PSYCHOLOGICAL SCIENCE

Research Report

SHEPARD'S UNIVERSAL LAW SUPPORTED BY HONEYBEES IN SPATIAL GENERALIZATION

Ken Cheng

Macquarie University, Sydney, Australia

Abstract—*An animal that is rewarded for a response in one situation (the S+) is likely to respond to similar but recognizably different stimuli, the ubiquitous phenomenon of stimulus generalization. On the basis of functional analyses of the probabilistic structure of the world, Shepard formulated a universal law of generalization, claiming that generalization gradients, as a function of the appropriately scaled distance of a stimulus from S+, should be exponential in shape. This law was tested in spatial generalization in honeybees. Based on theoretically derived scales, generalization along both the dimensions of the distance from a landmark and the direction to a landmark followed Shepard's law. Support in an invertebrate animal increases the scope of the law, and suggests that the ecological structure of the world may have driven the evolution of cognitive structures in diverse animals.*

When doing one thing in one stimulus situation leads to a reward, an animal is likely to do the same thing in similar but discriminably different stimulus situations. This is the ubiquitous phenomenon of stimulus generalization. Although different mechanistic explanations of generalization have been proposed (Cheng, Spetch, & Johnston, 1997; Gluck, 1991; Reid & Staddon, 1998; Shepard, 1958a, 1958b), this report is concerned with functional arguments. Shepard (1987) analyzed the structure of the world regarding generalization and formulated a universal law. His argument invoked evolutionary grounds: Organisms should have been driven in the course of evolution to generalize in a way that reflects the structure of the world. Hence, the law should apply to any organism that generalizes. It is the functional, and not the mechanistic, arguments that make it reasonable to claim universality for Shepard's law. So far, however, the law has been confirmed only in humans and pigeons. I tested the generality of the law in a very different animal, the invertebrate honeybee.

SHEPARD'S LAW

I explicate Shepard's law using the example of spatial location, the manipulated variable in the present experiments; Shepard's arguments, however, apply to all dimensions of stimuli. Suppose that an animal finds food in a container at one location (S+). When the animal returns, the container is at a noticeably different location. Will the animal still "bet" on finding food in the container? Underlying this question is the assumption that the animal can discriminate the two locations. Shepard's law does not apply when the animal has trouble discriminating between stimuli (Ennis, 1988; Nosofsky, 1986, 1988; Shepard, 1986, 1988). The question is whether the two locations have the same consequence of concern—in this case, whether the container contains food. In Shepard's (1987) formulation, some region around S+ is the consequential region, and the problem of generalization is estimating the probability that an encountered location is in the consequential region, given that S+ is in the consequential region. Shepard showed that the generalization gradient remains nearly exponential under a wide range of assumptions about the consequential region. In short, the structure of the world dictates that animals "should" follow Shepard's law in generalization.

The law states that $y = e^{-kx}$. In this formulation, y is the probability or rate of responding, relative to responding at S+, and k is a scaling parameter for the exponential function. The value of k depends on how steeply the generalization gradient drops off, and also on what units are used to label the x-axis. In testing Shepard's law, determining how the x-axis should be scaled is crucial.

Two methods may be used to calculate the appropriate scale for the x-axis. One is multidimensional scaling (Shepard, 1965). In the one-dimensional case, many overlapping generalization gradients with different S+ locations are obtained. The x-axis is then adjusted subject to monotonicity to make the gradients as similar as possible when the S+s for all gradients are lined up. The method is nonarbitrary because at each point along the x-axis are found (typically) the y values for several gradients. Thus, adjusting one point on the x-axis affects several functions. Shepard (1987) used multidimensional scaling to scale 12 different data sets, all of which supported the exponential law.

A second method is to use a theoretically specified scale (Cheng, 1999). In the present study, I adopted this method because the data obtained were insufficient for multidimensional scaling. Aside from being sometimes the only method possible, this method has another advantage: Because the scale for the x-axis is specified completely by theory, no scale adjustments (except the experimental slope parameter, k) are made to fit data.

The data Shepard (1987) presented were all on vertebrate animals. In the present study, I tested the generality of Shepard's law on an invertebrate animal. Honeybees make excellent subjects because they learn many things (Bitterman, 1996; Capaldi, Robinson, & Fahrbach, 1999; Cheng, 2000; Collett & Zeil, 1998). Free-flying worker bees can be readily trained to find a location defined by landmarks within a laboratory (Cheng, 2000). The animals are first trained to find reward in a container at a constant location defined by nearby landmarks. Occasional tests then present a single container, without reward, at various locations. These tests should replicate the training situation as much as possible, except for the change in location. An animal's likelihood of searching the container gives a measure of how much it "bets" on the location. In previous work (Cheng, 1999), I obtained three spatial generalization gradients from bees using such methodology. The x-axis was then scaled according to extant theory on how the honeybee uses a landmark to localize (Cartwright & Collett, 1982, 1983). The discrepancy of a location from S+ was based on a weighted combination of discrepancies in the direction to the

---

Address correspondence to Ken Cheng, Department of Psychology, Macquarie University, Sydney NSW 2109, Australia; e-mail: kcheng@axon.bhs.mq.edu.au.

landmark and the distance from the landmark. Shepard's law was confirmed in all cases.

The experiments here extended this earlier work by testing generalization of distance and direction to a landmark separately. Also, in the earlier study (Cheng, 1999), I based the calculation of distance discrepancy solely on the retinal size projected by the landmark. Honeybees, however, use both projected retinal size (Cartwright & Collett, 1983) and motion parallax (Lehrer, Srinivasan, Zhang, & Horridge, 1988) as means of judging the distance to a landmark (for a review, see Cheng, 2000). In the present work, both these methods of measuring distance were used to calculate the scale for distance to a landmark.

In Series 1, honeybees were trained to find sugar water near a cylindrical bottle. Occasional tests then offered tap water at different distances from the landmark, all in the same direction from the landmark. In Series 2, honeybees were trained to find sugar water 20 cm from the landmark. Occasional tests then offered tap water at the same distance but in a different direction from the landmark. I hypothesized that the directional scale would be uniformly circular, meaning that directional discrepancy would increase linearly up to 180° and then decrease linearly from 180° to 360°. To test this assumption, gradients with S+ at different directions from the landmark were obtained. If the scale were uniformly circular, then all generalization gradients would look similar untransformed.

METHOD

Subjects

The honeybees lived in a maintained hive nearby. Subjects were recruited from a feeding station set up just outside the laboratory. A sugar solution (2 parts sugar : 3 parts water) was put in bottle caps (2.5-cm diameter) on the entrance plank to the lab. The plank led the foragers through a 2-cm-square entrance with a guillotine-style door. Subjects (5 for each version of each experiment) were painted for individual identification.

Setups

The rewarded location was on a table (120 cm × 75 cm × 73 cm high) in a laboratory with plenty of directional cues given by doors, windows, and furniture. In each experiment, the bees were trained with one setup, and then tested with five different setups, including the training setup (Fig. 1). The training setup was moved to different parts of the table from trial to trial, so that the arrays shown in Figure 1 were the only valid predictors of reward location. The landmark was a cylindrical bottle (7 cm in diameter, 15 cm high) wrapped in light-blue poster paper. The ring or long strip of cardboard was light yellow.

In a (failed) attempt to create two gradients of different slopes, two experiments on distance generalization were run. In one, the reward container sat on a piece of light-blue poster paper (7 × 5 cm) in training and tests; in the other, the blue paper was absent and the cap sat on the yellow strip. The S+ location for distance generalization (Fig. 1a) was 5 cm from the edge of the landmark. In one version of each experiment, the goal and test locations were east of the landmark, whereas in a second version, the locations were west of the landmark. The test locations were in the middle of the demarcated regions shown in Figure 1a (5 to 45 cm from the landmark). The demarcation lines were drawn in pencil, too thin for the bees to see.

For direction generalization, two versions of each of three experiments were run. The blue poster paper was under the cap in training and tests. The two versions differed in the locations used: (a) 20 cm north, northeast, east, southeast, and south of the landmark or (b) 20

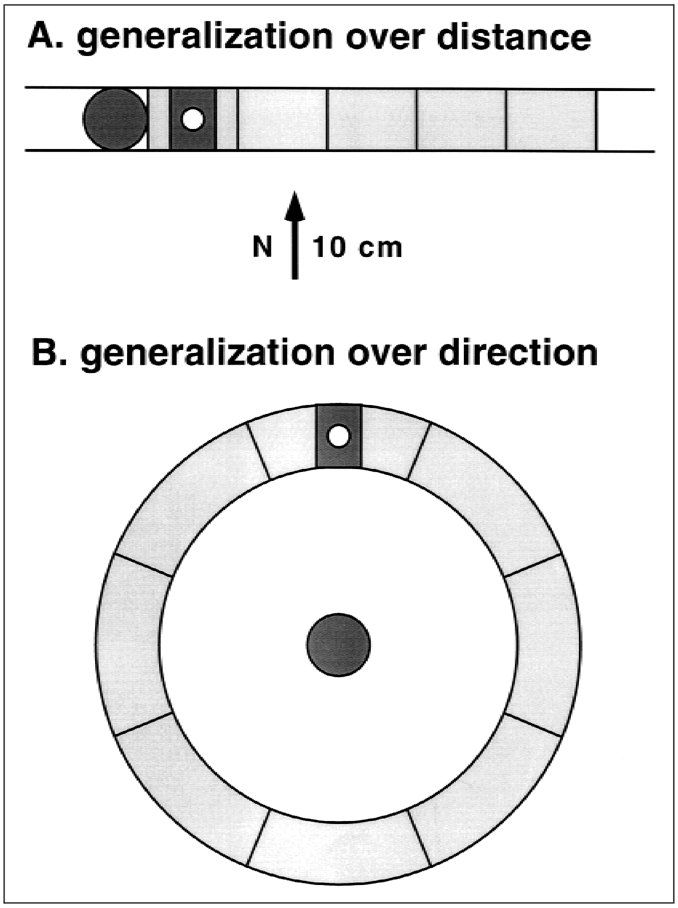

Fig. 1. One version of setups used to study (a) generalization over distance to a landmark and (b) generalization over direction to a landmark. The dark circle indicates the landmark used, a light-blue cylinder. The light-gray strip (a) and ring (b) show strips of light-yellow cardboard paper. In (a), the strip continued the length of the table. Regions on the yellow strip or ring were marked in pencil, with lines too thin for bees to see. The small open circle indicates the location of the reward during training. On test trials, the cap used to hold the reward on training trials was filled with tap water, and could be in the training location or in one of the other demarcated areas. The dark rectangle indicates light-blue poster paper, which was placed under the cap wherever it was located (except that in one experiment testing generalization over distance it was never present). North was the direction from which the bees entered the lab through a window. The location of the setup on the table changed from trial to trial. In other versions of the experiments, the locations in (a) were west instead of east of the landmark and different training and test locations were used in (b).

cm north, northwest, west, southwest, and south of the landmark. The octants of the ring were demarcated in pencil. The test location was always at the middle of an octant. The three experiments differed in S+ location, which was at the north, east or west, or south.

Procedure

During training, a bee was first moved a few times from the entrance to the target location while it was feeding. After that, the forager learned to walk through the gangplank and fly toward the table, and was then trained for 30 to 50 trials. Bees returned after unloading their food at the hive (typically after 5–12 min). All bees in each experiment were trained together.

After training, each bee was tested individually approximately once every three trials. Tap water in a new cap replaced the sugar water. Tests were videotaped from above, and lasted 60 s from the time the bee came into view on camera. After each test, the subject was given a standard training trial. Three tests at each location were given to each subject.

Data Analysis

The video record was counted frame by frame, either on a video monitor or on a computer monitor after the record was fed into the computer. The number of frames that the bee was in the region of the tap water, as demarcated in Figure 1, constituted the data. Statistical tests were considered significant at an alpha of .01.

RESULTS

Series 1: Distance Generalization

All the generalization gradients, from both versions of the two experiments, were very similar, and thus combined for analysis. For curve fitting, results from all subjects were first averaged and then expressed as a proportion of responding at S+. The x-axis was scaled in five different ways. In one, the scale was left untransformed (linear scale). Two scales were based on the retinal size projected by the landmark, and two were based on motion parallax (Fig. 2), these being the cues that bees use to judge distance (see the introduction). For both retinal size and motion parallax, height and width differences were equally weighted, and were combined in either a euclidean or a city-block metric, making four different but similar theoretical scales. Exponential curves fit the data well, without any significant discrepancies, on the theoretical scales, but not on the linear scale (Fig. 3), confirming Shepard's law. In fact, the exponential fit was good no matter how height and width were weighted on the theoretical scales. The scales based on retinal size and parallax are similar to one another (Fig. 3). No matter how parallax and retinal size are combined by the honeybee, distance generalization can be fitted by an exponential function, a finding that supports Shepard's law.

Series 2: Direction Generalization

The generalization gradients from the three experiments (Fig. 4) look similar in shape when they are lined up at their respective S+s. Statistical comparisons confirm this impression. First, the four gradients from both versions of the experiments with S+ at north and south were compared, with the latter gradients reflected, so that the S+ locations lined up in the mixed-model analysis of variance (ANOVA). The ANOVA (on frame counts as the dependent measure) revealed only a main effect of location, $F(4, 56) = 21.72$, indicating that the gradients were similar. No interaction was near significance ($ps > .3$). Next, all six gradients from both versions of each experiment were compared. Only the three locations nearest S+ were taken from the four gradients with S+ at north and south. In two separate ANOVAs, these gradients were compared with three locations from each of the other two gradients, with S+ at east and at west. In one test, the locations north, northeast, east, and north, northwest, west (the *left arms*) were included. In the other test, the locations south, southeast, east, and south, southwest, west (the *right arms*) were included. Using frame counts as the dependent measure, the ANOVA with the right arms of the gradients with S+ at east or west revealed only significant main effects of gradient, $F(2, 22) = 16.47$, and of location, $F(2, 44) = 40.78$. The ANOVA with the left arms of the gradients with S+ at east or west revealed only a significant main effect of location, $F(2, 44) = 25.08$. No interaction was near significance ($ps > .16$). The gradients were thus similar in shape when lined at their respective S+s. This result implies that the scale for direction is uniformly circular. A scale for testing Shepard's law was thus delivered by the data.

Curve fitting was done on a linear x-axis using the average of the gradients with S+ at north and south. Data from the east and west semirings were combined as well to give just one data set ranging from 0° to 180° from S+. Figure 5 shows that the exponential function fit the data well, but the Gaussian, a function sometimes thought to characterize generalization gradients, resulted in misfits at three data points.

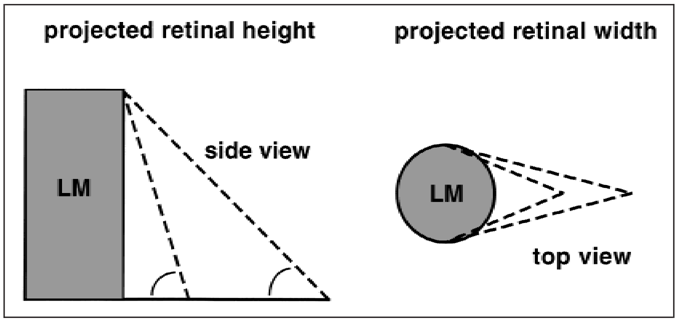

Fig. 2. Measuring projected retinal height and projected retinal width. All measurements were made from the center of the cap, at table level. Two possible cap locations are shown in each panel, at the ends of the dotted lines away from the landmark (LM). Retinal height refers to the angle between the ground and the top of the landmark from a location. Retinal width refers to the angle projected by the width of the landmark. Motion parallax was measured by the change in retinal height when an animal moved 1 mm closer to the landmark, or in the direction to the right edge of the landmark when an animal moved 1 mm to the right. The measurements were transformed to values on the x-axis (representing psychological distance from S+) by taking the difference of each measure from the corresponding measure at S+. Each scale was rescaled to give a maximum value of 1.

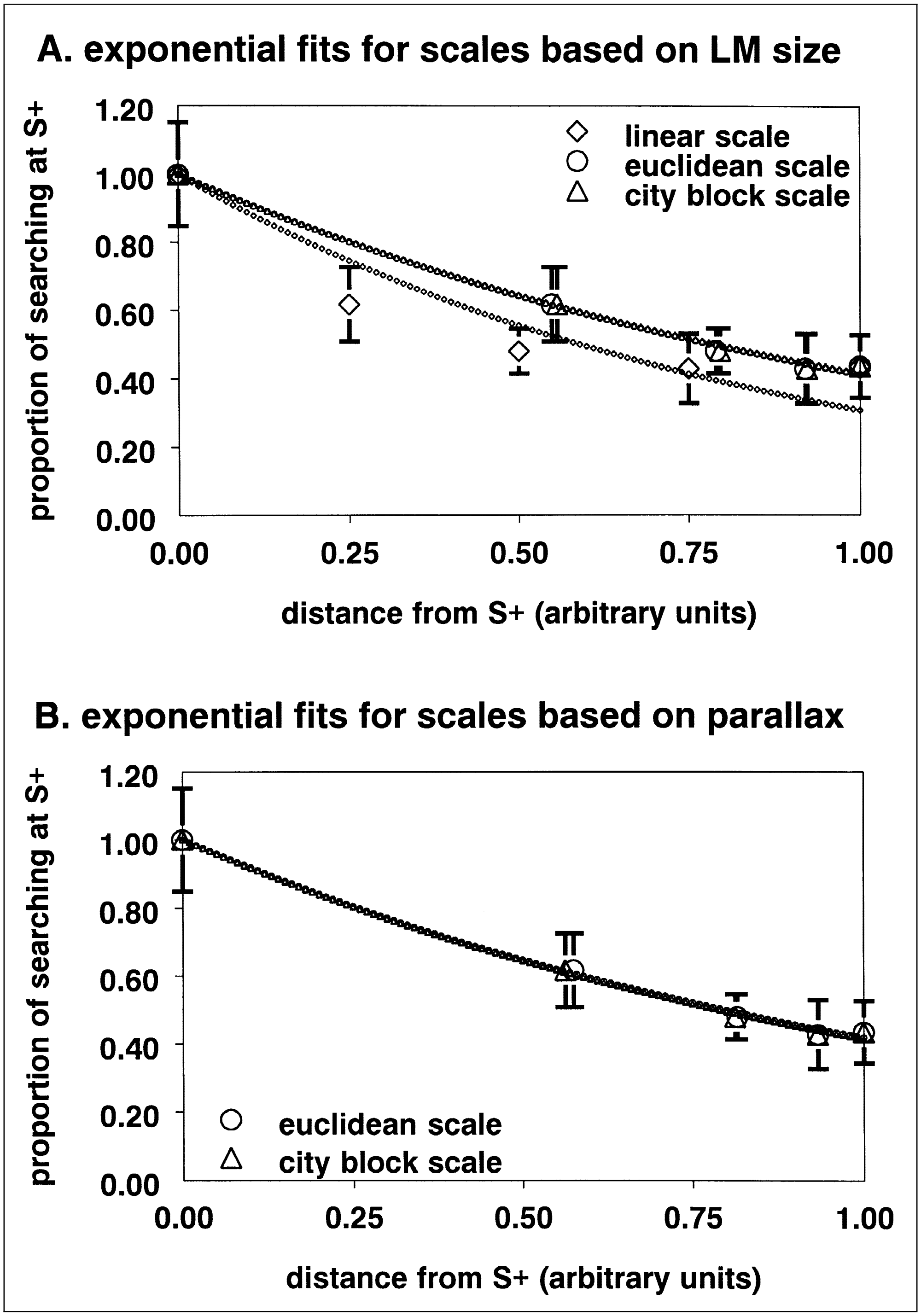

Fig. 3. Best exponential fits for distance generalization (small symbols) based on projected landmark (LM) size (a) and motion parallax (b). Fits were calculated by the least-squares criterion. The data (large symbols) are expressed as proportion of searching relative to searching at S+. Error bars show 95% confidence intervals. The data were obtained by first counting the number of frames in which the bee was in the region in which the cap of tap water was offered; then both means and confidence intervals were relativized (divided) by the mean count on tests at S+. In (a), the best exponential fit based on a linear scale is shown as well. The data points from corresponding locations in each series (e.g., the second data point from the left for all curve fits, representing the location second closest to the landmark) do have the same *y* value. That they appear to differ is a visual illusion.

DISCUSSION

In these experiments, honeybees were tested on spatial generalization. They were trained to find food at one location near a landmark, and then tested without reward at locations differing in distance (Series 1) or direction (Series 2) from the usual reward location. Scales for distance to a landmark were derived from extant theory on how bees measure distance from a landmark. The generated scales all looked similar, so that parameter choices did not matter in fitting the data. The scale for direction to a landmark was uniformly circular, as indicated by the data. With these unadjusted theoretically specified scales, generalization gradients in both series were fitted by exponential functions, supporting Shepard's law.

These experiments extend my previous findings supporting Shepard's law in an invertebrate animal (Cheng, 1999). My earlier work did not separate the dimensions of distance and direction. These experiments did. That the exponential form was found along dimensions of both distance and direction, based on different theoretically derived scales, adds to the generality of the findings. Shepard's law is claimed to be universal on functional grounds: Probabilistic properties of the world should have driven animals, in the course of evolution, to adopt the exponential strategy for generalization. Until these recent experiments, however, the law had been confirmed only in vertebrate animals. These experiments on invertebrates add support to the generality of Shepard's (1987) arguments, and suggest that the structure of the world might have shaped the evolution of cognition in diverse animals.

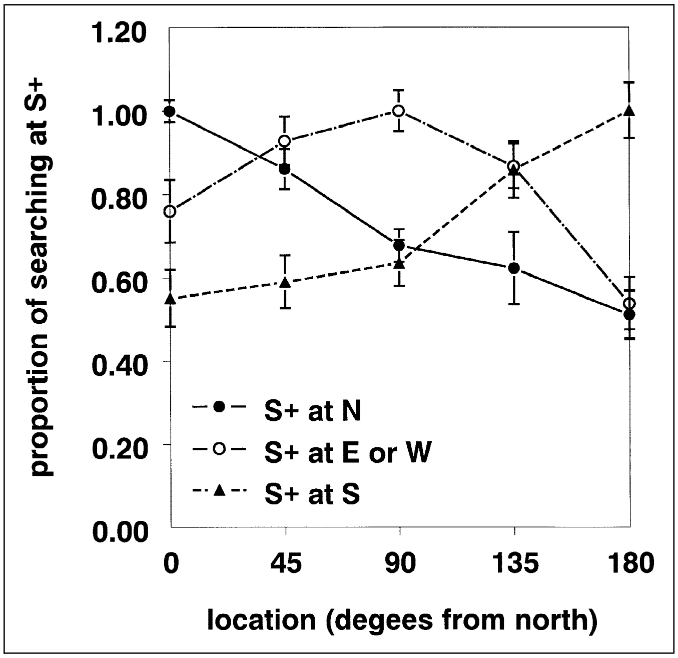

Fig. 4. Generalization gradients from Series 2 expressed as a proportion of searching at S+ (±SEM). Data from two setup versions (one with locations east of the landmark and one with locations west of the landmark) have been combined to make each gradient. The data were obtained by first counting the number of frames in which the bee was in the region in which the cap of tap water was offered; then both means and SEMs were relativized (divided) by the mean count on tests at S+. Directions of S+ are measured from the landmark shown in Figure 1b.

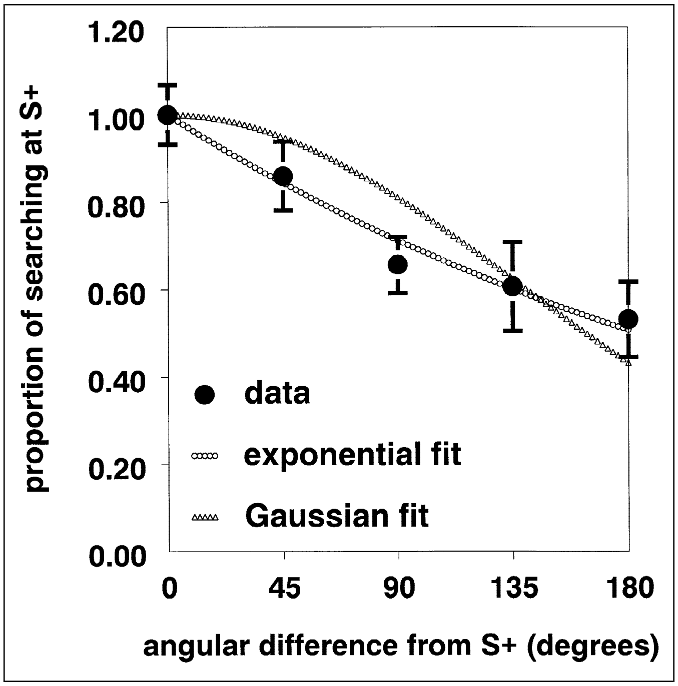

Fig. 5. Best exponential and Gaussian fits of directional generalization. Fits were calculated by the least-squares criterion. The data (large symbols), expressed as proportion of searching relative to searching at S+, are presented with the 95% confidence intervals. The data were obtained by first counting the number of frames in which the bee was in the region in which the cap of tap water was offered; then both means and confidence intervals were relativized (divided) by the mean count on tests at S+. The *x*-axis is left untransformed.

Acknowledgments—The work reported here was supported by a research grant from the Australian Research Council. I am indebted to Roger Shepard for helpful comments. Thanks are also due to Dagong Zhang for data collection and analysis.

REFERENCES

Bitterman, M.E. (1996). Comparative analysis of learning in honeybees. *Animal Learning & Behavior*, *24*, 123–141.

Capaldi, E.A., Robinson, G.E., & Fahrbach, S.E. (1999). Neuroethology of spatial learning: The birds and the bees. *Annual Review of Psychology*, *50*, 651–682.

Cartwright, B.A., & Collett, T.S. (1982). How honeybees use landmarks to guide their return to a food source. *Nature*, *295*, 560–564.

Cartwright, B.A., & Collett, T.S. (1983). Landmark learning in bees. *Journal of Comparative Physiology A*, *151*, 521–543.

Cheng, K. (1999). Spatial generalization in honeybees confirms Shepard's law. *Behavioural Processes*, *44*, 309–316.

Cheng, K. (2000). How honeybees find a place: Lessons from a simple mind. *Animal Learning & Behavior*, *28*, 1–15.

Cheng, K., Spetch, M.L., & Johnston, M. (1997). Spatial peak shift and generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 469–481.

Collett, T.S., & Zeil, J. (1998). Places and landmarks: An arthropod perspective. In S. Healy (Ed.), *Spatial representation in animals* (pp. 18–53). Oxford, England: Oxford University Press.

Ennis, D.M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, *117*, 408–411.

Gluck, M.A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science*, *2*, 50–55.

Lehrer, M., Srinivasan, M.V., Zhang, S.W., & Horridge, G.A. (1988). Motion cues provide the bee's visual world with a third dimension. *Nature*, *332*, 356–357.

Nosofsky, R.M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R.M. (1988). On exemplar-based exemplar representations: Reply to Ennis (1988). *Journal of Experimental Psychology: General*, *117*, 412–414.

Reid, A.K., & Staddon, J.E.R. (1998). A dynamic route finder for the cognitive map. *Psychological Review*, *105*, 585–601.

Shepard, R.N. (1958a). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65*, 242–256.

Shepard, R.N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*, 509–523.

Shepard, R.N. (1965). Approximation to uniform gradients of generalization by monotone transformations of scale. In D.I. Mostofsky (Ed.), *Stimulus generalization* (pp. 94–110). Stanford, CA: Stanford University Press.

Shepard, R.N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, *115*, 58–61.

Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Shepard, R.N. (1988). Time and distance in generalization and discrimination: Reply to Ennis (1988). *Journal of Experimental Psychology: General*, *117*, 415–416.

(RECEIVED 7/1/99; REVISION ACCEPTED 1/13/00)