mackintosh (1964*b*); 15–16, Sutherland *et al.* (1965); 17–20, Messenger (1977); 21, Fiorito & Scotto (1922); 22, Boal (1993); 23–24, Fiorito (unpublished); 25–31, Boal (unpublished).

Table 2(*b*). *Methods*

Reference number	8 Positive reinforcement	9* Negative reinforcement	10* Reinforcement attached to stimuli	11* Position in tank	12* Support of stimuli	13* Movement of stimuli	14 Experimenter in view
1	Crab	Shock (8–12 V)	No	High	Hand	'Gently'	Yes
2	Sardine	None	No	High	Hand	~ 3 cm, 3 × /sec	Yes
3	Sardine	None	No	High	Hand	~ 3 cm, 3 × /sec	Yes
4	Sardine	None	No	High	Hand	~ 3 cm, 3 × /sec	Yes
5	Sardine	Shock (8 V)	No	High	Hand	~ 3 cm, 3 × /sec	Yes
6	Sardine	Shock (0–8 V)	No	High	Bracket	Rotation	Yes
7	Sardine	Shock (0–8 V)	No	High	Bracket	Rotation	Yes
8	Crab	None	No	High	Motor	~ 3 cm, 1 × /sec	Yes
9	Sardine	Shock (0–8 V)	No	High	Bracket	Rotation	Yes
10	Sardine	Shock (0–8 V)	No	High	Bracket	Rotation	Yes
11	Sardine	None	No	High	(Hand)	(Hand)	(Yes)
12	Fish	None	No	High	Bracket	Rotation	Yes
13	Sardine	Shock (8 V)	No	High	Bracket	Rotation	Yes
14	Sardine	Shock (8 V)	No	High	Bracket	Rotation	Yes
15	Sardine	Shock (8 V)	No	High	Bracket	Rotation	Yes
16	Sardine	Shock (8 V)	No	High	Bracket	Rotation	Yes
17	Sardine	None	No	High	Hand	3 × /sec	Yes
18	Sardine	None	No	High	Hand	3 × /sec	Yes
19	Sardine	None	No	High	Hand	3 × /sec	Yes
20	Sardine	None	No	High	Hand	3 × /sec	Yes
21	Fish	Shock (12 V)	Yes	Low	Hand	25 × /min	Yes
22	Squid	Water Squirt	No	High	Bracket	Placement	No
23	Anchovy	Shock (10 V)	Yes	Low	Hand	Placement	Yes
24	Anchovy	Shock (10 V)	Yes	Low	Hand	Placement	Yes
25	Anchovy	None	No	High	Hand	1·5 × /sec	Mixed
26	Anchovy	None	No	High	Hand	1·5 × /sec	Mixed
27	Crab	None	No	High	Hand	1·5 × /sec	Mixed
28	Anchovy	Shock (8 V)	No	High	Hand	1·5 × /sec	Mixed
29	Anchovy	Shock (8 V)	Yes	Low	Hand	1·5 × /sec	Yes
30	Anchovy	None	Yes	Low	Hand	1·5 × /sec	Yes
31	Anchovy	Shock (8 V)	Yes	Low	Hand	1·5 × /sec	No

Table 2(*c*). *Procedures*

Reference number	15* Inter-trial interval	16* Trials per session	17* Sessions per day	18* Total days	19* Total trials	20* Pre-training trials	21* Correction trials	22 Order (maximum repeats)
1	2 h, then 30 min	5–14	—	9	80	0	No	Alternating
2	—	7–16	1	4	44	4	Unlimited	—
3	—	10	1	6	60	28	Unlimited	—
4	—	10	1	4	120	6	Max 5	Random (2)
5	5 min	4	2	6	48	—	S+ alone	—
6	15 min	5	2	∼ 30	300	> 20	No	Random (3)
7	15 min	5	2	—	—	> 20	No	Random (3)
8	3–4 h	3	1	27	80	10	No	Random
9	15 min	5	2	10	100	> 20	No	Random (3)
10	15 min	5	2	17	170	> 20	No	Random (3)
11	30 min	5	2	—	—	10	S+ alone	Random
12	20 min	10	1	55	Mean 52	20	No	Random
13	15–20 min	5	2	6	Mean 56	30	S+ alone	Random (3)
14	15–20 min	5	2	7	Mean 66	—	S+ alone	Random (3)
15	20 min	4–5	2	10	90	> 20	No	Random (3)
16	20 min	4–5	2	4	36	10	No	Random (3)
17	10 min	8	2	10	56	12	Unlimited	‘Semi-random’
18	10 min	8	2	10	56	0	Unlimited	‘Semi-random’
19	10 min	8	2	10	24	12	Unlimited	‘Semi-random’
20	10 min	8	2	10	24	12	Unlimited	‘Semi-random’
21	5 min	5	3	10	40	> 5	Max 2	Random
22	9 min	5	2	20	100	25	No	Random (3)
23	2–3 min	5	3–4	5	28	> 5	Max 2	Random
24	2–3 min	5	3–4	5	56	> 5	Max 2	Random
25	15 min	8	1	10	80	0	Max 5	Random
26	15 min	8	1	10	80	0	Max 5	Random
27	15 min	8	1	9	72	0	No	Random
28	5 min	5	2	9	90	0	No	Random
29	5 min	5	3	6	80	7	Max 3	Fellows (1967)
30	3 min	5	3	7	95	5	Max 3	Random
31	3 min	5	3	4	55	5	Max 3	Random

Table 2(*d*). *Reported results*

Reference number	23 Success criterion	24* Variances reported	25* Final N	26 Response rate (%)	27 Information reported	28 Final or cumulative (c) success (%)
1	—	Yes	1	75	Raw data	71
2	—	Yes	1	—	Responses × time	92
3	—	Yes	1	—	Cumulative responses	28c
4	—	No	6	100	Cumulative responses	38 horiz. (+) c 88 vert. (+) c
5	—	No	4	—	Mean × time	91
6	—	No	8	—	Final success	93
7	—	Yes	8	—	Final success	75
8	—	No	6	—	Mean × time	87
9	—	Yes	8	—	Responses × time	84
10	—	Yes	8	—	Responses × time	54
11	18 out of 20	Yes	8	—	Errors to criterion	27·5 errors
12	8 out of 10	Yes	10	—	Mean errors	19·03 errors
13	15 out of 20	No	16	—	Means to criterion	75
14	15 out of 20	Yes	16	—	Means to criterion	80
15	10 days	Yes	16	—	Cumulative responses	72c
16	4 days	No	16	—	Cumulative responses	65c
17	7 days	No	6	100	Mean × time	88
18	7 days	No	6	100	Mean × time	96
19	3 days	No	6	100	Mean × time	50
20	3 days	No	6	100	Mean × time	47
21	5 out of 5	No	44	> 74	Trials to criterion	—
22	—	Yes	4	69	Responses × time	62 biv. (+) c 38 gastr. (+) c
23	10 out of 10	—	4	94	—	90c
24	10 out of 10	—	5	88	—	76c
25	—	—	6	94	—	74 biv. (+) c 27 gastr. (+) c
26	—	—	2	94	—	56c
27	—	—	11	72	—	34c
28	—	—	8	71	—	49 horiz. (+) c 53 vert. (+) c
29	—	—	3	59	—	70 grey (+) 80 white (+)
30	—	—	2	92	—	67, 67
31	—	—	2	18, 77	—	78, 93

Table 2(*e*). *Results and interpretations*

Reference number	29 Adequate evidence provided	30** Evidence for learning	31 Reasons/Comments
1	No	—	Preference rewarded, no improvement
2	No	—	Preference rewarded, no improvement
3	Yes	No	No improvement
4	Yes	Weak	No improvement but significantly different in horizontal choices
5	Yes	Yes	Significantly better than chance, improvement
6	Yes	Maybe	Significantly better than chance, balanced design, but improvement unreported
7	Yes	Maybe	Significantly better than chance, balanced design, but improvement unreported
8	Yes	Yes	Significantly better than chance, semi-balanced design, improvement
9	Yes	Yes	Significantly better than chance, semi-balanced design, improvement
10	Yes	Weak	Chance success, no improvement, but apparent transfer to new task
11	Yes	Yes	Significantly better than chance, balanced design, improvement across reversals
12	Maybe	Maybe	Improvement across successive reversals but continuing success not tested
13	Yes	Maybe	Significantly better than chance, balanced design, but improvement unreported
14	Yes	Maybe	Significantly better than chance, balanced design, but improvement unreported
15	Yes	Yes	Significantly better than chance, balanced design, improvement
16	Yes	Maybe	Significantly better than chance, balanced design, but improvement unreported
17	No	—	Improvement but preference rewarded
18	No	—	No improvement, preference rewarded
19	Yes	No	No improvement, chance success
20	Yes	No	No improvement, chance success
21	Maybe	Maybe	Significantly better than chance, continued success, balanced design, but nonsignificant criterion
22	Yes	No	Chance success, no improvement
23	No	—	Preference rewarded, no improvement
24	Yes	Yes	Significantly better than chance, improvement
25	Yes	Weak	Chance success, weak improvement but significant effect on transfer
26	Yes	Weak	Chance success, weak improvement
27	Yes	No	Poor success, no improvement
28	Yes	No	Chance success, no improvement
29	Yes	Yes	Significantly better than chance, improvement
30	Yes	Yes	Significantly better than chance, improvement
31	Yes	Yes	Significantly better than chance, improvement

Variables included in the Correspondence Analysis are indicated by asterisks on their column numbers.
**Evidence for Learning (column 30) was superimposed onto the results of the analysis (see text).

Table 3. *Relationships of procedural variables to learning*

(Variables not listed were not included in statistical analyses.)

Column in Table 2	Procedural variables	Direction of association with learning	Significance $^{**}P < 0{\cdot}01$ $^{*}P < 0{\cdot}05$	Importance in correspondence analysis
3	Space		*	
4	Start location	+ +	**	
5	Dividers	+	*	
6	Stimuli			
	Brightness	+ +	**	×
	Hue			
	Orientation			
	Shape			
	Size			
9	Negative reinforcement	+	**	× ×
10	Reinforcement on stimuli	+	**	×
11	Position in tank	High –	**	
12	Support of stimuli	Hand –	**	
13	Movement	–	*	
15	Intertrial interval	Longer –	**	× ×
16	Trials per session	More – –	**	
17	Sessions per day	More + +	**	× ×
18	Total days			
19	Total trials			
20	Pretraining			
21	Correction trials	–	*	×
24	Variances reported			
25	Final N			

(3) *Training procedures*

(*a*) *Training tanks*

Tank size (*a*-3) appeared to affect experimental results (see Table 2*a*). No strong evidence for learning was found in the most narrow tanks (30 cm wide). One possible explanation is that when an octopus 'attacks', its arms radiate out to both sides of its head and body. Even when the line of approach is directly at one stimulus, one of the side arms may encounter the alternative stimulus first. In this case, the animal will sometimes move its body centre over so as to envelop the alternative stimulus. This problem is avoided in the wider tanks; of course, adequate tank width must depend on the size of the octopus.

This effect from tank size is only weakly apparent statistically (Mann–Whitney $U = 52{\cdot}5$, d.f. $= 1$, $P = 0{\cdot}06$), perhaps because the experiments finding no learning included not only all the experiments in the narrowest tanks but also the experiment in the widest tank (Sutherland & Muntz, 1959). In this latter experiment, the test tank was not a home tank, so subjects were moved at the start of each session. In no other experiment were animals moved for training trials.

The effect of tank size appears to be separate from the effect of the distance between stimuli. Fiorito & Scotto (1992) obtained evidence for learning when they presented objects as close together as Boal (1993 and unpublished Experiment 1), who found no evidence for learning. The widths of the tanks in Fiorito & Scotto were twice that of Boal, however. It may be that in the wider tanks, octopuses could approach stimuli either directly or from the outside, whereas in the narrower tanks, the stimuli were directly against the outside walls of the tank, forcing the octopuses to a more central path.

In most experiments, trials did not begin until the subject was in a predetermined start location (*a*-4), usually a den made of bricks. In Boal's unsuccessful experiments (1993 and unpublished Experiments 1 and 2), stimuli were presented at the opposite end of the tank from wherever the octopus was currently located. All of the experiments that demonstrated satisfactory evidence for learning used specific start locations. The difference in outcomes with and without specific start locations is significant (Mann–Whitney $U = 40{\cdot}0$, d.f. $= 1$, $P < 0{\cdot}01$).

Either fixed or temporary dividers (*a*-5) separating the two stimuli were provided in some experiments (see both 'simple' and 'complex'). Either served to eliminate Tarzan-like responses to both stimuli at once. Simultaneous responses to both objects were more of a problem in the narrower tanks. Boal (1993 and unpublished) always scored simultaneous responses as errors while Fiorito (1992 and unpublished) scored such responses as a mistrial rather than an error. It is not known to what extent this difference affected reported success rates, but the effect is unlikely to be substantial. None of the other published reports specified how such responses were scored.

In some experiments, additional screen barriers were fixed across the width of tanks, between the octopuses' homes and locations where the stimuli were presented ('complex'). A transparent door in the screen was raised a few seconds after the objects had been rotated or placed into the tank. This procedure insured that subjects did not simply grab the first object seen. Not all experimenters used this method. Experiments with successful outcomes were slightly but not significantly more likely to use dividers than were unsuccessful experiments (Kruskall–Wallis KW $= 7{\cdot}331$, d.f. $= 3$, $P = 0{\cdot}06$).

(*b*) *Stimuli* (*a*-6, *a*-7)

Discriminations based on object brightness were significantly associated with evidence for learning (Mann–Whitney $U = 40{\cdot}0$, d.f. $= 1$, $P < 0{\cdot}01$). None of the other discriminating characteristics (hue, orientation, shape and size) was associated with experimental outcome. None of the experiments demonstrating an absence of learning or only weak learning offered discriminations that could be solved using brightness cues.

Pre-existing preferences for some stimuli over others have been demonstrated in simultaneous, non-rewarded trials (Table 1). Both the strength and the direction of preferences appear to be affected by contextual factors such as background brightness (*a*-7), direction of motion, and number of trials. Many papers report preferences apparent during or after training, or determined during sequential presentations; however, some of these results appear to contradict each other (e.g. smaller object preferred, Boycott & Young, 1956; larger object preferred, Rhodes, 1963; Sutherland *et al.*, 1965). I have generalized across papers only when preferences were demonstrated in simultaneous, non-rewarded trials, or when consistent results were reported by multiple authors. A review of object preferences can be found in Wells (1978).

It is not certain that any learning occurs as a result of training when animals are rewarded for a pre-existing preference. A high degree of success may only indicate a strong pre-existing preference. Even more worrisome, improvement across training trials is not necessarily evidence for learning. Fiorito & Scotto (1992) found that a preference for red over white of 53 % on day 1 increased to a 78 % preference on day 5 when given non-rewarded trials across all 5 d (Table 1). Success significantly greater than chance, even coupled with evidence for improvement, is thus not adequate evidence that learning occurred when pre-existing preferences are rewarded.

In some experiments, a balanced design was implemented in which half of the subjects were rewarded for choosing one stimulus while the other half were rewarded for choosing the second stimulus. In some of these experiments, only the combined success of the two groups was reported. If one of the stimuli were preferred, it is impossible to know how much learning occurred as a result of training and how much of the reported success was due to the high success of the group rewarded consistent with preference. One might think that exceptional success of one of the groups would be counterbalanced by exceptionally poor success of the other group; however, performances as unbalanced as 88 % and 38 % have been reported (Sutherland & Muntz, 1959). In this case, the average, 63 %, is in the range of other reportedly successful demonstrations of learning (e.g. Sutherland *et al.*, 1965). In fact, the authors provide some evidence that these animals did demonstrate learning. To avoid this problem, I recommend that success for the two groups be reported separately.

All the reviewed experiments except Boal (1993 and unpublished) used artificial stimuli. While it may be argued that natural objects ought to be easier to discriminate because of their increased complexity, multiple relevant cues, and inherent saliency, these possible advantages appear to be offset by strong and persistent pre-existing preferences. The apparent preference for bi-valves over gastropods perhaps reflects natural dietary preferences (Mather, 1991*b*).

(*c*) *Protocol*

The responses octopuses make to stimuli are less clear-cut than, for example, a pigeon's peck. Sometimes subjects swim or crawl towards the stimuli and reach out tentatively with one arm or several to touch one or both stimuli. At the other extreme, sometimes subjects launch an all-out pounce and quickly and completely envelop just one of the stimuli. In the latter case, there is no problem in scoring the response; it is responses like the former that can cause considerable difficulty and provide room for possible subjectivity. Some researchers used dividers between stimuli to clarify subjects' choices (see above). This method is not fail-safe since it is nearly impossible to provide a divider longer than an octopus can reach yet short enough that the subject can see both stimuli from any realistic start location. In the reviewed experiments, response criterion was seldom stated explicitly. Boal (1993) used the perhaps overly conservative criterion of the first object touched by an arm or sucker, and Mackintosh & Mackintosh (1964*b*) used the criterion of a 'grasp' of one of the objects; all others simply report scoring which of the objects was 'attacked'. With enough space, 'envelop' is probably a reasonable criterion, where envelop means the mouth is close to or touches one of the stimuli and the interbrachial web surrounds that same object.

It is a bigger problem to score the more tentative responses that arise in difficult discriminations; a workable compromise could involve rewarding the subject with the benefit of the doubt but scoring the response conservatively. In several recent experiments (J. G. Boal unpublished, Experiment 3, Fiorito & Scotto, 1992; Fiorito unpublished, Experiment 1), trials were video-taped. Experimenters other than the trainer reviewed the video-taped trials, a procedure that can verify consistent scoring of outcomes.

Positive reinforcement (*b*-8) in all experiments has been food. Octopuses will work for a variety of different food rewards. The primary concern is practicability of delivery. The food chosen must remain intact when cut small and skewered onto the delivery device. Early experimenters used live crabs as food rewards but this forced long intertrial intervals since octopuses often hoard them (Pack, 1979) and consume them comparatively slowly. Crabs are also generally larger than the fish rewards used. J. G. Boal (unpublished) found that response rates were not significantly higher when rewards were live crabs as compared to bits of previously frozen anchovy, despite the octopuses' clear preference for the former. The successes attained by Rhodes (1963) using crab rewards are not notably superior to those reported by other researchers using fish rewards. In most experiments, response rates were not provided, so further comparison is not possible.

Negative reinforcement (*b*-9) in the form of electric shock was used in many, but not all, experiments. Several researchers report decreasing response rates with shock, especially in difficult discrimination tasks (Sutherland *et al.*, 1963; Sutherland & Muntz, 1959; J. G. Boal unpublished). When shock was omitted, however, the animals did not readily change their behaviour, especially when trained against pre-existing preferences (*ibid.*). Sutherland introduced shock in stages, from 0 to 8 V, to surmount these difficulties (Sutherland *et al.*, 1963, followed by Muntz *et al.*, 1962). For all experiments reviewed, experiments using shock were significantly more likely to have successful experimental outcomes than were experiments not using shock (Mann–Whitney $U = 36{\cdot}0$, d.f. $= 1$, $P < 0{\cdot}01$).

In most experiments, food and shock were provided with probes separate from the rods supporting the discriminanda. This procedure requires a small delay between a subject's response and the arrival of the reinforcement. The octopuses become very alert to the motions of the experimenter in order to try to anticipate and avoid potential shock. To avoid these limitations, reinforcement has been attached directly to the stimuli (*b*-10) in some experiments. This procedure was more common in experiments with octopuses given successive rather than simultaneous presentations. In simultaneous experiments, attaching food or electrodes directly to stimuli resulted in significantly greater success in experimental outcomes than when reinforcement was provided separately (Mann–Whitney $U = 55{\cdot}0$, d.f. $= 1$, $P = 0{\cdot}05$); indeed, all experiments with reinforcement attached directly to stimuli demonstrated evidence for learning.

If positive and negative reinforcement are directly attached to the stimuli, however, subjects could solve discrimination problems directly if they detected the reinforcers from a distance. Boyle (1986*b*) reported that octopuses became active with even quite low concentrations of food material in the water, Lee (1992) demonstrated chemotaxis in a *Y*-maze, and Joll (1977, as cited by Boyle 1986*a*) suggested that distance chemoreception was in fact quite significant for finding prey in the field. Crancher *et al.* (1972) designed a device in which octopuses could reach into a tube that extended up out of the water in order to obtain food. J. G. Boal (unpublished) attempted to adapt this procedure to discrimination training by providing a choice of tubes. The experiment was unsuccessful because once traces of food existed on the inside of one of the tubes, all animals focused all further attention on that same tube, regardless of spatial location of other cues (unpublished). In test trials, Boal found that attaching food to the backs of both stimuli disrupted performances on a learned discrimination (unpublished, Experiment 3*b*). Evidently, chemical cues from food are detected and can affect octopus behaviour.

Electrodes placed in seawater set up an electric field that could be detectable to octopuses unless the current to the electrodes was completely shut off. Although I have found no experimental data testing the possible influence of electrical current on choice behaviour (but see J. G. Boal unpublished Experiment 3*b*), when electrodes are attached directly to stimuli, it would be prudent to check frequently for current leakage past the source switch.

Post-training sessions demonstrating very short response times with clear movement trajectories or success in trials without reinforcement would provide evidence that octopuses are not adopting such alternative solutions to discrimination problems. Of the reviewed experiments, Fiorito & Scotto (1992), Fiorito (unpublished Experiment 1) and Boal (unpublished Experiments 2 and 3) are vulnerable to these alternate explanations of success. Response times have not been reported in any of the experiments I reviewed. In Boal's experiments, successful animals were tested without reinforcement; success reported is that during non-reinforced test trials.

It is important to note the distinction between training for learning, when any method that works is fine, and testing for learning, when rigorous controls are necessary. Of course, if the behaviour of the animal during training is of interest, it would be prudent to provide both positive and negative reinforcement on separate probes from the stimuli, as most researchers have done.

Bitterman (1975) referred to rough wrestling matches with octopuses in order to recover stimuli grasped during trials. If shock is given, the animals quickly release the objects, and more quickly with greater familiarity with shock. Rhodes (1963) noted, and I have also found, that even without shock or when reinforced with food, octopuses will let go of stimuli on their own, given a little time. Initially, a little time may mean several minutes and some encouragement by the experimenter in the form of a bit of gentle tugging. After a few days, most animals spontaneously release objects within 30 s.

Presentation method varied across experiments. Boycott & Young (1956) suspended stimuli on threads, which caused rotation problems with two-dimensional stimuli. In all other experiments, stimuli were fastened to rods. In most experiments, the objects were at least several centrimeters above the floor of the tank; in some experiments, objects were placed directly on the substrate, facilitating a more natural attack response (see Table 2*b*, column 11, 'high' and 'low'). All of the experiments that used bottom placement had successful experimental outcomes; the difference in outcomes between high and low placement is significant (Mann–Whitney $U = 105{\cdot}0$, d.f. $= 1$, $P = 0{\cdot}05$).

In many experiments, objects were supported directly by hand; in others, objects were attached to brackets (*b*-12). If optimistic experimenters unwittingly bias experimental outcomes, success with hand presentations would be greater than success when brackets or machines are used. On the other hand, if non-standard presentations introduce extraneous cues which distract subjects, success with standardizing brackets or machines would result in improved success as compared to hand presentations. For the reviewed experiments, the use of brackets of some kind resulted in significantly more successful experimental outcomes than when stimuli were supported by hand (Kruskall–Wallis KW = 7·742, d.f. = 3, $P = 0{\cdot}05$), which provides preliminary evidence for the distraction hypothesis.

In most octopus discrimination experiments, subjects choose between moving stimuli (*b*-13). Boycott & Young (1950, p. 26) reported that, '...movement is absolutely necessary; a crab which remains still, even in full view of an octopus, is not attacked'. While there is good contradictory evidence to this initial observation (Wodinsky, 1971; Sutherland, 1963), motion does appear to attract the attention of the octopuses and facilitate the detection of stimuli. Sutherland reported that if animals never see the stimuli in motion, they require a large number of pretraining trials with the positive shape alone before they regularly attack it (Sutherland, 1963; Sutherland *et al.*, 1963). In all of the experiments reviewed here, octopuses were exposed to some movement of stimuli, ranging from the placement of the stimuli into their tanks at the start of each trial to continuous movements up and down throughout trials. Many researchers provided motion by hand. A slight but insignificant negative relationship was found between movement rate and outcome (Mann–Whitney $U = 109{\cdot}5$, d.f. = 1, $P = 0{\cdot}08$). Published experiments using hand movements were slightly less successful than experiments in which stimuli were not moved by hand during trials (Mann–Whitney $U = 21{\cdot}0$, d.f. = 1, $P = 0{\cdot}10$).

Experimenters were in view of their subjects during trials (*b*-14) in most of the reviewed experiments. Boycott & Young (1956) acknowledged the undesirability of this arrangement but found it practically unavoidable. They reported similar responses of octopuses when different researchers manipulated the figures and when tank lids remained in place except for providing an opening for the introduction of stimuli. In J. G. Boal (unpublished Experiment 1), stimuli were presented in the front or the back of the tank, depending on the initial location of the subject. When stimuli were presented in the front, the experimenter was in view, whereas when the stimuli were presented in the back, the experimenter was out of sight. No difference was found in success between the front and the back of tanks. These subjects were unsuccessful on the discrimination task, either way, however. Because experimenters were in view in virtually all of the reviewed experiments (exceptions: Boal, 1993 and unpublished Experiment 3*b*), it was not possible to test statistically for any differences in experimental outcomes as a result of inadvertent visual cues by experimenters.

An obvious concern with hand presentations (especially 'high', un-anchored positions), hand movements, and experimenter visibility during trials is the possibility that experimenters might have unwittingly provided extraneous cues facilitating (or hindering) learning performances. Octopuses are undoubtedly highly attentive to activity in the laboratory (see discussion of holding tanks, above). While it would be quite interesting if octopuses could solve discriminations using some reliable cue or cues that experimenters unwittingly provided, it would mean that little if anything can currently be concluded about their ability to solve simultaneous discrimination problems.

In none of the published papers is it reported which person or how many people did the actual training. Since no experiment was fully automated, there is room for inter-experimenter variability affecting outcomes. Boycott & Young (1956) reported that, 'figures manipulated by different observers, without consultation, produced similar responses by an octopus' (p. 493). Likewise, Boal found no significant differences in subjects' successes when trained with guest experimenters (unpublished Experiment 1 *c*). Neither of these two experiments demonstrate simultaneous discrimination learning (see Table 2*e*, 30 and 31); it would be interesting to know whether one would find similar results in experiments that did provide strong evidence for learning.

Experimental outcomes could also be affected by intra-experimenter variability, either from one day to the next or from one animal to the next. For example, I always moved left to right down the row of octopus tanks when I trained my subjects. In hindsight, I suspected that I was more patient in initial trials, possibly due to increasing boredom and fatigue. In a post-hoc analysis of success versus tank order for all subjects trained in Boal (1993) on either two- or three-object discrimination tasks, and unpublished Experiments 1 and 2 ($n = 61$), this hypothesis was supported; animals in the left-most tanks were more likely to be successful (average ranks in first through fourth quarters of subjects were 1·84, 2·05, 2·50 and 2·71; the probability of ordering this extreme is 1/24, $P = 0·04$). (Assignments of newly arriving animals to particular tanks was haphazard.) The only experiment that deviated strongly from this pattern was Boal (1993) Experiment 6, in which the experimenter could neither see nor be seen by subjects during trials, and stimuli were not hand-held (rank order 3, 4, 1, 2). This finding suggests that octopuses' performances can be affected by small differences in presentations.

(*d*) *Timing*

Intertrial intervals (*c*-15) varied between experiments. Positive results were found with a wide range of intertrial intervals (see Table 2). In a short experiment, Boal compared responses of experienced subjects (unpublished Experiment 1 *b*) to presentations of a single white plastic square, using intertrial intervals of 5 or 15 min (2 d; two sessions per day; four trials per session but the first trial not counted; hand presentations). Animals were rewarded with fish for each attack. Responses were scored from 0 to 3 based on speed and whether the responses involved the whole body or just one or two arms. The octopuses were significantly more responsive with intertrial intervals of 5 min than they were with intertrial intervals of 15 min (Wilcoxon Signed Ranks Test, $T+ = 34·5$, $P < 0·02$). In the reviewed experiments, no significant relationship was found between intertrial interval and experimental outcome, however.

Both the number of trials per session (*c*-16) and the number of sessions per day (*c*-17) varied between experiments. Clearly, satiation must set outside limits, which must vary with the size of the octopuses, the size of the food rewards, and the water temperature. Successful outcomes were positively associated with fewer trials per session (Mann–Whitney $U = 133·5$, d.f. = 1, $P < 0·01$) and more sessions per day (Mann–Whitney $U = 30·0$, d.f. = 1, $P < 0·01$).

Activity cycles affect the time of day that trials can profitably be given. I find that *Octopus vulgaris* is quite inactive in the early afternoons. The reviewed experiments with *O. vulgaris* were all conducted during the day, but more precise times were not provided. It is not known if there are circadian rhythms in learning (Bradley & Young, 1975; Kovaceric & Rakic, 1971; see discussion in Wells *et al.*, 1983).

The total number of days (*c*-18) or trials (*c*-19) required for octopuses to learn a simultaneous discrimination depends on the difficulty of the task. In the reviewed experiments, total numbers of trials range from 24–300. Few researchers report whether training continued without interruptions 7 d a week. Fiorito & Scotto (1992) trained their subjects only during a normal 5-d work week and did not note any decrement in performance between Friday and Monday. Boal (1993) also found no consistent pattern due to weekends without trials. There was no significant relationship between the total length of experiments, measured by trials or time, and final outcome.

(*e*) *Training trials*

It has been common practice, and not only in octopus work, to pretrain subjects (*c*-20) by presenting the positive stimulus alone for some number of trials before presenting the two stimuli simultaneously. In most of the reviewed experiments, subjects were pretrained with the positive stimulus. In some cases, a set number of trials was given (indicated directly in Table 2*c*); in others, subjects were required to meet some predetermined criterion, such as a total of 5, or perhaps 5 out of 5, prompt responses (indicated as '> 5', for example, in Table 2*c*). Interestingly, Sutherland & Muntz (1959) report that, 'Where animals did give evidence of learning by the method of simultaneous discrimination, learning was extremely quick, and seemed to be complete after about 10 presentations of the positive figure on its own in association with reward [pretraining]. No animal gave any evidence of improving its performance in the course of the simultaneous discrimination training' (p. 122). Sutherland *et al.* (1963) found similar results when training without punishment; one group of animals was successful 70% of the time on the first day of simultaneous trials following pretraining. They did not improve thereafter until shock was added. Of the experiments I reviewed, in no case was simultaneous discrimination learning demonstrated when the animals had not been pretrained with the positive stimulus. No relationship was found between the number of pretraining trials and experimental outcome.

A pretrained animal is facing a different problem from that in which the identity of the positive stimulus is initially unknown. Unless its performance improves across the simultaneous presentations, the animal has demonstrated only the discrimination of a familiar, rewarded stimulus from an unfamiliar and non-rewarded or punished stimulus, rather than demonstrating a learned discrimination between two simultaneously presented stimuli. It was not always reported whether performances improved significantly during the course of discrimination trials.

Correction trials (*c*-21) have been used by some experimenters as a means of interrupting position preferences (e.g. Sutherland & Muntz, 1959). In some experiments, after an incorrect response, experimenters presented the rewarded stimulus alone in the same location; in others, they repeated the trial exactly, using both stimuli, either for a set maximum number of repetitions or indefinitely (up to eight or more times). In theory, it is important that data analyses include only the first trial when correction trials are given; otherwise the alternate strategy of 'same location if no reward' will be reinforced. In fact, over the course of training, a reduction in the number of repetitions required after an initial failure was not found by either Sutherland & Muntz (1959) or by J. G. Boal (unpublished Experiment 1). In the reviewed experiments, the negative relationship between number of correction trials and experimental outcome was not significant (Mann–Whitney $U = 110{\cdot}0$, d.f. = 1, $P = 0{\cdot}10$).

In two-object simultaneous presentations, location is usually irrelevant to the discrimination task; the rewarded object may be either on the right or the left. In order to ensure that alternative response strategies are not inadvertently rewarded or interpreted as discrimination learning, care must be taken with the order in which presentations are made (left, right) (*c*-22). In most papers reviewed here, presentation was randomized, although in at least one case order was actually haphazard (personal communication), which is not at all the same thing. Several papers do not report their method. I found that midway through training I could predict with an accuracy of better than 70% which object the subject would take on any given trial simply by noting which eye was looking out of the brick den at the start of the trial (unpublished Experiments 2 and 3*b*). Although this predictability was not maintained, it is a clear reminder that presentation order must be determined before the session begins. Fellows (1967) published a useful set of left/right series that controls for spurious success associated with plausible alternate choice strategies.

Success criterion (*d*-23) must, of course, depend upon the question asked, and requires careful choice. In some of the reviewed experiments, criterion was defined solely as attaining a set level of success. A single performance, alone, cannot normally be considered adequate evidence for learning. If experiments are long or open ended, it is likely that such a criterion will eventually be met even without learning. If variances in daily performances are high, the probability increases that spurious successful performances will occur on any one day. High variance was evident in the experiment of Boal (1993), where success ranged from 10 to 100% for one individual in 10 d of trials (indeed, in two consecutive days). In this experiment, no evidence for improvement was demonstrated. Variabilities in performances within individuals were not reported in any of the other papers I reviewed; some also failed to include any evidence for improvement. A more convincing criterion for learning would include a consistent as well as non-random level of success.

Variability between subjects was also rarely reported explicitly (exceptions: Sutherland *et al.*, 1963; Fiorito & Scotto, 1992). When sample sizes are as small as usual in the reviewed experiments, homogeneity between individuals in preferences and choice behaviour cannot be assumed between subjects. Fortunately, most papers did account statistically for variability between subjects (*d*-24). Combined totals of choices of all subjects for each stimulus provide little interpretable information (e.g. portions of both Sutherland & Muntz, 1959; Experiment 4, and Messenger, 1977). I recommend that either tests for significant differences between individuals be shown to be negative, or behaviour be compared within individuals, before and after training, using paired statistics. Non-parametric tests such as the Sign Test or the Wilcoxon Signed Ranks Test provide appropriate measures.

In many experiments, not all subjects that began experiments finished them. The sample size (*d*-25) reported in Table 2*d* is the final sample size. Some animals simply quit responding reliably; some failed to learn. Animals dropped from a learning experiment because of failure to learn should be carefully recorded, of course.

Response rates (*d*-26) were seldom reported. Low response rates may arise from poor water quality, extreme temperatures, satiation, or from repeated failure. A serious concern in data analysis when trials are successive, low response rates are less of a concern here. Boal (1993) found that for difficult tasks, presentations of a single, always rewarded stimulus randomly interspersed between training trials resulted in improved response rates.

V. ANALYSIS OF EXPERIMENTAL OUTCOMES

I performed statistical analyses in order to determine if there were any consistent relationships between procedures and outcomes. In my analyses, I adopted conservative criteria for the satisfactory demonstration of discrimination learning. Each experiment was evaluated on the basis of the information provided explicitly about simultaneous discrimination training (*d*-27–28; *e*-29–31). Firm evidence for discrimination learning consisted of (*a*) evidence against simply rewarding preference, (*b*) improvement in success with simultaneous training, (*c*) final performances significantly better than chance, and (*d*) sustained final success as evidence against spurious successful performances. Many of the experiments reviewed were not designed specifically to address discrimination learning; partially as a consequence of this, not all provide sufficient data to be certain that discrimination learning occurred.

Five experiments were excluded from further consideration. Four of these were excluded because all animals were rewarded for a known pre-existing preference. In three of these, no evidence for improvement across trials was provided (Boycott & Young, 1956; Messenger, 1977; Experiment 1*a*, Sutherland & Muntz, 1959; Experiment 1). In the fourth (Messenger, 1977; Experiment 1*b*), improvement with time was demonstrated; I chose to exclude this experiment in the light of Fiorito & Scotto's (1992) demonstration of non-rewarded strengthening of preferences with time (Table 1). The fifth experiment excluded was Dawes, *et al.* (1993), in which only the abstract was published.

In my analysis of the remaining 22 experiments, I included all the procedural variables where there was some variability between published experiments, and few missing values. The variables included are indicated by asterisks by their column numbers (Table 2). Variables were coded categorically and in such a way as to reduce some of the heterogeneity between variables (see Appendix). My results apply to coded variables, not original values, therefore. I chose not to correct for the apparent parabolic relationship between tank width and experimental outcome because such an assumption rested on only one experiment in an exceptionally large tank.

I used ordination techniques (Correspondence Analysis; Pielou, 1984) to examine which procedural variables were important to the coded experimental outcomes (*e*-30) in the remaining 22 experiments. This is not a statistical method but rather a tool for better understanding relationships between experiments and procedural variables. Experimental outcome was not included as a component in the analysis; it was superimposed onto the axes afterwards.

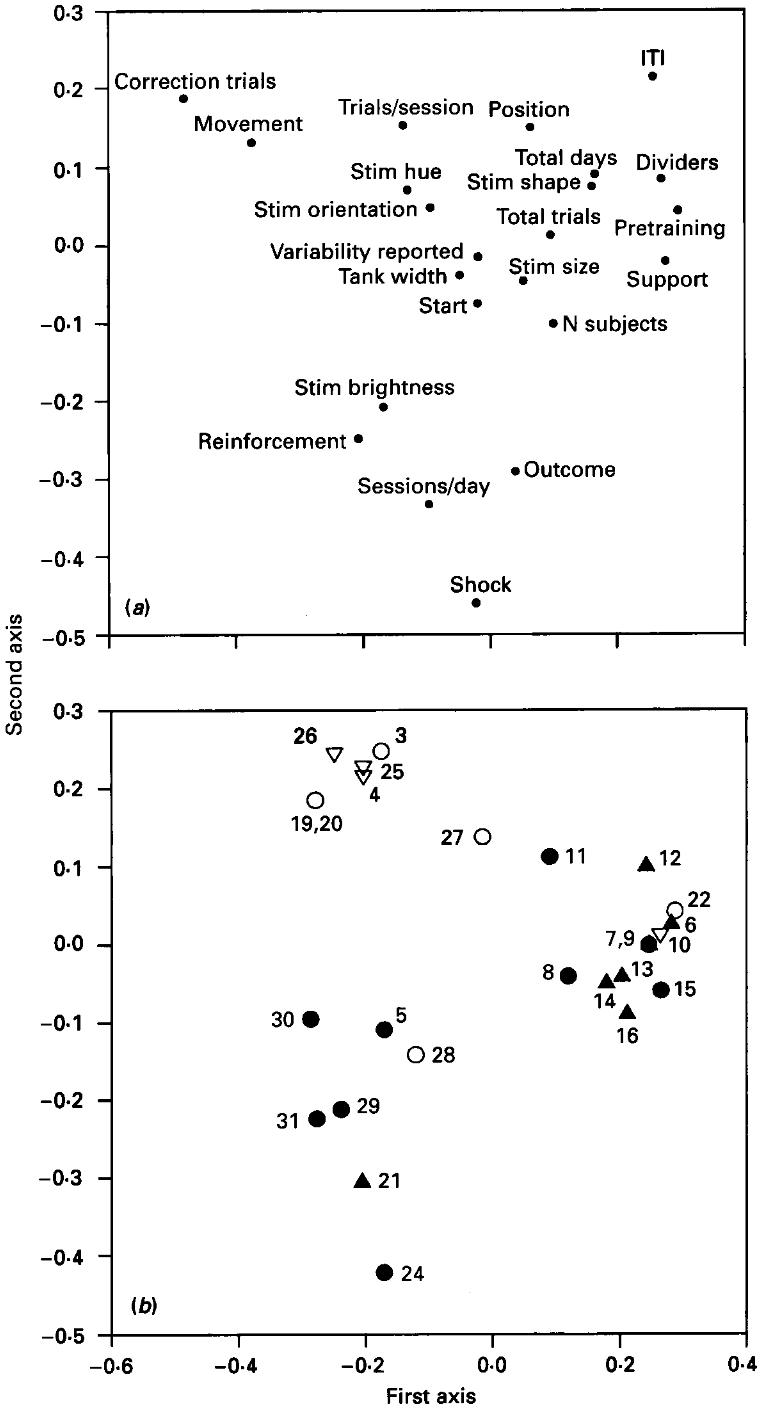

Fig. 1. Graphical representation of the results of a Correspondence Analysis of the reviewed experiments. One set of axes was computed; for ease of viewing, (*a*) procedural variables, and (*b*) individual experiments were plotted separately. In (*b*), evidence for learning (outcome) for each experiment is indicated: ●, yes; ▲, maybe; ▽, weak; and ○, no.

The first axis explained 34% of the variability in the data set but only 0·8% of the variability in experimental outcome, indicating a weak relationship between outcome and the overall structure of the data set. For the first axis, the primary components were correction trials (−31%), movement of stimuli (−21%), and number of pretraining trials (14%) (Fig. 1*a*). Learning was weakly associated with fewer correction trials, less movement of stimuli, and fewer trials per session.

The second axis explained only 19% of the variability in the data set but 42% of the variability in experimental outcome. For the second axis, the primary components were negative reinforcement (−34%), number of sessions per day (−18%), and length of the intertrial interval (10%) (Fig. 1*a*). Further components included number of correction trials (7%), reinforcement attached directly to stimuli (−6%), and discriminations based on object brightness (5%). Learning was associated with greater negative reinforcement, more sessions per day, and shorter intertrial intervals, as well as fewer correction trials, reinforcement attached directly to stimuli, and discriminations based on object brightness.

The third axis provided little further information, explaining only 12% of the variability in the data set and 0·5% of the variability in experimental outcome.

In the space defined by the first and second axes, experiments sorted into three clumps (Fig. 1*b*). For experiments with negative values on the first axis (more correction trials, more movement of stimuli, and less pretraining), there was clear separation in the direction of the second axis between experiments with outcomes indicating good evidence for learning (first clump) and poor or no evidence for learning (second clump). For these experiments, experimental outcome can be well explained by the procedural variables determining the second axis (negative reinforcement, number of sessions per day, and intertrial interval).

For experiments with positive values on the first axis (fewer correction trials, less movement of stimuli, and more pretraining), there was no separation of experiments by outcome (third clump) (Fig. 1*b*). The second axis has little explanatory power for this group of experiments.

Three experiments demonstrating no evidence for learning were not included in the second clump, described above. All were performed in the narrowest tanks. Had tank width been coded to take into account the apparent parabolic relationship between tank width and outcome, these points may have been grouped more closely with others showing similar outcome.

In a further analysis, Spearman correlation coefficients were computed for all procedural variables against outcome. The only variables with coefficients greater than 0·50 were trials per session (−0·67), sessions per day (0·63), object brightness (0·60) and specific start location (0·58).

VI. DISCUSSION

In this review, I have uncovered procedural differences that appear to influence experimental outcome significantly. Are any of these procedural variables likely to have been subjectively influenced? What evidence do we have for inadvertent cues influencing experimental outcomes?

The variables most related to experimental outcome can be seen in Table 3. Most are clearly objective factors: sessions per day, object brightness, trials per session, start location, use of shock and intertrial interval. Stimuli supported and moved by hand, and those held up in the water column rather than stabilized against the tank bottoms, were all slightly negatively associated with learning; clearly success was not enhanced by these three potentially subjectively influenced variables. I conclude that any subjective biases of optimistic researchers have not systematically inflated experimental successes.

Octopus performances do appear to be influenced by inadvertent cues, however. When presentations were manual, moved, and held high in the water column they were almost certainly less standardized than in cases where brackets, motors and bottom placement were implemented. These seemingly minor variations between presentations appear to have compromised success. Even subtle differences resulting from the order in which subjects received each trial apparently influenced outcomes (Boal experiments). As far as I know, experimenters were in view of subjects in all experiments other than Boal (1993 and unpublished Experiments 1*c* and 3*b*); it remains possible that experimenter's eyes or body provided systematic cues. Careful experimental work remains to be done to clarify exactly what procedural differences are responsible for influencing octopuses' choice.

Eight of the 32 experiments were performed by me; it is worth considering whether my own experimental results are in some way atypical and biased my statistical analyses. Eliminating these experiments reduced the data set to the point where a Correspondence Analysis was no longer reliable. In Fig. 1*b*, we can see that my experiments are not clumped together, nor are they scattered outliers. I believe, therefore, that my possible over-representation in the data set has not distorted my conclusions.

Can we be completely certain, in any of the experiments, that discrimination learning did occur? If the experimenter was, in fact, hidden from view during Rhodes' (1963) mechanized experiment, we would have firm evidence that octopuses can learn to discriminate simultaneously presented stimuli without any potential help from extraneous cues provided by the experimenter. Unfortunately, Rhodes did not specify this important detail.

Most octopus discrimination experiments have used successive rather than simultaneous presentations of stimuli. Using successive presentations, Maldonado (1963, 1965) demonstrated discrimination learning in octopuses, with fully automated experiments. The experimenter remained out of sight in these experiments; however, reinforcements were attached directly to the back of stimuli, providing another potential source of inadvertent cues. I have not found any discrimination learning experiment using octopuses as subjects that fully controlled for all potential sources for extraneous cues.

It is possible that all 16 of the successful simultaneous experiments I reviewed here could have been systematically biased by subjects viewing the experimenters during trials, so that they could have produced evidence for discrimination learning through unsystematic, unintentional cuing. I consider it more reasonable, however, to conclude that octopuses are indeed capable of simultaneous visual discrimination learning but that their behaviour is sensitive to the exact experimental procedures implemented.

VII. CONCLUSIONS

There is reasonable evidence that octopuses can learn to solve simultaneous visual discrimination problems successfully. Although there is good evidence that some procedural variables do have significant influences on experimental outcomes, I found no evidence that sub-optimal experimental designs biased experimental outcomes in any significant and systematic way. Nevertheless, Bitterman (1966, 1975) certainly raised important procedural concerns, and I also found many imperfections in experimental designs. I cannot rule out that performances in simultaneous discriminations are subject to important and as yet unknown inadvertent experimenter cues. A well designed experiment addressing the issue of cuing in simultaneous discrimination experiments remains to be done.

I found no reason to reject the conclusions about learning in most of the papers I reviewed. Of the experiments I rejected from my analysis, one was intended only as a preliminary, exploratory attempt at simultaneous discrimination training (Boycott & Young, 1956), and in another, the authors noted the confounding influence of object preference (Sutherland & Muntz, 1959).

There remains the question of why animals were successful in some experiments and not others. I found several possible reasons including inadequate tank size (Boal, 1993; and unpublished Experiment 1), insufficient number of trials (Messenger, 1977; Experiments 1*c* and 1*d*) and strong pre-existing preferences coupled with lack of negative reinforcement (Boal, 1993, and unpublished Experiments 1*a* and 1*b*; Sutherland & Muntz, 1959, Experiment 2).

I have several procedural recommendations as a result of my analysis. I recommend providing tanks with windows and of sufficient width, given the size of animals, for easy dichotomous choice. I also recommend requiring a specific start location, administering shock as negative reinforcement, and using separate probes to supply all reinforcements. Learning appears to be facilitated by more training sessions per day with fewer trials each. I suggest that naive animals be tested for pre-existing preferences for particular stimuli before experiments begin. I encourage future published experiments to include such information as variability in performances, and initial success as well as final success, so that readers can confirm for themselves that performances did indeed improve as a result of training. There is no reason why future experiments could not be mechanized to assure standardized presentations. In any case, the use of 'blind' experimenters in testing and scoring subjects' responses would provide important assurance of the objectivity of experimental results.

VIII. SUMMARY

I have presented a review and critique of the procedures employed in simultaneous discrimination training experiments using octopuses as subjects. Procedural variables were analysed statistically for their influence on experimental outcome. The variables most significantly associated with successful discriminations included use of a specific start location for subjects, shock as negative reinforcement, fewer trials per session, more sessions per day, and discriminations based on stimulus brightness. No experiment controlled all potential sources of inadvertent cues, and subjects' performances appeared to be sensitive to exact procedural details. The most common

practice diminishing evidence for learning involved reward that coincided with the subject's pre-existing preferences. I found no evidence that sub-optimal experimental designs biased experimental outcomes in any significant and systematic way. Although there is insufficient reason for rejecting results of published simultaneous discrimination training experiments, careful, conclusive experiments remain to be performed.

IX. ACKNOWLEDGEMENTS

I would like to thank G. Coppola, D. A. Eckerman, G. Fiorito, M. Scardi, R. H. Wiley, and two anonymous experimenters for their assistance with this project. G. Fiorito provided valuable information about his experiments, including the details of one previously unpublished experiment. B. U. Budelmann, R. T. Hanlon, A. Packard, R. H. Wiley and J. Z. Young all contributed helpful comments on drafts of this paper. I am grateful to The National Institute of Child Health and Human Development, postdoctoral fellowship 5F32HD07686, for my financial support. This work would not have been possible without the generous assistance of staff and the use of facilities at the Stazione Zoologica 'Anton Dohrn' of Naples, Italy, the University of North Carolina at Chapel Hill, North Carolina, and the University of Texas Medical Branch in Galveston, Texas.

X. REFERENCES

ANDERSON, R. C. (1987). Cephalopods at the Seattle Aquarium. *International Zoo Yearbook* **26**, 41–48.

ANGERMEIER, W. F. & DASSLER, K. (1992). Inhibitory learning and memory in the lesser octopus (*Eledone cirrhosa*). *Bulletin of the Psychonometric Society* **30**, 309–310.

BIEDERMAN, G. B. & DAVEY, V. A. (1993). Social learning in invertebrates. *Science* **259**, 1627–1628.

BITTERMAN, M. E. (1966). Learning in the lower animals. *American Psychologist* **21**, 1073.

BITTERMAN, M. E. (1975). Critical commentary. *Invertebrate Learning, Vol. 3 Cephalopods and Echinoderms* (ed. W. C. Corning, J. A. Dyal and A. O. D. Willows), pp. 139–145. Plenum Press, New York.

BOAL, J. G. (1993). An assessment of complex learning in octopuses. Doctoral Dissertation: The University of North Carolina, Chapel Hill, North Carolina.

BOYCOTT, B. B. (1954). Learning in *Octopus vulgaris* and other cephalopods. *Pubblicazione della Stazione Zoologica di Napoli* **25**, 67–93.

BOYCOTT, B. B. (1965). Learning in the octopus. *Scientific American* **212**, 42–50.

BOYCOTT, B. B. & YOUNG, J. Z. (1950). The comparative study of learning. *Symposium of the Society for Experimental Biology* **4**, 432–453.

BOYCOTT, B. B. & YOUNG, J. Z. (1956). Reactions to shape in *Octopus vulgaris* Lamarck. *Proceedings of the Royal Society of London, Series B* **126**, 491–547.

BOYLE, P. R. (1986*a*). Neural control of cephalopod behavior. In *The Mollusca. Vol. 9: Neurobiology and Behavior, Part 2* (ed. A. O. D. Willows), Academic Press, New York.

BOYLE, P. R. (1986*b*). Responses to waterborne chemicals by the octopus *Eledone cirrhosa* (Lamarck, 1798). *Journal of Experimental Marine Biology and Ecology* **104**, 23–30.

BRADLEY, E. A. & MESSENGER, J. B. (1977). Brightness preference in *Octopus* as a function of the background brightness. *Marine Behavior and Physiology* **4**, 243–251.

BRADLEY, E. A. & YOUNG, J. Z. (1975). Are there circadian rhythms in learning by *Octopus*? *Behavioral Biology* **13**, 527–531.

CRANCHER, P., KING, M. G., BENNETT, A. & MONTGOMERY, R. B. (1972). Conditioning of a free operant in *Octopus cyanea* Gray. *Journal of the Experimental Analysis of Behavior* **17**, 359–362.

DAVIS, H. & BALFOUR, D. A. (1992). *The Inevitable Bond: Examining Scientist-Animal Interactions.* Cambridge University Press, New York.

DAWES, J., FERNANDES, J. & ROBERTSON, J. D. (1993). *Octopus vulgaris* can learn by visual observation. *Anatomical Record, Suppl. 1b* **16**, 46.

DEWS, P. M. (1959). Some observations on an operant in the octopus. *Journal of the Experimental Analysis of Behavior* **2**, 57–63.

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

FIORITO, G. (1993). Social learning in invertebrates. *Science* **259**, 1629.

FIORITO, G. & SCOTTO, P. (1992). Observational learning in *Octopus vulgaris*. *Science* **256**, 545–547.

FIORITO, G., VON PLANTA, C. & SCOTTO, P. (1990). Problem solving ability of *Octopus vulgaris* Lamarck (Mollusca, Cephalopoda). *Behavioral and Neural Biology* **53**, 217–230.

HANLON, R. T. & FORSYTHE, J. W. (1984). Visual discrimination training of laboratory reared octopuses. *American Malacological Bulletin* **2**, 92.

JOLL, L. M. (1977). The predation of pot-caught western rock lobster (*Panulirus longipes cygnus*) by octopus. *Western Australia Department of Fish and Wildlife Report* **29**, 1–58.

KOVACERIC, N. K. & RAKIC, L. J. (1971). Circadian rhythm and visual discrimination in *Octopus vulgaris* Lamarck. *Archiv Bioloskih nauka, Beograd* **23**, 3–4.

LEE, P. G. (1992). Chemotaxis by *Octopus maya* Voss et Solis in a Y-maze. *Journal of Experimental Marine Biology and Ecology* **153**, 53–67.

MACKINTOSH, J. (1962). An investigation of reversal learning in *Octopus vulgaris* Lamarck. *Quarterly Journal of Experimental Psychology* **14**, 15–22.

MACKINTOSH, N. J. & MACKINTOSH, J. (1963). Reversal learning in *Octopus* with and without irrelevant cues. *Quarterly Journal of Experimental Psychology* **15**, 236–242.

MACKINTOSH, N. J. & MACKINTOSH, J. (1964*a*). Performance of *Octopus* over a series of reversals of a simultaneous discrimination. *Animal Behaviour* **12**, 321–324.

MACKINTOSH, N. J. & MACKINTOSH, J. (1964*b*). The effect of overtraining on a nonreversal shift in *Octopus*. *Journal of Genetic Psychology* **106**, 373–377.

MALDONADO, H. (1963). The positive learning process in *Octopus vulgaris*. *Zeitschrift für vergleichende Physiologie* **47**, 191–214.

MALDONADO, H. (1965). The positive and negative learning process in *Octopus vulgaris* Lamarck. Influence of the vertical and median superior frontal lobes. *Zeitschrift für vergleichende Physiologie* **51**, 185–203.

MATHER, J. A. (1991*a*). Navigation by spatial memory and use of visual landmarks in octopuses. *Journal of Comparative Physiology A* **168**, 491–497.

MATHER, J. A. (1991*b*). Foraging, feeding, and prey remains in middens of juvenile *Octopus vulgaris* (Mollusca: Cephalopoda). *Journal of Zoology* **224**, 27–39.

MESSENGER, J. B. (1977). Evidence that *Octopus* is colour blind. *Journal of Experimental Biology* **70**, 49–55.

MIKHAILOFF, S. (1920). Experiences reflexologiques; experiences nouvelles sur *Eledone moschata*. *Bulletin of the Institute for Oceanography, Monaco* **379**, 1–11.

MUNTZ, W. R. A., SUTHERLAND, N. S. & YOUNG, J. Z. (1962). Simultaneous shape discrimination in *Octopus* after removal of the vertical lobe. *Journal of Experimental Biology* **39**, 557–566.

NIXON, M. (1969). The time and frequency of responses by *Octopus vulgaris* to an automatic food dispenser. *Journal of Zoology* **158**, 475–483.

PACK, J. E. (1979). *Octopus*: apparatus and procedures for laboratory housing and for research on visual wavelength discrimination. Masters Thesis: Stephen F. Austin State University.

PACKARD, A. (1972). Cephalopods and fish: the limits of convergence. *Biological Review of the Cambridge Philosophical Society* **47**, 241–307.

PAPINI, M. R. & BITTERMAN, M. E. (1991). Appetitive conditioning in *Octopus cyanea*. *Journal of Comparative Psychology* **105**, 107–114.

PIELOU, E. C. (1984). *The Interpretation of Ecological Data*. John Wiley & Sons, New York.

RHODES, J. M. (1963). Simultaneous discrimination in *Octopus*. *Pubblicazione della Stazione Zoologica di Napoli* **33**, 83–91.

ROSENTHAL, R. A. & FODE, K. L. (1963). Three experiments in experimenter bias. *Psychological Reports* **12**, 491–511.

SANDERS, G. D. (1975). *The Cephalopods*. In *Invertebrate Learning. Vol. 3: Cephalopods and Echinoderms* (ed. W. C. Corning, J. A. Dyal and A. O. D. Willows, pp. 1–101. Plenum Press, New York.

SCHILLER, P. H. (1948). Studies in learning in the octopus. *Report of the Committee on Research for the National Academy of Sciences* 158–160.

SIEGEL, S. & CASTELLAN, N. J. JR. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.

SUBOSKI, M. D., MUIR, D. & HALL, D. (1993). Social learning in invertebrates. *Science* **259**, 1628–1629.

SUTHERLAND, N. S. (1963). The shape discrimination of stationary shapes by octopuses. *American Journal of Psychology* **76**, 177–190.

SUTHERLAND, N. S. & MUNTZ, W. R. A. (1959). Simultaneous discrimination training and preferred directions of motion in visual discrimination of shape in *Octopus vulgaris* Lamarck. *Pubblicazione della Stazione Zoologica di Napoli* **31**, 109–126.

SUTHERLAND, N. S., MACKINTOSH, N. J. & MACKINTOSH, J. (1963). Simultaneous discrimination training of *Octopus* and transfer of discrimination along a continuum. *Journal of Comparative and Physiological Psychology* **56**, 150–156.

SUTHERLAND, N. S., MACKINTOSH, N. J. & MACKINTOSH, J. (1965). Shape and size discrimination in *Octopus*: the effects of pretraining along different dimensions. *Journal of Genetic Psychology* **106**, 1–10.

THORPE, W. H. (1956). *Learning and Instinct in Animals.* Methuen, London.

VAN HEUKELEM, W. F. (1977). Laboratory maintenance, breeding, rearing and biomedical research potential of the Yucatan octopus (*Octopus maya*). *Laboratory Animal Science* **27**, 852–859.

VON UEXKULL, J. (1894). Physiologische untersuchungen an *Eledone moschata*. IV. Zur analyse der functionen des centralnervensystems. *Zeitschrift für Biologie* **31**, 584–609.

WALKER, J. J., LONGO, N. & BITTERMAN, M. E. (1970). The octopus in the laboratory: handling, maintenance, training. *Behavioral Research Methods and Instrumentation* **2**, 15–18.

WELLS, M. J. (1962). *Brain and Behaviour in Cephalopods.* Heinemann, Toronto.

WELLS, M. J. (1966*a*). Learning in the octopus. *Symposium of the Society for Experimental Biologists* **20**, 477–507.

WELLS, M. J. (1966*b*). The brain and behavior of cephalopods. In *Physiology of the Mollusca*, vol. II (ed. K. M. Wilbur and C. M. Yonge), pp. 547–590. Academic Press, New York.

WELLS, M. J. (1978). *Octopus: Physiology and Behaviour of an Advanced Invertebrate.* Chapman and Hall, London.

WELLS, M. J., O'DOR, R. K., MANGOLD, K. & WELLS, J. (1983). Diurnal changes in activity and metabolic rate in *Octopus vulgaris*. *Marine Behavior and Physiology* **9**, 275–287.

WILKINSON, L. (1990). SYSTAT: The System for Statistics. Evanston, IL: SYSTAT, Inc.

WODINSKY, J. (1971). Movement as a necessary stimulus of *Octopus* predation. *Nature* **229**, 493–494.

YOUNG, J. Z. (1961). Learning and discrimination in the octopus. *Biology Reviews* **36**, 32–96.

YOUNG, J. Z. (1962). Repeated reversal of training in *Octopus*. *Quarterly Journal of Experimental Psychology* **14**, 206–222.

YOUNG, J. Z. (1977). Brain, behavior and evolution of cephalopods. *Symposium of the Zoological Society of London* **38**, 377–434.

XI. APPENDIX

Coding of variables for the Correspondence Analysis: (5) Space: width, 1 = 30 or 40, 2 = 60, 3 = 100. (6) Start: 1 = no, 2 = yes. (7) Dividers: 1 = no, 2 = simple, 3 = complex. (8) Stimuli: 1 = no, 2 = yes, separate variables for brightness, hue, orientation, shape and size. (11) Negative reinforcement: shock, 1 = none, 2 = 0–8 progressive, 3 = 8, 4 $\geqslant$ 8. (12) Reinforcement located on stimuli: 1 = no, 2 = yes. (13) Position of stimuli in tank: 1 = low, 2 = high. (14) Support: 1 = hand, 2 = bracket. (15) Movement: 1 = placement only, 2 = rotation into tank, 3 = motorized movement, 4 = slow hand movement, 5 = hand movement at $3 \times /s$. (17) Intertrial interval: 1 = 1–5 s, 2 = 6–10 s, 3 = 11–15 s, 4 = 16–20 s, 5 = 21–30 s. (18) Trials per session: 1 = 1–3, 2 = 4–6, 3 = 7–10. (19) Sessions per day: 1 = 1, 2 = 2, 3 = 3, 4 $\geqslant$ 3. (20) Total days: 1 = 1–7, 2 = 8–14, 3 = 15–21, 4 = 22–30. (21) Pretraining: 1 = 0, 2 = 1–6, 3 = 7–12, 4 = 13–18, 5 = 19–24, 6 = 25–30. (22) Correction trials: 1 = no, 2 = S+ only, 3 = fewer than 5, 4 = 5 maximum, 5 = unlimited. (25) Number of subjects: 1 = 1, 2 = 2–4, 3 = 5–8, 4 = 9–12, 5 = 13–16, 6 $\geqslant$ 16. (26). Total trials: 1 = 0–33, 2 = 34–66, 3 = 67–100, 4 = 101–133, 5 > 133. (29) Variance: 1 = no, 2 = yes. (32) Simultaneous discrimination learning: outcome, 1 = no, 2 = weak, 3 = maybe, 4 = yes. Variables missing or outside coded ranges were given the average code for their group.

ERRATUM

Table 2e, page 172, row 30, column 30. Should say “No” instead of “Yes.”
Figure 1, page 183, lower graph. Experiment 30 should have open circle, not filled circle.