

Processing of hierarchic stimulus structures has advantages in humans and animals

Martina Siemann, Juan D. Delius

Allgemeine Psychologie, Universität Konstanz, D-78434 Konstanz, Germany

Abstract. Carmesin and Schwegler (1994) have determined theoretically that a linear hierarchical stimulus structure can be encoded by a parallel network of minimal complexity. The experiments reported here compare the efficiency with which humans and pigeons process sets of stimulus pairs embodying different inequality structures. Groups of subjects of each species were taught to discriminate all 10 pairwise combinations of 5 stimuli with an operant conditioning method. For one group, the reward/punishment allocations within the pairs agreed with a linear hierarchy. For a second and third group, the reinforcement allocations of one or three, respectively, of the stimulus pairs deviated from such ordering. The time it took the subjects to learn the tasks as well as the final choice latencies and/or error rates increased with the number of deviating inequalities. The results agree with the assumption that both humans and pigeons encode stimulus inequality structures with parallel processing neural networks rather than with a sequentially processing algorithm.

1 Introduction

Most higher animals, including humans, are deluged with sensory information in the course of normal life. As information processing is associated with definite costs, any reductions of redundancy within this incoming stream generally signifies an advantage in fitness. Indeed, there is evidence that advanced animals have evolved several strategies to cope with this problem. A tendency towards information chunking (Wickelgren 1979; Terrace 1987) and information categorization (Medin and Smith 1984; Herrnstein 1990), both involving a clustering of item information into larger units, are well-documented examples. Organization of item information in an economic hierarchical form offers itself as a further such strategy in certain contexts (Riley 1976). The six item inequalities $A > B$, $A > C$, $A > D$, $B > C$, $B > D$, and

$C > D$, for example, can be conveniently transformed without loss into the more concise, summary proposition $A > B > C > D$. As the number of items increases, the economy of a single hierarchical ordering is magnified. With 25 items, for example, the number of pairwise inequalities rises to no less than 300. Conversely, assuming a linear hierarchical organization of items, a minimum number of pairwise relations, for example $A > B$, $B > C$, $C > D$, and $D > E$ are sufficient to specify the ordering $A > B > C > D > E$. This, in turn, allows inference, in this instance, of six further pairwise relations between items. Among them is the $B > D$ relationship that plays an important role as the critical conclusion in many experiments concerned with deductive thinking in humans and animals.

A transitive inference experiment based on this latter schema was carried out by Fersen et al. (1991) using pigeons. The birds were taught to discriminate the stimulus pairs $A + B -$, $B + C -$, $C + D -$ and $D + E -$, where the letters stand for different small visual patterns, each displayed on a pair of response discs of a conditioning chamber, and the + and the - signs indicate that pecks directed to the corresponding patterns were either rewarded with a quantity of grain or punished with a period of darkness. When the animals performed well on these premise pairs, preferentially choosing the rewarded patterns, the inference test pair $B D$ was presented. The pigeons pecked the stimulus B in preference to D though neither reward nor punishment was issued during the corresponding trials. This and some further tests suggest that they had inferred the hierarchy $A > B > C > D > E$ as underlying the premise pairs. Similar results have been obtained in equivalent experiments with humans (Werner et al. 1992; Siemann and Delius 1993).

Motivated by Fersen et al.'s findings, Carmesin and Schwegler (1994) proposed that the pigeons' transitive behavior was controlled by an adaptive neural network. They suggest that the transitivity shown during a test is a consequence of a network complexity minimization process. This led them to consider not only the above-mentioned linearly hierarchical set, but all the inequality assignment sets possible with the 10 stimulus pairs that can be assembled from 5 individual stimuli. Each of these

Table 1. Learning tasks for subject groups G0, G1 and G3. Task class numbering and minimal complexity scores taken from Carmesin and Schwegler (1994). Plus and minus signs symbolize reward and punishment. The numbers in the rightmost columns indicate the overall frequencies of these reinforcements appertaining to the different A to E stimuli. The normal, boldface and italic entries are explained in the text

| | A | B | C | D | E | + | - |
|--|----------------|----------------|----------------|----------------|----------------|---|---|
| <i>Group G0, task 1, parallel score = 4, serial score = 3</i> | | | | | | | |
| A | | A + B - | <i>A + C -</i> | <i>A + D -</i> | <i>A + E -</i> | 4 | 0 |
| B | | | B + C - | <i>B + D -</i> | <i>B + E -</i> | 3 | 1 |
| C | | | | C + D - | <i>C + E -</i> | 2 | 2 |
| D | | | | | D + E - | 1 | 3 |
| E | | | | | | 0 | 4 |
| <i>Group G1, task 7, parallel score = 5, serial score = 4</i> | | | | | | | |
| A | | A + B - | <i>A + C -</i> | <i>A + D -</i> | | 3 | 1 |
| B | | | B + C - | <i>B + D -</i> | <i>B + E -</i> | 3 | 1 |
| C | | | | C + D - | <i>C + E -</i> | 2 | 2 |
| D | | | | | D + E - | 1 | 3 |
| E | <i>E + A -</i> | | | | | 1 | 3 |
| <i>Group G3, task 12, parallel score = 7, serial score = 4</i> | | | | | | | |
| A | | A + B - | <i>A + C -</i> | | | 2 | 2 |
| B | | | B + C - | <i>B + D -</i> | | 2 | 2 |
| C | | | | C + D - | <i>C + E -</i> | 2 | 2 |
| D | <i>D + A -</i> | | | | D + E - | 2 | 2 |
| E | <i>E + A -</i> | <i>E + B -</i> | | | | 2 | 2 |

12 sets constitutes a multiple pair discrimination task that can actually be presented to experimental subjects. To characterize the complexity of the minimal theoretical parallel network capable of encoding each of these tasks optimally, Carmesin and Schwegler derived a special measure, a so-called minimal complexity score. If the pairwise inequalities within a task are completely reducible to a linear hierarchy, the score turns out to be small as compared with scores related to tasks that cannot be reduced in this way. For comparison purposes, the same authors also computed quantitative scores characterizing the minimal theoretical program capable of producing perfect task solutions when these are processed by a serial Turing machine (see their Table I). The reader is referred to Carmesin and Schwegler's paper for theoretical arguments and precise derivations.

The experiments we now report extend studies on the transitive inference competences of both pigeons (Fersen et al. 1991; Siemann 1993b; Siemann et al., manuscript submitted) and humans (Werner et al. 1992; Siemann and Delius 1993; Siemann 1993a) carried out in our laboratory. Of course, these experiments were inspired by the above-mentioned theoretical study. Using both humans and pigeons, the experiments determine actual performance on a critical selection of the tasks examined there. Different subject groups learned tasks where either all stimulus pairs conformed with a linear hierarchy or where one or three pairs, respectively, deviated from such a structure. These three tasks, each consisting of 10 pairs and laid out in detail in Table 1, were chosen to be crucial for the above-mentioned parallel/serial processing distinction.

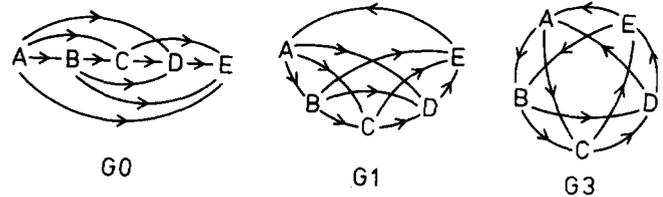


Fig. 1. Graphic representations of the task structures G0, G1, and G3. The arrowheads represent inequalities, not directions

Notice that the four boldface printed pairs composed of symbolically adjacent stimuli are the same for all three tasks. They are those commonly serving as premise pairs in transitive inference problems and from which subjects usually infer an A to E hierarchy. The other six pairs composed from symbolically non-adjacent stimuli normally serve as inference tests. Three of these pairs, printed in italics, consist of stimuli bridging just one symbolic position. Their reinforcement allocations are also the same for all three tasks. The reinforcement allocations of the remaining three stimulus pairs, printed in normal type and bridging more than one symbolic position varied between tasks. While the G0 task complies with a linear structure, the G3 task agrees with a circular structure; the G1 task, however, reflects an odd stimulus structure (Fig. 1). Note that, according to Carmesin and Schwegler (1994), the minimal complexity with a parallel network processing increases from G0 through G3 but that with a sequential Turing processing there is only a complexity increment from G0 to G1, while G1 and G3 tasks have identical complexity indices.

2 Method

2.1 Humans

Thirty human subjects (14 male, 16 female, mean age 22.8 years) took part in the experiment. Payment was related to choice accuracy performance. Subjects were randomly allocated to three groups of 10. Each group had to learn to discriminate 10 stimulus pairs assembled from 5 patterns. These were simple geometric figures, all of similar complexity. Half of the subjects in each group were confronted with the stimulus pairs as listed in Table 1; the other half received the pairs with the reinforcement allocations reversed (F + E -, E + D -, D + C -, etc.). This compensated for any interfering spontaneous stimulus preferences.

The tasks were implemented as computer games. A personal computer (Compaq 386) and a flat-screen monitor (Zenith) with an infrared touchscreen (Carrol) were used. Subjects were instructed to respond as accurately and as fast as possible. A trial began with the side by side presentation of two stimuli on the screen. Subjects had to choose one of the two stimuli by tapping it with a pencil. The time which elapsed from the appearance of the stimuli to the choice reaction was recorded as latency. The stimulus selected moved into an opening centered at the top of the screen. If the choice was correct, the stimulus was seen to fit into the template. If the choice was incorrect, it did not. A coin was accordingly either added or removed from a symbolic kitty. A virtual button appeared simultaneously and had to be tapped to call up the next trial. The right-left position of the positive and negative stimulus was determined quasi-randomly from trial to trial (Gellermann 1933).

Each subject participated in a single learning session. The session consisted of successive blocks of 100 consecutive trials, involving 10 presentations of each of the 10 stimulus pairs. To facilitate learning during the initial block, pairs were presented in runs involving first 5, then 3 and finally 2 consecutive pair repetitions. From the second block onwards, the pairs were randomly sequenced. The session ended when the subjects achieved a criterion of 70% correct responses on each of the pairs within one block of trials or, alternatively, when they had completed 12 blocks.

At the end of the session, the subjects answered a questionnaire concerning what they thought the experiment was about and how they had coped with the stimulus pairs. Finally, they were asked to lay out 5 cards, each bearing one of the stimuli, into any sequence(s) their recent experience suggested to them. They could ask for additional cards if they felt they needed them.

2.2 Pigeons

Nine pigeons of local homing stock were used. Birds were housed in individual cages in a well-ventilated room with a 12 h light-on 12 h light-off schedule and maintained on a food deprivation diet until they achieved 80% of their normal weight. The birds were randomly allocated to three groups, each dealing with one of the

above-mentioned tasks, and were trained in a cage similar to the home cages. A small panel, containing two horizontal, side-by-side pecking keys and two food dispensers, was attached to a cage wall opening. The stimuli (5 different geometric figures) were displayed with two miniature 7 × 5 red light-diode matrices that were mounted immediately below the transparent keys. All experimental events were steered and recorded by a Commodore 64 computer with a Dela interface programmed in Psychobasic (Xia et al. 1991).

Pecks directed to the positive stimulus led to 2–4 grains of millet being delivered directly onto the corresponding key. If the animals chose the negative stimulus, the whole corresponding matrix turned to red for 3 s. Following that, they were permitted to correct themselves. Sessions took place daily and consisted of 100 trials, 10 with each stimulus pair, presented in a quasi-random sequence, with the side position of the positive stimulus similarly randomized (Gellermann 1933). During the first 5 sessions the response requirement for reinforcement was gradually increased from 1 to 5 pecks. This latter FR5 schedule was then retained for the remaining 20 sessions.

3 Results

3.1 Humans

Since the results of the subgroups with reversed reinforcement allocations did not yield any significant differences, they are henceforth lumped together. The notation laid out in Table 1 is accordingly used indiscriminately as applying to both data sets.

The number of learning blocks to achieve the criterion differed markedly between the three groups: group G0 subjects required a mean of 3 (range 2–4) blocks, group G1 subjects a mean of 5.1 (2–10) blocks and group G3 subjects a mean of 6.3 (2–12) blocks. Indeed, two subjects belonging to this last group reached the 12-block limit without achieving the 70% accuracy criterion. These acquisition differences between the groups are statistically significant [$H(2,30) = 7.78$; $P < 0.05$].

Even when the subjects had learned to discriminate the respective stimulus pairs to the set accuracy ($> 70\%$ choices correct, last block), there were still performance differences between the groups. Figure 2 presents the corresponding data where the pairs are grouped according to the symbolic distance they incorporate (D1: AB, BC, CD and DE, boldface in Table 1; D2: AC, BD and CE, italicized in Table 1; D3: AD and BE; D4: AE, normal print in Table 1). This form of data reduction was chosen because, in transitive inference experiments, it is usual to find a so-called symbolic distance effect, such that the error rates and choice latencies decrease with increasing distance between the stimuli making up a pair (Potts 1972; Trabasso and Riley 1975). It must be conceded, though, that this way of proceeding does not quite do justice to the G3 data since this group dealt with a circular stimulus structure (Fig. 1). However, group G0 dealing with a linear structure task could in

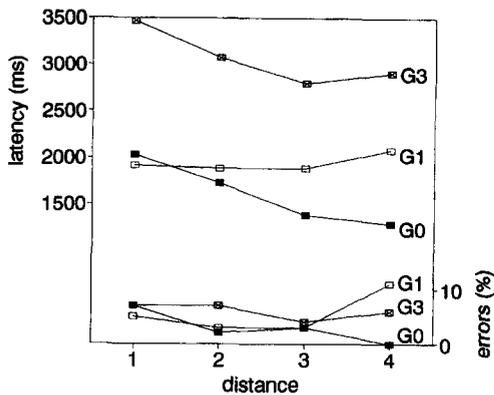


Fig. 2. Mean error rates and choice latencies of groups G0, G1, and G3 as a function of the symbolic distance characterizing the various stimulus pairs

fact be expected to exhibit the above-mentioned distance effect.

As subjects had been trained to fulfill an accuracy criterion on each of the stimulus pairs, it is not surprising that the overall choice accuracies of the groups do not evince statistically significant differences. Analyzing the different distances separately reveals that only D4 is associated with a significant group difference [$H(2,28) = 8.4$; $P < 0.05$]. The overall choice latencies, on the other hand, increase from G0 to G3, with G3 standing out quite clearly on this count. These differences are highly significant [$H(2,28) = 13.9$; $P < 0.001$]. Moreover, the choice latencies differ significantly for each of the distances [D1: $H(2,28) = 13.1$; $P < 0.01$; D2: $H(2,28) = 12.2$, $P < 0.01$; D3: $H(2,28) = 11.7$; $P < 0.01$; D4: $H(2,28) = 14.1$; $P < 0.001$]. As expected, group G0 shows a clear symbolic distance effect, the latencies decreasing as the distance separating the stimuli paired increases ($L = 295$, $P < 0.01$). A similar trend is unexplainably (see above) also present in the G3 data ($L = 223$, $P < 0.01$). The G1 data reveal no significant trend.

The questionnaire answers and the card sorting yielded additional information. Nine out of the ten group G0 subjects reported that they eventually dealt with the task by referring back to a serial order that they had memorized. These subjects did indeed also lay out the cards in the correct order. In group G1, only six subjects reported such a strategy, but of those, only four were able to lay the cards out in the correct sequence, three of them asking for a further stimulus card (A) to compose the $E > A$ pair deviating from the linear hierarchy. However, two subjects who did not report any particular strategy, and who did not lay the cards correctly, nevertheless performed at an above median level in terms of both speed and accuracy. Such a dissociation between knowledge and performance is actually a quite common finding in transitive inference experiments (Siemann and Delius 1993). The dominant strategy reported by group G3 subjects was that they memorized the stimuli pairwise using verbal markers. One subject said that he clustered the stimuli into transitive triads (e.g. $D > E > A$; $C > D > E$, etc.). None of the

G3 subjects, however, laid out the cards to match the correct overall structure.

3.2 Pigeons

All three animals in groups G0 and G1 and one animal in group G3 completed the target of 25 sessions. Two animals in group G3 ceased to respond after three and four sessions, respectively. Figure 3 displays the mean error rates in percent as a function of the progression of blocks of five sessions. Groups G1 and G3 clearly exhibited a slower learning than group G0. The mean error rate of this latter group is significantly below the 50% chance level (binomial, $P < 0.01$).

The percent error rates were averaged across the last block of five learning sessions and plotted as a function of the symbolic distance characterizing the various stimulus pairs (see above under Humans). They are shown in Fig. 4. The accuracy functions produced by the pigeons reveal similarities with the corresponding speed functions of humans (Fig. 2). The error rates of group G0 reveal a clear symbolic distance effect ($L = 88.5$; $P < 0.01$). Group G1 evinces strikingly increased error rates for the distance D4 over the other distances. As with the human

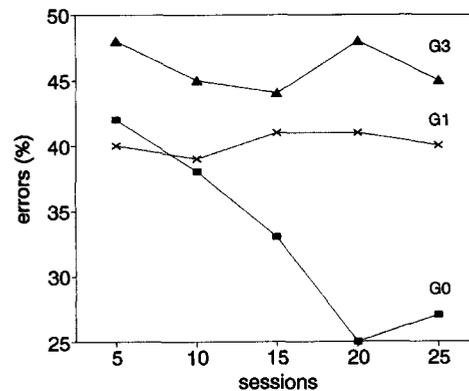


Fig. 3. Learning performance of pigeons averaged for blocks of five sessions

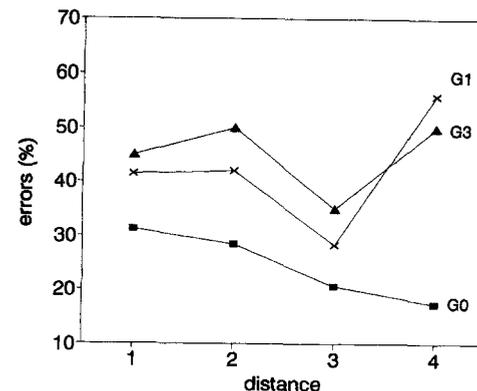


Fig. 4. Mean error rates for group G0, G1, and G3 pigeons as a function of the symbolic distance characterizing the various stimulus pairs

subjects, this seems reasonable, since the D4 pair is the only one that deviates from a linear hierarchy. The group G3 pigeon showed nearly random choices. The overall performance of the groups in any case differs significantly when the two abortive pigeons of group G3 are assigned the worst ranks [$H(2,9) = 7.26, P < 0.05$].

4 Discussion

The results show that regardless of the species, a stimulus structure embodying a linear hierarchical order facilitates both the acquisition and retrieval of pairwise inequalities between stimuli. Mounting exceptions from such a structure impair both learning and memory performance. However, most human subjects also managed to cope with the most difficult task, G3, at the expense of long reaction times, while the pigeons really failed, either by not discriminating properly or by ceasing to respond. Awareness of the structure underlying the various multiple discriminations by human subjects is comparably affected by departures from a hierarchy. Generally, the results are in agreement with the minimal complexity ranking that Carmesin and Schwegler (1994) derive for parallel network solutions. The clear-cut performance difference between groups G1 and G3 contradicts the identical complexity that those authors establish for sequential algorithms coping with these tasks.

At a more intuitive level, the greater difficulty of the G1 task as compared with the G0 task is perhaps not surprising. Subjects clearly learn the latter task by gradually recognizing the consistent order of the stimuli, while the former task includes one freak pair relationship that needs to be separately coded. But at the same intuitive level, one is inclined to consider task G3 as relatively easy since it involves a smooth circular structure. Subjects could be initially expected to recognize that structure when dealing with the pairs formed by adjacent stimuli (Fig. 1, G3, outer pairs) and then to extend the principle to the pairs consisting of nonadjacent stimuli (inner pairs). But as others have suggested before us (Henley et al. 1969; von Fersen et al. 1991), humans and pigeons are clearly not prone to grasp any but linearly hierarchical structures. It has been speculated that this may represent an adaptation to the admittedly also disputed fact (Archer 1992) that real hierarchies in human and animal societies are as a rule linear (von Fersen and Delius 1992). Of course, Carmesin and Schwegler (1994) probably have put a finger on a less involved, more powerful reason, namely computational economy.

A further aspect, though certainly related to the neural network approach, also has to be briefly considered. In connection with the study of Fersen et al. (1991), several rather simple learning models have been shown to adequately reproduce transitive responding (Couvillon and Bitterman 1992; Wynne et al. 1992). Since a conditioning technique was also used to implement the present experiments, it is important to attend to the total allocations of reward and punishment to the five stimuli. These are listed down the two last columns of Table 1

under the headings + and -. For G0, the figures are compatible with a consistent stimulus value hierarchy, A always being rewarded and E always punished, the other stimuli falling suitably in between. For G1, there is a reinforcement equivalence of A and B (predominantly rewarded) and again of D and E (predominantly punished), C being intermediate. In G3, all five stimuli are the same in the sense that they are as often rewarded as they are punished. Without going into any specific modelling, it is patent that the gross reinforcement allocations to the individual stimuli in the G0 task are directly conducive to a generally adequate discrimination behavior, whereas that is only very partially the case for task G1 and not at all so for task G3.

Nevertheless, the conditioning explanations referred to are not fully in agreement with the Carmesin and Schwegler (1994) network proposals. While the pigeons' performance on all three tasks can be adequately simulated with simple conditioning models, the partially successful human performance on task G3 cannot. Learning models such as Luce's β -operator (Luce 1959) or Rescorla and Wagner's (1972) model are based on the assumption that the associative value of a given individual stimulus increases after reward and decreases after punishment. Thus, the G3 task in which each stimulus is as often rewarded as it is punished inevitably yields identical values and consequently chance stimulus choices (Siemann 1993a). For the Carmesin and Schwegler network, however, the G3 task is complex but not impossible. With certain modifications though, the conditioning models can also yield precise predictions for the performance of both species on any of the three tasks, although then at the cost of increased complexity (Siemann 1993a; Siemann and Delius, manuscript submitted).

Regardless of these issues, the present results support the view that, firstly, in humans and animals, relational stimulus structures are likely to be processed by neural network-like parallel processing and that, secondly, linearly hierarchical stimulus structures inherently have processing advantages that are in consonance with computational economy arguments.

Acknowledgements. The research was supported by the Deutsche Forschungsgemeinschaft. We thank H.-O. Carmesin and H. Schwegler for providing us with an early draft of the accompanying paper and for helpful discussions, A. Niemuth for help with the manuscript, L. Xia for technical assistance and Julia Delius for editing the English.

References

- Archer J (1992) Ethology and human development. Harvester Wheatsheaf, New York
- Carmesin HO, Schwegler H (1994) Parallel versus sequential processing of relational stimulus structures. *Biol Cybern* 71:523-529
- Couvillon PA, Bitterman ME (1992) A conventional conditioning analysis of "transitive inference" in pigeons. *J Exp Psychol [Anim Behav Proc]* 18:308-310
- Fersen L von, Delius JD (1992) Schlußfolgerndes Denken bei Tauben. *Spektr Wiss* July:18-22
- Fersen L von, Wynne CDL, Delius JD, Staddon JER (1991) Transitive inference formation in pigeons. *J Exp Psychol [Anim Behav Proc]* 17:334-341

- Gellermann LW (1933) Chance orders of alternating stimuli in visual discrimination experiments. *J Genet Psychol* 42: 206–208
- Henley N, Horsfall R, De Soto C (1969) Goodness of figure and social structure. *Psychol Rev* 76:194–204
- Herrnstein RJ (1990) Levels of stimulus control: a functional approach. *Cognition* 37:133–166
- Luce DR (1959) Individual choice behavior. Wiley, New York
- Medin DL, Smith EE (1984) Concepts and concept formation. *Annu Rev Psychol* 35:113–138
- Potts GR (1972) Information strategies used in the encoding of linear orderings. *J Verbal Learn Behav* 103:431–439
- Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF (eds) *Classical conditioning II: current research and theory*. Appleton-Century-Crofts, New York, pp 64–99
- Riley CA (1976) The representation of comparative relations and the transitive inference task. *J Exp Child Psychol* 22:1–22
- Siemann M (1993a) Transitive Inferenz, experimentelle Untersuchung einer kognitiven Leistung. Dissertation, University of Konstanz
- Siemann M (1993b) Transitive inferences in pigeons and humans. In: Elsner N, Heisenberg M (eds) *Gene, brain, behavior*. Thieme, Stuttgart, p 857
- Siemann M, Delius JD (1993) Implicit deductive responding in humans. *Naturwissenschaft* 80:364–366
- Trabasso T, Riley CA (1975) The construction and use of representation involving linear order. In: Solso RL (ed) *Information processing and cognition. The Loyola Symposium*. Erlbaum, New York, pp 381–410
- Terrace HS (1987) Chunking by a pigeon in a serial learning task. *Nature* 325:149–151
- Werner U, Köppl U, Delius JD (1992) Transitive Inferenz bei nicht-verbaler Aufgabenstellung. *Z Exp Angew Psychol* 19:662–683
- Wickelgren WA (1979) *Cognitive psychology*. Prentice Hall, Englewood Cliffs
- Wynne CDL, Fersen L von, Staddon JER (1992) Pigeons' transitive inferences are the outcome of elementary conditioning principles: a response. *J Exp Psychol [Anim Behav Proc]* 18:313–315
- Xia L, Wynne CDL, Münchow-Pohl F von, Delius JD (1991) Psychobasic: a Basic dialect for the control of psychological experiments with the Commodore-64 and Dela interfacing. *Behav Res Method Instr Comp* 23:72–76