ORIGINAL RESEARCH

Wolves, Dogs, Rearing and Reinforcement: Complex Interactions Underlying Species Differences in Training and Problem-Solving Performance

Harry Frank

Abstract Frank and Frank et al. (1982–1987) administered a series of age-graded training and problem-solving tasks to samples of Eastern timber wolf (*C. lupus lycaon*) and Alaskan Malamute (*C. familiaris*) pups to test Frank's (Zeitschrift für Tierpsychologie 53:389–399, 1980) model of the evolution of information processing under conditions of natural and artificial selection. Results confirmed the model's prediction that wolves should perform better than dogs on problem-solving tasks and that dogs should perform better than wolves on training tasks. Further data collected at the University of Connecticut in 1983 revealed a more complex and refined picture, indicating that species differences can be mediated by a number of factors influencing wolf performance, including socialization regimen (hand-rearing vs. mother-rearing), interactive effects of socialization on the efficacy of both rewards and punishments, and the flexibility to select learning strategies that experimenters might not anticipate.

Keywords Wolves · Dogs · Socialization · Cognition · Problem-solving · Training · Reinforcement

H. Frank (✉)
Department of Psychology, The University of Michigan-Flint,
4488 Hillside Drive, Ann Arbor, MI 48105, USA
e-mail: hfrank@umich.edu

Introduction

Frank's (1980) theoretical model of canine information processing argued that natural selection has favored the evolution of a duplex, or bicameral, system of information processing in the gray wolf (*C. lupus*). The more primitive "instinctual" component is a repertoire of relatively closed behavioral programs. Whether the behaviors regulated by this subsystem are genetically preprogrammed or are locked in during early development by innate teaching mechanisms, they (a) are elicited by very specific stimulus configurations and (b) exhibit little plasticity. The more recently acquired "cognitive" subsystem, which may have evolved in tandem with the rise of cooperative group hunting, is characterized by such complex capacities as insight into means-ends relationships, imagery, cognitive mapping, foresight, and serial organization of behavior.

In contrast, the model suggested that the evolutionary environment of domestic dogs (*C. familiaris*) was a world increasingly shaped to meet human needs, capacities, and morphology and, concomitantly, human mediation of dog-environment transactions. These changes in the evolutionary landscape both relaxed the adaptational pressures that favored cognitive complexity in wolves (e.g., a mental map of the actions that open a gate has no advantage if the manipulanda are at human height or require an opposable thumb) and introduced pressures favoring selection for *tractability*, responsiveness to a broad band of stimuli (e.g., verbal cues) and sufficient behavioral plasticity to permit shaping and reinforcement.

Based on this evolutionary scenario, the author predicted that domestic dogs should perform better than wolves on *training* tasks, in which (a) cues are arbitrarily selected by the experimenter, (b) reinforcement is

administered by the experimenter, and (c) the to-be-learned behavior has no perceptible, functional connection with the outcome (i.e., the reinforcement). Conversely, it was predicted that wolves should perform better on *problem-solving* tasks, in which both the cues and the reinforcement are perceptibly intrinsic to the activity.

A series of age-graded learning experiments (summarized in Frank and Frank 1987) conducted on samples of four Eastern timber wolf (*C. l. Lycaon*) pups and four Alaskan Malamute pups, both foster-reared by the same female wolf under identical conditions, supported the hypotheses: As predicted, Malamute pups performed better in tests of inhibition, leash-training, and cue-discrimination in a T-maze. Wolf pups performed better on a barrier test (a version of Köhler's 1925/1927 detour test adapted for dogs by Scott and Fuller 1965), a complex puzzle-box test, and a 6-unit T-maze. However, a number of factors suggested that some of the training-task experiments should be replicated:

(1) Although the wolf and dog pups were reared under virtually identical conditions, the domestic pups were highly socialized to humans, but the wolf pups were essentially unsocialized to humans.

(2) Two of the training-task experiments used a choke chain as an aversive stimulus. This is a customary device for training domestic dogs, but Frank and Frank (1983) noted that several wild species have been observed to respond in largely programmed fashion to neck restraint (e.g., the "freeze" response reported in some small mammals or persistent panic) and suggested—more generally—that tolerance for restraint in wild animals is maladaptive. It was therefore considered possible that the original experimental procedure may in these instances have run afoul of species-specific constraints on learning (see Hinde and Stevenson-Hinde 1973; Seligman and Hager 1972) or otherwise exploited innate action patterns that biased the results in favor of the authors' predictions.

(3) Contrary to the model's prediction, wolves performed better than dogs on a discrimination learning test classified by the experimenters as a training task, though the difference was not statistically significant.

The replication studies were conducted in 1983 at the University of Connecticut's Biobehavioral Sciences Program under the sponsorship of Benson Ginsburg. *Ss* were seven wolf pups donated by the Ross Park Zoo (Binghamton, NY) at 6 days of age, hand-reared, and wholly socialized to human contact and companionship.

Inhibition training

Summary of original experiments (Frank and Frank 1983)

Experiment 1: inhibition test

This was a passive inhibition task requiring *Ss* to remain for a prescribed length of time on a plywood platform (100 × 100 × 10 cm) situated in one corner of the enclosed plywood arena (2.44 m high) shown in Fig. 1.

Method

Apparatus for this experiment comprised a choke chain tied to a length of 1/8-in (0.32 mm) nylon cord, which ran through an eyebolt 45 cm above the platform to a pulley 1.8 m above the platform and then through a series of eyebolts to an observation window in the wall of the arena.

Testing began on the Monday nearest the pups' 7-week birthday and continued for 10 weekdays following the procedures described by Scott et al. (1967). On Day 1, *S* received a preliminary trial in which it was held in position by the rope and choke chain for 5 s. On Day 2, *Ss* received a preliminary trial and two training trials in which each *S* was placed on the platform by one experimenter and corrected (punished) by a sharp tug on the choke chain if it
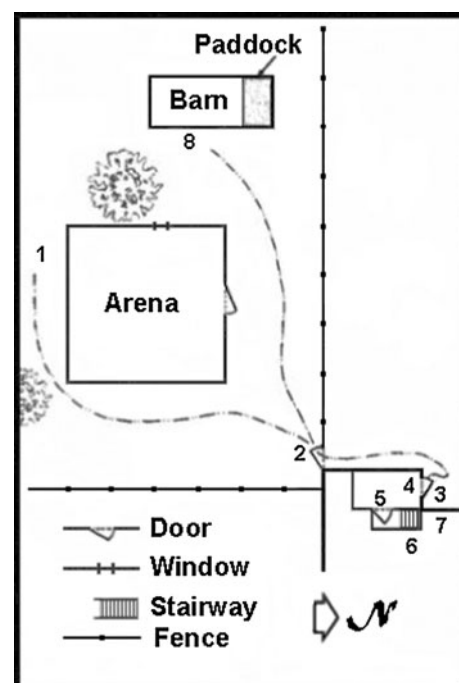

Fig. 1 Leash training course

Table 1 Total intratrial best-performance times (min)

	Wolves ($N = 4$)	Malamutes ($N = 4$)
$\bar{X}$	4.13	12.33
$\hat{\sigma}^2$	5.31	16.37

stepped off the platform before 15 s had elapsed. Punishment was administered from outside the arena by the second experimenter posted at the tinted observation window in the west wall of the arena. From Day 3 to 10, Ss were each administered a preliminary trial and five training trials. Time criteria for the training trials increased daily (depending on S's performance the preceding day) from 15 to 120 s, and Ss were allowed up to five corrections per trial. S's score for each trial was its best performance, the longest time spent on the platform without or prior to correction.

Results

Sample means and variances of best performance totals appear in Table 1. The difference of means was significant ($t = 3.52$, $df = 6$, $p \cong 0.006$), with Malamutes, as predicted, averaging longer times on the platform.

Experiment 2: leash training

This was an active inhibition task, requiring Ss to extinguish such unacceptable responses as jumping, biting the leash, and tugging and to substitute an alternative behavioral pattern, e.g., walking at the trainer's side with no tension on the leash. The course (See Fig. 1) was approximately 150 m long and incorporated as nearly as possible the gates, doors, stairs, and similar physical features described by Scott et al. (1950).

Method

Leash training began on the Monday nearest the pups' 11-week birthday and proceeded for 10 consecutive weekdays. On Days 1 and 2, each pup was carried from the paddock to the starting point of the course (point 1 in Fig. 1). From there, S was led (using a standard leather training leash and choke chain) to the enclosure gate (point 2) and then carried into the experimenters' kitchen (point 7). The pup was fed a teaspoon of sardines, carried back into the enclosure (point 2) and led back to the home barn (point 8), where it was given an additional serving of fish. Performance was not scored on the return trip. Beginning on Day 3, each S was led from the starting point (1) through the gate (2), two doors (3 and 5), up a short flight of stairs (6) into the kitchen and given administered food reward. The pup was

then led back to point 8, where (on Days 3–5) it was administered a second serving of fish.

Scores were demerits assigned for a number of faults, including balking (in open or doorways), fighting the leash, position errors (e.g., dragging behind or tugging ahead), interference with the handler, and vocalization. Vocalization demerits were ultimately dropped from the analysis, because no vocalizations were recorded for any wolf pup.

Results

The experiment focused on improvement over the 10-day training period. Means scores for the two groups on Days 1 and 10 and mean improvement (Day 1–10, Ss as own controls) are shown in Table 2 below.

As predicted, the Malamutes averaged fewer demerits on Day 10 ($t = 4.37$, $df = 6$, $p \cong 0.002$) and showed greater improvement ($t = -3.23$, $df = 6$, $p \cong 0.009$) than the wolves. No difference was predicted for Day 1 performance, and the difference between wolf and Malamute Day 1 means was not significant.

Replication study

Experiment 1: inhibition test

Method

Apparatus The outdoor testing arena used in the original study was reproduced inside the project's testing facility[1] and incorporated one exterior and one interior wall of the building.

Procedure Fours Ss were administered the same protocol followed in the original study. Three Ss were punished by felt pellets shot from a match-quality air rifle by a sniper located outside the building and shooting through a 5×15 cm slot in a plywood panel, with which we had replaced one windowpane.[2]

---

[1] The testing facility was a rehabilitated chicken coop located in Mansfield Depot on the site of an experimental chicken farm operated by the University of Connecticut in the 1950s.

[2] Various alternative punishment media were considered. Similar studies have used electric grids surrounding the platform. However, our experience suggested that wolf pups would very quickly learn to jump beyond any grid that was small enough in area to be economically feasible. By chance, one the project's volunteer assistants (Dawn Littleton) was a nationally ranked air-rifle competitor and suggested we use felt pellets, which are ordinary shot through an air rifle to clean the bore of lead fouling. A test of the pellets produced only a mild, though startling, sting when the pellets struck an unprotected hand at point-blank range. It is worthy of note that in several hundred shots, Ms. Littleton, shooting from a distance of

Table 2 Leash-training demerits (excluding vocalization) for wolves and malamutes

	Wolves ($N = 4$)			Malamutes ($N = 4$)		
	Day 1	Day 10	Improvement Day 1–10	Day 1	Day 10	Improvement Day 1–10
$\bar{X}$	7.25	6.5	.75	6	1	5
$\hat{\sigma}^2$	1.5833	5.6667	4.25	0.6667	0.6667	2.6667

Table 3 Total intratrial best-performance times (min)

	Wolves			Malamutes
Rearing	Hand		Mother	Mother
Reinforcer	Choke chain ($N = 4$)	Pellets ($N = 3$)	Choke chain ($N = 4$)	Choke chain ($N = 3$)
$\bar{X}$	6.12	9.08	4.13	12.33
$\hat{\sigma}^2$	6.07	4.22	5.31	16.37

Table 4 p-values for a priori comparisons of inhibition test performance

	Species	Reinforcer	Socialization
Pooled	$\cong 0.003$	>0.20	$\cong 0.006$
Within			
Species (wolves)	–	$\cong 0.03$	$\cong 0.04$
Reinforcer			>0.15
Socialization		>0.10	–
Reinforcer (choke chain)		–	
Species			
Socialization	$\cong 0.006$		–
Interaction			
Socialization × Reinforcer	N/A		$\cong 0.02$
Species × Reinforcer		>0.05	N/A

Results and discussion

Means and sample variances for all four groups are shown in Table 3. Group effects were significant in a simple one-way analysis of variance [$F(3,11) = 6.078$, $p \cong 0.01$], but since the data included only one group of Malamutes (mother-reared and reinforced by choke chain) and only hand-reared wolves were subjected to pellet punishment, it was not possible to tease apart the effects of species, reinforcement, and rearing by analysis of variance. Furthermore, "rearing" and "socialization" are not synonymous across species. The Malamutes were mother-reared but fully socialized to humans. Instead, the effects of species, reinforcer, and socialization were investigated by a series of a priori planned comparisons, results of which are summarized in Table 4.

Each comparison represented in Table 4 is the p-value for a linear combination of the form

$$c_1\bar{x}_1 + c_2\bar{x}_2 + \cdots + c_j\bar{x}_j$$

where ($c_j = 0$. The Pooled Species cell, for example, represents the combination

$$0.3333\bar{x}_1 + 0.3333\bar{x}_2 + 0.3333\bar{x}_3 - \bar{x}_4$$

and the entry is the p-value for a Student's transformation testing the hypothesis

$$\frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4 = 0$$

i.e., that the average of means for the three wolf groups equals the Malamute mean.

The most basic conclusion that we can draw from the results in Table 4 is that species trumps the other variables (reinforcer and socialization), thus supporting the findings of the original study. The pooled species comparison (wolves vs. Malamutes) is significant ($p \cong 0.003$), and even when the potentially confounding effects of socialization and reinforcer differences are eliminated, the species comparison (socialized wolves × choke chain vs. socialized Malamutes × choke chain) remains statistically significant ($p \cong 0.006$). Differences between wolf and Malamute performance drop to a nonsignificant level ($p \cong 0.085$) only in species × reinforcement interaction (hand-reared wolves × pellets vs. Malamutes × choke chain).

The significance probability ($p \cong 0.21$) for the pooled reinforcer combination (hand-reared wolves × choke chain + mother-reared wolves by choke chain + Malamutes × choke chain vs. hand-reared wolves × pellets) is difficult to interpret because of both the socialization × reinforcement interaction ($p \cong 0.023$) and confounding: All Malamutes and all unsocialized wolves were in the choke chain group. When species effects are

eliminated (hand-reared wolves × choke chain + mother-reared wolves × choke chain vs. hand-reared wolves × pellets), we obtain a significant ($p \cong 0.034$) combination. The explanation is evident from an examination of the group means in Table 3: In the pooled comparison, the choke chain average was sufficiently elevated by the Malamute scores to mask the difference between choke chain-punished wolves and pellet-punished wolves. If socialization effects are also partialed out of the mix (hand-reared wolves × choke chain vs. hand-reared wolves by pellets), the significance disappears ($p \cong 0.101$).

The pooled socialization combination (hand-reared wolves and Malamutes vs. mother-reared wolves) is also ambiguous. The unsocialized group included only choke chain-punished wolves, and all of the Malamutes were in the socialized group, so the significance ($p \cong 0.0062$) might be attributable to socialization, to species, or to the interaction of socialization and reinforcement. When the Malamute mean is dropped from the combination, the comparison of socialized and unsocialized Ss remains significant ($p \cong 0.041$), but when reinforcer differences are also eliminated (mother-reared wolves × choke chain wolves vs. hand-reared wolves × choke chain) the significance disappears ($p \cong 0.175$).

The analyses of the reinforcement and socialization comparisons would therefore suggest that these factors exert an effect only in concert, which is consistent with the significance ($p \cong 0.023$) of the socialization × reinforcement interaction comparison (Mother-reared wolves × pellets vs. hand-reared wolves × choke chain).

Experiment 2: leash training

Method

The leash-training course used for replication was conducted over a route that included the testing and food preparation areas of the test facility and the adjacent outdoor enclosure. The course was approximately the same length as in the original study and included the same number and type of obstacles (doors, stairs, etc.). The only procedural departure was that three of the Ss were restrained by a body harness instead of a choke chain.

Results and discussion

Mean demerits (excluding vocalization demerits) for all four groups are shown in Fig. 2. The original study focused on 10-day improvement, following Scott and Fuller (1965). However, since many of the wolf pups in both the original and replication study reached peak performance before Day 10, mean demerits for "best day" is also shown.
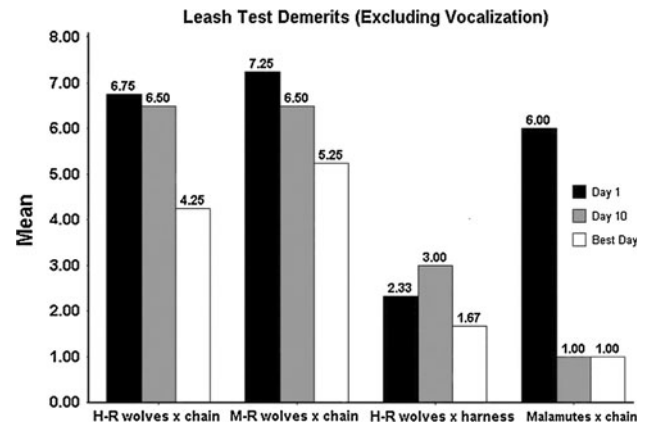

Fig. 2 Mean leash-test demerits for four groups of Ss

Changes in leash-test performance Leash-test data were submitted to the same analyses as the inhibition test data. In order to capture every subject's maximum change, I examined best-day changes (Day 1 demerits—Best-day demerits) as well as 10-day changes (Day-1 demerits—Day 10 demerits). These scores are characterized as "changes" rather than "improvements," because one group (See Fig. 2) averaged higher demerits on Day 10 than on Day 1. One-way ANOVA for differences among the four groups yielded significant F-ratios for both changes [$F(3,11) = 8.699$, $p \cong 0.003$ and $F(3,11) = 5.850$, $p \cong 0.012$, respectively]. Significance levels for the planned comparisons among groups are shown in Table 5. In each cell, the boldface entry is the p-value for 10-day change, and the italicized entry in parentheses is the p-value for best-day change.

As in the inhibition test, the most significant factor in the leash test is species. Whether we consider 10-day change or best-day change, species differences are significant whether pooled ($p \cong 0.00019$ for 10-day change; $p \cong 0.001$ for best-day change) or controlled for socialization and restraint ($p \cong 0.0009$ for 10-day change; $p \cong 0.01$ for best-day change). Although comparisons between Malamutes and wolves yielded highly significant differences ($p \cong 0.0005$ for 10-day change; $p \cong 0.0012$ for best-day change) even under conditions predicted to maximize wolf performance (hand-reared × harness-trained), this is interpreted as a scoring artifact, rather than further evidence of unassailable species differences. Day 1 performance for wolf pups in this group simply left very little room for improvement (see Fig. 2).

The pooled restraint comparisons (all choke chain-restrained Ss vs. hand-reared wolves × harness) are significant ($p \cong 0.015$ for 10-day change; $p \cong 0.013$ for best-day change), but the differences shrink to nonsignificant levels when species differences are eliminated (choke chain-restrained wolves vs. harness-restrained wolves).

Table 5 *p*-values for a priori comparisons of 10-day changes and best-day changes in leash-training demerits

	Species	Restraint	Socialization
Pooled	≅**0.0002** (*≅0.001*)	≅**0.01** (*≅0.01*)	>**0.20** (*>0.15*)
Within			
Species (wolves)	–	>**0.15** (*>0.05*)	>**0.15** (*>0.30*)
Restraint		–	>**0.30** (*>0.30*)
Socialization		>**0.20** (*>0.05*)	–
Restraint (choke chain)		–	
Species	–		
Socialization	≅**0.001** (*≅0.01*)		–
Interaction			
Socialization × Restraint	N/A		>**0.10** (*>0.15*)
Species × Restraint		≅**0.0005** (*≅0.001*)	**N/A**

Boldface entries are *p*-values for 10-day change; (*italicized*) entries are *p*-values for best-day change

Table 6 *p*-values for a priori comparisons of best-day leash performance

	Species	Restraint	Socialization
Pooled	≅0.002	>0.05	≅0.003
Within			
Species (wolves)	–	≅0.007	≅0.02
Restraint		–	>0.15
Socialization		≅*0.03*	–
Restraint (choke chain)		–	
Species	–		
Socialization	≅0.006		–
Interaction			
Socialization × Restraint	N/A		≅0.02
Species × Restraint		>0.15	N/A

Since there is no interaction between restraint and socialization ($p > 0.10$), it is not surprising that the differences remain at chance levels when socialization is factored out (hand-reared wolves × choke-chain vs. hand-reared wolves × harness).

Socialization differences are nonsignificant at any level of control-pooled, within species, or within species and restraint. In this regard, it might be noted in the choke-chain condition, Day 10 means for hand-reared and mother-reared wolves were actually identical ($\bar{X} = 6.5$).

Best-day performance Results of the leash training replication support Frank and Frank's (1983) conclusion that differences between wolf and dog improvement scores were attributable to species differences. This being said, it must be recalled that the comparisons in Table 5 are based on performance *changes*, which can obscure differences in *absolute* performance. As noted above, for example, Day 1 performance of the harness-trained wolf pups allowed little opportunity for improvement. Furthermore, changes in performance are not strictly comparable to the "best

performances" considered in the inhibition test analyses. Accordingly, a one-way ANOVA for group differences [$F(3,11) = 7.67$, $p \cong 0.0048$] and the same planned comparisons were performed on best-day scores.

The pattern of *p*-values summarized in Table 6 is virtually identical to the pattern of inhibition-test *p*-values in Table 4. The one exception (shown in italics) is the significant ($p \cong 0.027$) difference between leash-trained and harness-trained socialized wolves. Although choke chain reinforcement was used in both inhibition training and leash training, it was applied very differently in the two experiments. In the inhibition test, the choke chain was applied with a single sharp tug. In leash training, faults such as dragging, tugging, and balking created a continuous restraint and might therefore be more likely than an abrupt tug, followed by release, to trigger the sort of species-specific reaction to neck restraint that we considered a potential confounding factor in the original experiment. That a "misbehavior of animals" artifact may have contributed to differences between wolves and Malamutes in the original study is supported by the nonsignificant species × restraint interaction ($p \cong 0.176$) that compared best-day performance of Malamutes with best-day performance of socialized, harness-trained wolves.

Conclusions: inhibition learning

These results sustain the basic conclusion drawn by Frank and Frank (1983) that differences between wolves and dogs in both inhibition test performance and change in leash-training demerits were attributable to species differences—socialization differences and possible species-specific responses to choke-chain reinforcement notwithstanding. Accordingly, the replication further sustains the authors' (1983) broader conclusions that results of these experiments support the prediction that dogs should perform better than wolves on training tasks and, by extension, the

theoretical model of wolf and dog evolution from which the hypotheses derived.

Nevertheless, socialization and reinforcement were not without effect. In comparisons of both inhibition-test performance and best-day leash-test performance, there was significant within-wolves interaction between socialization and reinforcer, and in both experiments the combination of wolf socialization and use of some reinforcer other than choke chain reduced differences between Malamute and wolf performance to nonsignificant levels.

Motivation and insight in visual discrimination learning

The experiment was an oddity-learning task using a three-position Wisconsin General Test Apparatus (WGTA). A comparison of the original and replication studies is fully presented in Frank et al. (1989) and is summarized here to document more fully the contribution of Benson Ginsburg and the University of Connecticut to the author's research program and to correct a statistical error (see footnote 5) that favored the null hypothesis in the 1989 publication.
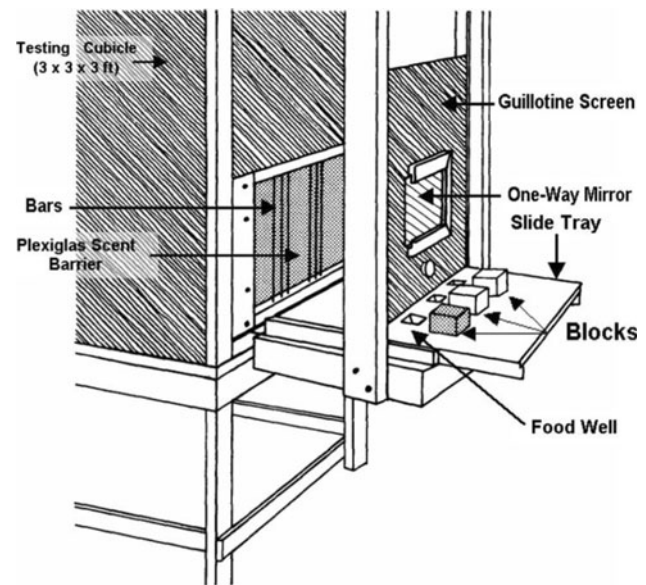
Method

The WGTA (Fig. 3) consisted of a plywood cubicle with a barred window and a shelf supporting a guillotine-type screen and a sliding tray with recessed food wells. On each trial, the experimenter put food in one of the three wells, covered it with either a white block or a black block, covered the empty wells with blocks of the opposite color,[3] then lifted the screen far enough to clear the blocks, slid the tray forward and lowered the screen. The bars on the window were spaced to create three openings (one centered directly in front of each food well) that could be adjusted according to the size of the S's head.

Testing began on the Monday nearest the Ss' 15-week birthday with 4 days of habituation and shaping, in which pups learned to reach through the bars and displace a solitary white or black block to get food from the well underneath.

Fifteen oddity discrimination trials were administered on Day 5 and proceeded for 2 weeks or until pups reached a criterion of 85% correct responses over two consecutive days (26 correct choices in 30 trials). In this phase of the experiment, the odd-colored block was the color to which the pup was shaped during habituation.

Fig. 3 Wisconsin General Test Apparatus

On the day following each pup's criterion performance, the color scheme was reversed. For pups that had been rewarded initially for displacing the white block, food was placed under a black block and the two empty wells covered with white blocks; pups that had learned to displace the black block were in this phase presented with a white block and two black blocks.

Results and discussion

Socialization and motivation Means and sample variances for Ss in both the original and replication study appear in Table 7.

In the initial study (mother-reared wolves and Malamutes) the only significant difference was in reversal learning, favoring the Malamutes ($t = 2.5$, $df = 6$, $p < 0.025$), which was consistent with our hypotheses and therefore occasioned little attention. It is seen in Table 7, however, that—contrary to prediction—wolf pups performed better than Malamute pups in discrimination

---

[3] In the original study, food was smeared into a groove in the underside of the unrewarded pair of blocks to prevent olfactory location. In the replication study, this precaution was refined: The grooves were replaced by countersunk 35-mm film containers with wire mesh lids, and the scent barrier shown in Fig. 3 was added to the WGTA.

Table 7 Trials to criterion in visual discrimination learning for wolves and alaskan malamutes

	Hand-reared wolves ($N = 7$)	Mother-reared wolves ($N = 4$)	Malamutes ($N = 4$)
Discrimination			
$\bar{X}$	55.71	82.5	97.5
$\hat{\sigma}^2$	278.57	2275.0	225.0
Reversal			
$\bar{X}$	117.86	191.25	150
$\hat{\sigma}^2$	1082.14	806.25	300

learning. The difference was not statistically significant, but was nevertheless worrisome because of the magnitude of the difference: Malamutes required almost 20% more trials (or, one Malamute standard deviation) than wolves to reach criterion, and it appeared that the difference may have failed to reach statistical significance only because of the extraordinarily high variance in wolf performance. This potential anomaly was noted by M. Kiley-Worthington (personal communication, August, 1982), who suggested that our significant result might be confounded by socialization differences. Her suggestion was supported in replication by the performance of the hand-reared wolves, which was significantly better than Malamute performance in both discrimination learning ($p < 0.005$) and reversal learning ($p < 0.05$).

Hand-reared wolves also performed better than mother-reared wolves. Mother-reared wolves required almost half again as many trials to reach criterion in discrimination learning, but the high variability of the mother-reared wolves precluded statistical significance. The difference was, however, significant ($p < 0.005$) in reversal learning. In this regard, Frank et al. (1989) noted that the mother-reared wolves were largely indifferent to food rewards (ranging from gourmet quality smoked oysters to chocolate fudge cake frosting) throughout the original series of experiments, whereas the hand-reared wolves and the Malamutes were highly motivated by food reward, the former often demonstrating extraordinary task persistence to earn even token reinforcement (e.g., a single nugget of dried kibble). In addition, the performance of mother-reared wolves showed relapses, exceeding chance (or even criterion) on 1 day and falling to random chance levels the next day. The authors therefore ascribed the inferior performance of the mother-reared pups to motivational variables, rather than task variables, and suggested various theoretical linkages between socialization and the incentive value of food reward.[4]

Insight and trial-and-error learning in WGTA performance
Motivational factors being equal (i.e., discounting the

performance of the mother-reared wolves), the superior wolf performance would seem to contradict the basic prediction that dogs should perform better than wolves on training tasks. During the course of the replication study, however, Victor Dennenberg (then a faculty member in the University of Connecticut's Biobehavioral Sciences program) suggested that I might have misclassified WGTA oddity learning as a training task. He pointed out that (1) even though there was no perceptible, *functional* connection between the cues and the to-be-learned behavior, there was a visible *spacio-temporal* connection and (2) even though the cues were arbitrarily determined by the experimenter, reinforcements were embedded in successful task performance, rather than administered by an external agent. He therefore proposed that this particular task might be amenable either to the sort of "insight" learning tapped by the problem-solving tasks or to the sort of "trial-and-error" learning tapped by the other training tasks. This line of reasoning implies two critical hypotheses:

(1) Wolves approach the WGTA task with insight, and dogs approach the WGTA task by trial-and-error.
(2) Insight produces more rapid acquisition than trial-and-error learning

Koffka (1925) pointed out that the hallmark of insight is a sudden, discontinuous increase in the frequency of correct responses (a "sharp descent" in errors; p 164), as contrasted with the gradual, incremental increases that characterize associative, trial-and-error learning. Since Ss learned at different rates, the most meaningful zero-point for comparison seemed to be the first day each S's score exceeded random chance. For $N = 15$ trials, P(success) $= 0.333$, and $\alpha = 0.05$, the critical value is 8 correct responses[5] ($p \cong 0.03$).

Figure 4 plots the mean number of correct responses over a 4-day period, from 2 days before the first day S exceeded chance performance (Day C) to 1 day after. In Fig. 4, it would appear indeed that Malamute performance improved in accordance with models of incremental learning, and wolf performance demonstrated the sort of jump from chance performance to greater-than-chance performance characteristic of sudden insight.

To test the hypothesis that wolf performance jumped more precipitously to greater-than-chance levels than Malamute performance, the percentage increase from Day C-1 to Day C was calculated for every subject and the percentages submitted to a Mann–Whitney test. Results ($p \cong 0.045$) support the hypothesis. This analysis

---

[4] Coppinger and Coppinger (2001) propose that selection for the ability to eat in proximity to people (and the consequent capacity to exploit a new niche, the village dump) produced a distinct breeding population of wolves from which domestic dogs evolved. This scenario suggests explanations for the unsocialized wolves' indifference to food not considered by Frank, Frank, and Hasselbach (1989). For several weeks after weaning, the mother-reared pups were fed individually with the experimenters in attendance, both to ensure equal rationing of medication (incorporated into their food) and to increase daily contact with humans. Rather than habituating them to human presence, it is possible that the stress associated with forced human proximity may have established a conditioned inappetence. In the experimental setting, the mere presence of humans may therefore have averted the unsocialized pups from food reward.

[5] The comparisons reported in Frank, Frank, and Hasselbach (1989) were based on a greater-than-chance threshold of 9 correct responses, the critical value for $\alpha = 0.01$ (p $\cong$ 0.0085), which the first author mistakenly believed to be the critical value for $\alpha = 0.05$.
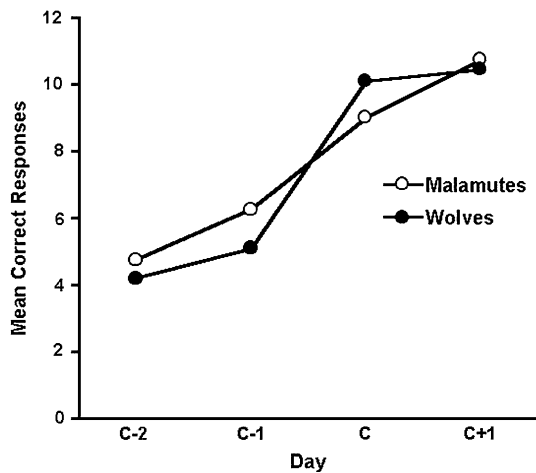
Fig. 4 Mean correct responses for wolves and Alaskan Malamutes 2 days preceding and 1 day following first day subject had at least 8 correct responses (Day C) in reversal learning

Table 8 Trials to reversal criterion for insightful and noninsightful subjects

	Insightful		Noninsightful	
	Wolves	Malamutes	Wolves	Malamutes
N	4	1	3	3
$\bar{X}$	108.0		147.5	
$\hat{\sigma}^2$	1,057.5		307.5	

examined only reversal learning, because discrimination learning was shaped, and Ss in both groups exceeded random chance on the first day of trials.

To test the hypothesis that insight yields more rapid acquisition than trial and error learning, mean trials to criterion were compared for "insightful" and "noninsightful" Ss. For purposes of this analysis, reversal leaning performance was defined as insightful if the subject's score on Day C was at least double its score on Day C-1. Mean trials to reversal (See Table 8) was significantly greater for noninsightful Ss than for insightful Ss ($t = 2.48$; $df = 9$; $p \cong 0.015$).

Conclusions: motivation and insight

The replication of the oddity learning and reversal learning experiment using hand-reared wolf pups illuminated results of the original study in several ways:

(1) As Dennenberg suggested, the task—most especially in the reversal phase—includes characteristics of both problem-solving and training, as defined earlier, and is therefore susceptible both to insight and to trial-and-error learning.

(2) Insight solutions achieve more rapid acquisition than trial-and-error solutions.

(3) Given a task that can engage either insight or trial-and-error, wolves are more inclined than Malamutes to adopt the insight strategy and therefore require fewer trials to reach criterion.

(4) Failure of wolves in the original study to perform better than the Malamutes was a motivational artifact rising from their socialization regimen.

The inadvertent inclusion of a test that could be approached by Ss as either a problem-solving exercise or a training exercise offered a unique opportunity to capture in a single task essential species differences implied by the (1980) theoretical model: that evolution favored (a) complex problem-solving capacities in the wolf and (b) trainability in the dog.

General conclusions

Replication research is often the neglected stepchild of science. Scientists become easily bored and prefer exploring new frontiers to plowing old fields. Furthermore, replication work is not professionally cost-effective: If the original research is supported, replication adds little to our knowledge and may not reach the threshold of originality demanded for publication. And, if one is replicating his own work, there is always the risk that the original results may be contradicted and discredited. The series of studies reported in the present paper approach the "sweet spot" of replication: The results support the conclusions drawn from the original research but offer sufficient refinement of those conclusions to yield something fresh—and to remind us that the natural world is wonderfully more complex than the experimental hypotheses we test to illuminate that world.

References

Coppinger R, Coppinger L (2001) Dogs. Scribner, New York

Frank H (1980) Evolution of canine information processing under conditions of natural and artificial selection. Zeitschrift für Tierpsychologie 53:389–399

Frank H, Frank MG (1982) Comparison of problem-solving performance in six-week-old wolves and dogs. Anim Behav 30:95–98

Frank H, Frank MG (1983) Inhibition training in wolves and dogs. Behav Process 8:363–377

Frank H, Frank MG (1987) The University of Michigan canine information-processing project (1979–1981). In: Frank H (ed) Man and wolf: Advances, issues and problems in captive wolf research. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 143–167

Frank H, Frank MG, Hasselbach LM (1989) Motivation and insight in wolf (*Canis lupus*) and Alaskan malamute (*Canis familiaris*): visual discrimination learning. Bull Psychonomic Soc 27(5): 455–458

Hinde RA, Stevenson-Hinde J (eds) (1973) Constraints on learning. Academic Press, New York

Koffka K (1925) The growth of mind: an introduction to child-psychology, 2nd edn. (R.M. Ogden, trans.). Harcourt Brace, New York

Köhler W (1927) The mentality of apes, 2nd edn. (E. Winter, trans.). Routledge and Kegan-Paul Ltd, London (Original work published 1925)

Scott JP, Fuller JL (1965) Genetics and the social behavior of the dog. University of Chicago Press, Chicago

Scott JP et al (1950) Manual of dog testing techniques. Roscoe B. Jackson Memorial Library (mimeoed), Bar Harbor, ME

Scott JP, Shepard JH, Werboff J (1967) Inhibitory training in dogs: effects of age at training in basenjis and Shetland sheepdogs. J Psychol 66:237–252

Seligman JEP, Hager JL (eds) (1972) The biological boundaries of learning. Appleton, Century, Crofts, New York