The study of behavioral dysfunctions

Staay, Franz Josef van der

5

The study of behavioral dysfunctions using animal models: summary, conclusions, and recommendations for future research

Introduction

Animal models of behavioral dysfunction serve two main aims, namely 1) enhance our understanding of the underlying substrates and mechanisms, i.e. the brain-behavior relation, and 2) assess the effects of putative neuroprotective, anti-degenerative, revalidation-supporting, and/or cognition-enhancing compounds or treatments.

This book describes studies that were performed to assess behavioral changes in rodent models of central nervous system (CNS) deficiency. Three types of deficiency model were considered: normal aging; CNS-specific lesions; and ischemia. Each of these models was used in a series of experiments, the results of which are summarized below.

Summary of the studies performed

Chapter 2: age-related changes in learning and memory in rats

Aging, the *"time-dependent decline in function which affects all tissues and organ systems"* (Hazzard, 1991, p.225) that ultimately leads to behavioral dysfunction, is receiving growing attention from the scientific community because the proportion of aged people in industrialized countries and developing countries is steadily increasing (Martin, 1991; Olshansky, Carnes & Cassel, 1993; Holden, 1996; Butler, 1997). Despite this attention, the processes underlying normal aging, let alone pathological aging, are still poorly understood. Animal models are an essential tool for investigating these processes.

Cognitive aging is, for example, reflected by changes in spatial orientation. Spatial cognition appears to be compromised in old humans (Evans et al., 1984; Kirasic, 1991; Simon et al., 1992; Uttl & Graf, 1993; Kirasic et al., 1996) and in patients suffering from senile dementias (Flicker et al., 1984; Liu, Gauthier & Gauthier, 1991). These age-related impairments have been found in both laboratory settings and real-life situations (Simon et al., 1992). To gain insight into these cognitive changes we investigated age-associated cognitive decline in rats, using the Morris water escape task which measures spatial orientation performance (Morris, 1984).

In a series of three experiments, described in Chapter 2.1, we compared the spatial discrimination performance of adult and aged outbred Janvier Wistar (WISRJ) rats, of young and old inbred Fischer 344 (F344), and of young and old hybrid Fischer 344*Brown Norway (FBNF1) rats. The aged rats of the WISRJ and FBNF1 strains showed a poorer acquisition of the platform escape behavior and a

weaker bias for the previous platform position in the probe trial than their younger counterparts. The aged rats of the Fischer 344 strain were unable to acquire the task. We concluded that F344 rats should not be used for gerontological or gerontopharmacological studies *on spatial discrimination learning in the Morris task.* However, if this strain is used, then the oldest rats to be investigated should be 20 months or younger.

In the first two experiments of Chapter 2.2, we assessed the effects of aging on the acquisition of the standard Morris water escape task by Winkelmann Wistar rats (WISW). These experiments revealed that, in this strain, clear age-related impairments in the acquisition of the task appeared between 19 and 24 months of age. Therefore, in the third experiment of Chapter 2.2, we used 3-month-old and 24-month-old WISW rats to assess the effects of age on working memory (WM) in the repeated acquisition modification of the Morris task. In this task, which has originally been designed by Whishaw (1985, 1987), each of four start positions in the pool is used randomly in each series of four trial pairs of a daily training session. The rats are randomly started from each of the four starting positions on each trial of a pair. For each daily session, the escape platform is in a different position. The decrease in escape latency and in the distance swum to reach the escape platform from the first to the second trial of a trial pair is considered a measure of spatial WM.

The young rats acquired the task within the first sessions. In contrast, the 24-month-old animals did not acquire the task, even after 12 daily training sessions. It is not clear whether this poor performance of the old rats reflects impaired WM, or whether they did not acquire the procedural aspects of the task. However, the results obtained in the standard Morris task showed that aged WISW rats can acquire the escape response, even though they do not attain the performance of their young counterparts.

Age-associated decreases in spatial discrimination performance are usually seen in cross-sectional studies, in which the performance of naive young animals is compared with that of naive aged animals. However, a few longitudinal studies have shown that spatial discrimination performance is sometimes preserved in the aged animal, if the animal acquires the task at a younger age (e.g. Beatty, Bierley & Boyd, 1985). In Chapter 2.3, we performed two experiments in which albino Wistar rats acquired the Morris water escape task for the first time at the age of 25 months. Retention performance in the Morris task was tested approximately 3 and 5 months later. The performance of the aged animals was not only preserved, but was actually better in the retention tests. The clearest improvement was seen about 3 months after the original acquisition, and the effect was more pronounced in the first than in the second experiment.

One of the factors responsible for the differences between the two experiments reported in Chapter 2.3 might have been the occurrence of genetic drift in the rat strain used (extensively discussed in Chapter 2.4). Although age-related deficits in spatial learning and memory performance in naive rats can replicably be found in rats in cross-sectional studies using the Morris water escape task (experiments 1, 2, and 3 in Chapter 2.1, experiments 1, and 2 in Chapter 2.2), this task appears to be unsuited for the evaluation of age-associated deficits of spatial memory performance in old Wistar rats (up to an age of 30 months) in longitudinal studies.

We found there to be an undesirably large variation in results between experiments when we used aged WISW rats. For this reason, we decided to investigate the replicability of spatial discrimination performance in the standard Morris water escape task in Chapter 2.4, by comparing the learning curves and the performance in a probe trial of 24-month-old outbred WISW control rats from 36 experiments. These experiments had been performed at our laboratory under strictly controlled conditions over a period of 71 weeks. There was a very high variability in the learning curves between

experiments. The initial performance, i.e. the performance during the first session, did not change systematically across the 36 experiments, whereas the final performance, i.e. the performance reached in the fifth training session, decreased over the 71-week period, when the platform escape latency and the distance swum to reach the platform, measured as number of line crossings, were considered. In fact, in the last experiments of the series, learning curves were no longer seen: the rats did not improve their performance across the acquisition sessions.

By contrast, the swimming speed and, in the probe trial, the bias for the quadrant where the platform had been positioned during training did not change. This indicates that, across experiments, *spatial orientation learning* decreased, whereas *motor performance* appeared to be unchanged. The most obvious factor to explain these differences between experiments is that the cohorts or shipments of rats were different. Mos and Hollander (1987), for example, found a wide variation in the survival characteristics of rats of the inbred WAG/Rij (WAG) and Brown Norway (BN) strains, and this was even true for successive cohorts. Short- and long-living cohorts were seen over the 5 years of their study but there was no consistent trend. The failure to find a trend is consistent with the fact that both the WAG and the BN strains are inbred. Our regression analyses (Chapter 2.4) indicated that genetic drift had occurred in the outbred WISW rat strain, as reflected by the shift in performance of the aged rats in the Morris water escape task.

Critical discussion of the model of the old animal

Dean and colleagues (1981), investigating age-associated changes in sensorimotor and cognitive behavior over the lifespan of C57BL mice in a cross-sectional study, concluded: *"The similarity of these results across the life span of the C57 mouse with those previously reported for other aged mammalian species demonstrates that certain common types of behaviors seem to be impaired selectively by increased age across mammalian species and raises the possibility that common neurological etiologies may exist for these behavioral deficits."* (p.427). According to Decker (1995), the normal, aging animal represents a reasonably isomorphic model for the conditions which produce age-associated deficits in humans.

The aging animal as model for the aging human thus certainly possesses a high face validity. In humans, it is very difficult to control for cohort differences between young and old subjects, especially with respect to health-related factors, but also with respect to education and events with a profound impact on the life of entire generations (Rodin, 1986), such as war. In animal studies these differences can be minimized; breeding and housing conditions can be controlled and the health status monitored. It appears that, as in humans, there is a dissociation between chronological and biological aging in rodents (Collier & Coleman, 1991). Animal studies can thus be of great help in determining the relative contributions of various factors to age differences in learning and memory performance.

However, most aging studies are based on the comparison of two age groups only (this also includes most of our investigations on aging). Typically, young or young-adult animals are compared with aged counterparts. This approach does not allow comparison of the *aging curves* for neurobiological and psychological parameters (Ingram, 1996). Instead, this approach is only suited to clarify whether the measures under investigation decrease with age. Multiple age groups are needed to obtain relevant information about the shape of the aging curve or curves for particular measures. Unfortunately, the slopes of the age regressions are not always linear (e.g. Ingram, 1983), and the most profound age-related decline often occurs over a relatively narrow age range, for example between the ages of 19

and 24 months for the Morris water escape performance of WISW rats (Chapter 2.2), or somewhere between the ages of 19 and 25 months for the reference memory (RM) performance of Brown Norway rats in the spatial holeboard discrimination task (van der Staay, van Nies & Raaijmakers, 1990). If one considers that different behaviors age at different speeds, then there should be very small differences (even down to 1 month or so) in the age of the animals investigated in order to obtain meaningful aging curves which closely cover the period in which a particular behavior shows a significant age-related change. This, of course, makes comparison of multiple age groups a strong experimental approach. In practice, however, this approach is often difficult (Ingram, 1996), because such studies are very expensive, extremely time consuming, and are restricted by the availability of aging and old animals.

A major problem of animal models is their validity: are similar processes and functions measured in animals and humans? Spatial discrimination tasks seem to fulfill the criteria for validity for age-related memory dysfunctions because in humans, memory for spatial information usually deteriorates with age (e.g. Light & Zelinski, 1983; Flicker et al., 1984; Liu, Gauthier & Gauthier, 1991).

It has sometimes been proposed that whether aged humans or animals have an impaired performance depends on the complexity of the task used (e.g. Doty, 1966; Goodrick, 1972; Meudell, 1983; Gower & Lamberty, 1993). In rats, it has clearly been shown that it is not task complexity *per se* that determines whether age differences are found. For instance, Stone (1929a,b) found that young and old rats performed equally well in rather complex tasks, and Soffié and Lejeune (1991) found that aged rats were slower than young rats in acquiring complex temporal discrimination tasks but the final level of memory performance was not different. A more critical variable than task complexity seems to be whether performance depends on the integration of spatial cues (Barnes, Nadel & Honig, 1980). In particular striking age differences in ”free choice” spatial discrimination tasks such as the circular platform (Barnes, 1979), the radial arm maze (Arendash et al., 1995), and the Morris water escape task (Gage, Dunnett & Björklund, 1984; Pelleymounter, Smith & Gallagher, 1987; Socci, Sanberg & Arendash, 1995) have been found.

There are other aspects that should be considered when evaluating the usefulness of tasks to study the influences of age on learning and memory. For instance, performance should not be influenced by age-related differences in sensory and motor capabilities, or by differences in motivation or problem-solving strategies used. Knowledge of what the ‘real’ age differences in learning and memory are and which behavioral variables represent ‘secondary’ effects of age that do not involve learning and memory ability is essential for a better understanding of the process of cognitive aging and of its neural correlates (van der Staay & Blokland, 1996a).

Valid interpretations of age differences found in a single task can only be made if it is certain that the age groups use the same strategy to solve the task, and that the age groups do not differ in terms of motivational or emotional factors or sensorimotor function. The use of different strains of rats may shed light on the generalizability of age effects on learning and memory (Spangler et al., 1994; Ingram, 1996). This is important when trying to unravel the neurobiological changes underlying changes in cognitive performance in aging rats and is essential for a valid extrapolation of findings to humans.

Motivation

A major point of concern in age comparison studies should be how to control for motivational differences between groups. In learning and memory tasks which use food-deprivation schedules to induce an appropriate level of motivation, animals should be differentially deprived. Goodrick (1968) advocated the use of differential deprivation schedules to control for differences in food motivation

when rats of different ages with different proportions of body fat are deprived to a comparable degree. Blokland and Raaijmakers (1993b) showed that a differential deprivation schedule could equalize the food motivation of rats of different ages. But what are the motivational factors that can affect performance in the Morris water escape task(s)? Swimming speed cannot be taken as an unbiased index of motivation because this measure might be affected by age-related motor impairments. Moreover, it is not known whether the poorer thermoregulation of aged rats (Lindner & Gribkoff, 1991) affects their motivation to escape from the water.

Similar considerations should be borne in mind when comparing animals with an experimentally induced deficit with normal control animals. Unfortunately, there is no simple way to assess the motivational level of different groups of animals because many factors, such as sensorimotor or cognitive impairments, may influence the procedures to measure the degree of motivation.

In conclusion, the aging animal as a model for the aging human might shed light on the aging process and the accompanying neurodegenerative changes (Dean et al., 1981). A particular advantage of aging studies with rodents is that the contribution of genetic and environmental factors to the aging process can be investigated and strictly controlled. Cross-sectional study designs are suitable for studying age-related behavioral changes. Unfortunately, the tasks available to assess the processes underlying the age-associated deteriorations in learning and memory of rodents in cross-sectional designs seem to be less suited to assess these processes in longitudinal studies (e.g. Beatty, Bierley & Boyd, 1985; Chapter 2.3; but see van der Staay, Krechting, Blokland & Raaijmakers, 1990). It will be a task for behavioral scientists to develop test systems which can be used to monitor animals longitudinally. More than two age groups should be included in cross-sectional studies (Spangler et al., 1994) in order to study the rate of decline of performance due to aging, because aging proceeds at a different rate for different behavioral domains (Gage et al., 1988; Gage, Dunnett & Björklund, 1989; van der Staay, Blokland & Raaijmakers, 1990; Collier & Coleman, 1991).

Chapter 3: behavioral consequences of lesioning the nucleus basalis magnocellularis (nbm) in rats

Concomitantly with the steady increase in the number of aged people as a consequence of the increased life expectancy (Martin, 1991; Holden, 1996), there is an increase in the number of patients suffering from Alzheimer's disease (Anderson, 1986). Although the activity of many neurotransmitter systems is decreased in this group of patients (Edwardson et al., 1986; Whalley, 1989), the strongest decrease has been found to occur in the cholinergic system (Jacobs & Butcher, 1986; Fibiger, 1991; Bierer et al., 1995). There is a severe loss of cells in the nucleus basalis of Meynert (nbM) in patients suffering from Alzheimer's dementia. This cell loss leads to a profound decrease in cortical cholinergic projections and the resulting cholinergic hypofunction has been suggested to be one of the major causes of the cognitive impairments found in Alzheimer's patients (Coyle, Price & DeLong, 1983; Jacobs & Butcher, 1986; Davison, 1987; Bierer et al., 1995). It has been hypothesized that animals with experimentally induced lesions of the nbm, the subprimate analogue of the nbM in humans (Smith, 1988), might mimic the neurodegenerative processes associated with Alzheimer's disease and might provide an animal model of this disease. We used rats to investigate the consequences of bilateral lesions of the nbm, induced by ibotenic acid on sensorimotor and cognitive performance.

In Chapter 3.1 we assessed whether bilateral lesioning of the nbm in rats affects spatial learning in the holeboard, which allows the simultaneous assessment of spatial WM and RM. Both the WM and RM of

the lesioned rats were impaired compared with those of intact or sham-lesioned control rats. This finding supports the notion that the nbm has a role in both WM and RM.

In a series of three experiments described in Chapter 3.2, we used the seven-choice task in an eight-arm radial alley maze, a task that is sensitive to age-associated impairments, to study the influence of experimental lesions of the nbm on spatial learning and memory performance in rats. The seven-choice task is a win-stay task which taps spatial RM. In the first experiment, bilateral lesioning of the nbm by ibotenic acid disrupted the acquisition of the seven-choice task in the radial alley maze. However, this finding could not be replicated: in the third experiment bilateral lesions of the nbm at the same set of lesion coordinates as in the first experiment had no effect on learning. A second set of lesion coordinates had also no effect.

In the second experiment, rats received nbm lesions after they had acquired the seven-choice task. We then determined the effects of the lesions on the retention, acquisition of a new problem, and re-acquisition of the originally acquired problem. Under these conditions, the nbm lesions did not affect performance in the discrimination task.

It is conceivable that lesioning of the nbm in aged rodents provides a model of Alzheimer's disease that shares more aspects of the Alzheimer symptomatology than lesioning of the nbm in young rodents would do. In order to test this hypothesis and to assess the effects of aging, of ibotenic acid-induced lesions of the nbm, and of the interaction between age and lesion, we tested young and aged Wistar rats in a battery of behavioral tests in Chapter 3.3. The battery consisted of a seven-choice task in an eight-arm radial alley maze, and a series of sensorimotor tests. We detected clear age-associated impairments in the sensorimotor tests and in the acquisition of the seven-choice task, but lesioning of the nbm did not affect the performance of the rats in the battery of sensorimotor tasks. Only a transient effect was found on the acquisition of the seven-choice task. All rats were eventually able to acquire this task, but the nbm-lesioned rats made more errors before they reached criterion. The effects of the lesion were similar in both age groups. Thus our hypothesis that aged, nbm-lesioned rats would provide a better model of Alzheimer's dementia than young nbm-lesioned rats was not confirmed.

Our results do not support the notion that the cortical cholinergic activity originating in the nbm is critically involved in memory. There are two main explanations for this unexpected finding. Either the lesion was not large enough and caused too little damage, or the task was not sensitive enough to detect lesion-induced deficits, or there was an interaction between both factors which contributed to the inconclusive findings of the present study.

Critical discussion of the nbm-lesioned rat as model for Alzheimer's disease

The nbm-lesioned rodent has been suggested as an animal model for Alzheimer's disease in that there is a similar loss of cells in the basal forebrain accompanied by a decrease in cortical cholinergic activity (Smith, 1988). However, about two decades of intense use of this model have shown that there are a number of inconsistencies and critical issues that need to be resolved.

First, when the behavioral consequences of neurotoxic or immunotoxic lesions of the nbm are compared across studies, there appears to be a discrepancy between the selectivity of the lesion to damage cholinergic projections to the cortex and the severity of behavioral impairments (Fibiger, 1991). For example, lesions of the nbm induced by quisqualic acid substantially reduce cortical ChAT activity, but cause no or only weak effects on learning and memory, whereas nbm lesions induced by ibotenic acid disrupt cognitive performance even though cortical ChAT activity is less reduced than it is

by quisqualic acid-induced lesions (e.g. Robbins et al., 1989; Steckler et al., 1993). A similar failure to affect cognitive performance with nbm lesions induced by the highly selective cholinergic immunotoxin 192 IgG-saporin (Torres, et al, 1994) questions the interpretation that the behavioral effects seen after ibotenic acid lesions are due to cholinergic denervation.

Second, the behavioral effects of ibotenic acid lesions appear to depend more on the size of the lesion, than on the degree of depletion of cortical ChAT (Dekker, Connor & Thal, 1991). The larger the lesions are, the greater the deficits induced. Moreover, it is difficult to compare studies, even those that used the same neurotoxin to produce lesions, because of the extreme variety of lesion coordinates used and of the different volumes of immunotoxins or neurotoxins injected. For example, ibotenic acid lesions of the nbm have been induced by using coordinates in the anterior-posterior orientation ranging from 1.0 mm anterior to bregma to 3.1 mm posterior to bregma, from 1.8 mm to 3.5 mm lateral from the midline, and from 6.8 mm to 7.7 mm dorsal from the brain surface. An example of strong deviations from the lesion coordinates which Wenk, Cribbs, and McCall (1984) considered as optimal to reduce cortical ChAT activity, namely AP -0.9 mm, L 2.6 mm and DV 6.8 mm, is provided by the study by Ohara and co-workers (1997). Their basal forebrain lesions, which were intended to damage the nbm in adult Wistar rats, were produced by injecting 1.5 µl ibotenic acid per side at the coordinates AP -3.1 mm, L ± 1.8, and DV 7.4 mm.

The volume of neurotoxin injected also varied between studies, ranging from 0.35 µl to 1.5 µl per side. Instead of single injections of a large volume of the neurotoxin or the immunotoxin, some investigators produced lesions by multiple injections of the toxin into the target area (e.g. Steckler, et al, 1993). This might increase the size of the resulting lesion and contribute to the heterogeneity of results reported.

It is conceivable that these different methods to lesion the nbm result in damage to a heterogeneous set of basal forebrain nuclei and to adjacent structures. Ibotenic acid lesions of the rostral globus pallidus in rats, for example, severely impaired acquisition of the Morris water escape task, whereas lesions restricted to the nbm did not affect acquisition (Meyer & Coover, 1996). Similar findings were reported by Everitt and co-workers (1987), who compared the re-acquisition of a conditional visual discrimination task by rats that had acquired the task before lesions were induced in the nbm or different areas of the globus pallidus. They found that animals with the largest lesions in the dorsal and ventral globus pallidus showed the severest cognitive deficits.

Alternatives to lesion-induced deficits

Schuurman and Traber (1989b) suggested that the old rat might serve as animal model for senile dementia. The model is based on the similarity between the behavioral symptoms of old rats and those of patients suffering from Alzheimer's disease, such as sensorimotor dysfunctions (e.g. Markowska et al., 1990), decreased social activity (e.g. Spruijt, 1991), and cognitive impairments. However, this model does not mimic the neuropathological changes underlying Alzheimer's dementia. Cholinergic hypofunction of basal forebrain nuclei and a decrease in cortical cholinergic functions probably is not a general symptom in aged rats, although reports are contradictory in this respect. For example, Fischer, Gage, and Björklund (1989) found that cell size and cell number decreased in the basal forebrain nuclei of aged rats, whereas biochemically determined cortical and hippocampal ChAT activity, which can be considered as a structural marker (Sherman & Friedman, 1990), was preserved. They found that age-related impairments in Morris water escape performance were unrelated to hippocampal and cortical ChAT activity, but were correlated with the cell sizes and number of cholinergic cells in the medial septum and with the number of cholinergic cells in the diagonal band of Broca (dbB) and the striatum (Fischer, Gage & Björklund, 1989). By contrast, Sherman and Friedman (1990) found an age-

associated increase in hippocampal and cortical ChAT activity in aged mice, whereas van der Staay (1989) found an age-related increase in ChAT activity in the hippocampus, but not in the frontal and parietal cortices of Brown Norway rats.

It is conceivable that aged rats with lesions in the nbm mimic a broader range of the behavioral deficits seen in Alzheimer's patients than do young, nbm-lesioned rats. Aged rats usually suffer from sensorimotor impairments, reduced social activity, and impaired cognitive functions. The cholinergic dysfunction that is characteristic of Alzheimer's symptomatology can be induced experimentally by lesioning the nbm in these animals. However, we did not find aged rats with nbm lesions to have increased validity as model for Alzheimer's dementia (see Chapter 3.3).

In conclusion, the role of the cholinergic projections from the basal forebrain nuclei, in particular the nbm, still has not been elucidated unambiguously. As Dunnett, Everitt, and Robbins (1991) pointed out, the hypothesis of a significant involvement of this cholinergic system in cognitive processes must satisfy two major criteria. First, the cognitive impairments must be due to disruption of cholinergic processes as opposed to non-cholinergic processes, and second, they must be due to damage in the cholinergic basal forebrain nuclei, in particular the nbm, as opposed to damage in other areas of the basal forebrain. Despite extensive research and the availability of specific tools such as 192 IgG saporin to lesion these cholinergic projections, the data are still contradictory.

Chapter 4: behavioral effects of stroke, induced by occlusion of the middle cerebral artery (MCA) in rodents

There is a strong need to investigate the behavioral consequences of stroke in animals in order to learn about the processes that lead to impairments. Stroke is one of the major causes of death in the industrialized countries (Hunter, Green & Cross, 1995), and many afflicted patients show long lasting or permanent functional impairments (Gorelick, 1995; Adams, Victor & Ropper, 1997). We performed a series of experiments with rats and mice as subjects to investigate the effects of stroke, induced by middle cerebral artery occlusion (MCA-O), on behavior.

Effects on sensorimotor functions

We assessed the effects of cerebral infarctions, induced by occlusion of the MCA, on sensorimotor functions in a series of experiments with rats as subjects. In Chapter 4.1 we tested Wistar Kyoto (WKY) rats with permanent unilateral occlusion of the MCA and sham-operated rats in a battery of simple behavioral tests 2, 16, and 37 days after surgery. In addition, we measured the motility of the animals over a period of 62 hours after the third test series. A subset of the tests appeared to be suitable to assess the effects of cerebral infarction, namely, grasping reflex of contralateral hindpaw, circling behavior, forelimb flexion, hindlimb flexion, and latency to fall off a square bridge. Except for the impaired grasping reflex of the contralateral hindpaw, there was spontaneous and complete recovery of function by the third test session, 37 days after surgery. However, some of the tests might not have been sensitive enough to detect the effects of the unilateral MCA-O on behavior and the rats were very inactive in some of the tests, so that reliable scoring of the effects was not always possible. A rat strain other than the WKY strain might be more suitable to study the behavioral consequences of MCA-O.

In order to test whether the sensorimotor impairments induced by MCA-O are strain dependent we carried out a series of three experiments with different rat strains (Chapter 4.2). We also investigated whether proximal (i.e. close to its origin) and distal occlusions (above the lenticulostriate branch) of the

MCA affected infarct volume and the behavioral impairments to a different extent, and whether there is a relation between infarct volume and behavioral deficits.

The pattern of sensorimotor malfunctions induced by proximal unilateral MCA-O appeared to be highly strain dependent. Of the eight strains tested, WISW rats, spontaneously hypertensive stroke-prone (SHR-SP) rats, and WKY rats were most severely affected. By contrast, Brown Norway (BN) rats showed only mild behavioral deficits after MCA-O. The second experiment confirmed that proximal occlusions induced slightly more behavioral malfunctions than distal occlusions did. Histological evaluation of the brain damage caused by proximal and distal MCA-O, confirmed that distal MCA-O damaged nearly exclusively cortical areas and spared the caudate/putamen. An exploratory analysis of the relation between infarct volume and behavioral deficits did not indicate that the severity of sensorimotor malfunctions can be predicted from the size of the infarct.

Effects on cognitive functions

We assessed the effects of cerebral infarctions, induced by MCA-O, on cognitive functioning in mice. In the first experiment of Chapter 4.3, male CFW1 mice acquired the standard Morris water escape task *before* half of the animals received a unilateral occlusion of the MCA. Retention was then measured in one session. In addition, the mice acquired a new platform position during daily training sessions on 4 consecutive days. In the second experiment of Chapter 4.3, naive male CFW1 mice acquired the water escape task *after* surgery. At the end of the fifth session, a probe trial was given. In both experiments the control group consisted of mice that had been sham-operated: their MCA was exposed surgically, but was left intact. Even though the infarcts induced by MCA-O covered the cranial part of the dorsomedial cortex (destroying substantial areas of the primary somatosensory cortex and smaller parts of the primary motor cortex) and part of the striatum, we did not observe marked behavioral impairments in the Morris water-escape task. Surgery *per se*, however, seemed to disrupt water-escape behavior.

The standard version of the Morris water escape task appeared to be insensitive to the effects of MCA-O. For this reason we investigated in Chapter 4.4 whether the WM version of the Morris water escape task, the repeated acquisition task, is more suited to assess MCA-O induced cognitive deficits. This task consists of trial pairs in which an animal is started twice from the same start position. Animals have mastered this task when they need less time to find the platform in the second of the two trials. Male C57BL mice were trained on this task with massed, spaced, or spaced delay trials in which there was a 90-minute delay between the first and second trials of a pair.

As soon as the mice had reached a stable baseline performance, the MCA was occluded or the mice were sham-operated and we studied the effects of the occlusion or sham operation on the re-learning of the repeated acquisition tasks. The mice trained with spaced trials learned the repeated acquisition task, whereas the mice trained with massed or spaced delay trials did not consistently show improved performance in the second trial of a pair, perhaps due to strong proactive interference in the massed trials condition or because the task was too demanding in the spaced delay condition. MCA-O hardly affected the performance during re-learning of this task, irrespective of the spacing condition of the trials, although surgery *per se* seemed to have a transient disruptive effect.

Critical discussion of the MCA occlusion induced infarct as an animal model of stroke

Rodents with strokes induced experimentally by occlusion of the major arteries in the brain are considered as models of stroke in humans. We found that the pattern of sensorimotor impairments

seen after unilateral permanent occlusion of the left MCA in rats was highly strain dependent and ranged from very mild dysfunctions to clear deficits. Only rat strains which are sensitive to the effects of infarcts are useful for the investigation of the effects of MCA-O on behavior; however, the order of sensitivity of the different strains to infarct induced dysfunctions might be affected by the behavioral tests used.

Although most of the sensorimotor deficits were observed contralateral to the afflicted hemisphere, ipsilateral dysfunctions were also seen. In Chapter 4.2 we found that the grasping reflex of the hindpaws was affected contralaterally and ipsilaterally in some of the strains tested. This effect was statistically confirmed for the SHR-SP and the WISW strains. We hypothesized that this phenomenon was due to enlargement of the ipsilateral hemisphere because of cytotoxic edema with a subsequent increase in intracerebral pressure (Persson et al., 1989) during the acute phase after infarction, i.e. the first days after MCA-O. Further support for an effect of unilateral MCA-O on the ipsilateral brain functions has recently been provided by Crespi and Pietra (1997). In *ex vivo* preparations of brain slices from frontal cortex, striatum, nucleus accumbens, and hippocampus they used differential pulse voltammetry, with treated carbon fiber micoelectrodes, to study the activity of catecholamines and 5-OH-indolamines. They found that the 5-OH-indole and the catechol levels in both the ipsilateral and the contralateral accumbens were increased, compared with those of a non-occluded sham group, 24 hours after occlusion.

Our observation of ipsilateral and contralateral disturbances of the hindpaw grasping reflex in two Wistar derived rat strains, the WKY and the SHR-SP strains (Chapter 4.2), extends the findings of Crespi and Pietra (1997). Taken together, these observations question the validity of the assumption that the brain regions contralateral to the occluded side can be considered as normal control areas. A better experimental approach would be to assess infarct-induced behavioral dysfunctions by comparing performance before and after occlusion. Behavioral tests which might be influenced by learning cannot be used in such a study design. Another appropriate, classical experimental design is the comparison of the behavior of MCA-occluded animals with that of sham-lesioned animals (e.g. Andersen, Andersen & Finger, 1991), especially, if learning processes affect successive testing scores (e.g. Yonemori et al., 1996). An intact control group should also be included in order to obtain an estimate of the effects of surgery *per se*.

Recovery of function

WKY rats showed a complete recovery of sensorimotor functions about 1 month after unilateral stroke, induced by occlusion of the left MCA, except for the grasping reflex of the contralateral hindpaw, which at that time was still slightly impaired (Chapter 4.1). Recovery of sensorimotor functions has also been observed by other investigators after permanent (e.g. Yamamoto et al., 1988; Markgraf et al., 1992), as well as transient occlusion of the MCA (Markgraf et al., 1997).

No effects of permanent MCA-O on cognitive functions

We did not find MCA-O to affect the spatial orientation performance of mice in the standard version (Chapter 4.3) or in the repeated acquisition version (Chapter 4.4) of the Morris water escape task. Brain structures, such as the hippocampal formation which are believed to be critically involved in spatial discrimination learning (Barnes, 1988b; Jarrard, 1993, 1995) are not damaged by permanent MCA-O. We did, however, see a transient effect of the operation itself, which involves craniotomy (Bederson et al., 1986; Welsh et al., 1987).

The central question, of course, is whether the MCA-occluded rodent provides a model for the stroke-related functional deficits seen in patients. According to Caplan (1995, p. 4-5), animal models of stroke suffer from a number of shortcomings:

- First, they are only poor replica of the situation in stroke patients. For example, most patients with MCA-O suffer from long lasting behavioral impairment (Adams, Victor & Ropper, 1997), although spontaneous recovery has been observed in patients after acute ischemic stroke (Biller et al., 1990). By contrast, we found nearly complete recovery of sensorimotor abilities in WKY rats after MCA-O, an observation that is in line with the results of other animal studies (e.g. Markgraf et al., 1992).
- Second, the brains and the craniocerebral vascular systems of rodents are different from those of humans. In fact, there are considerable differences between species and even between strains within species (e.g. van der Staay, Augstein & Horváth, 1996b, Chapter 4.2). Caplan's objection is probably justified if only one experimental approach is considered. However, nowadays the approach is to study the effects of experimental manipulations of a brain structure in two or more species (i.e. a 'comparative approach', including humans, if possible) in order to try to *generalize* about brain structures, functions, behavior, and how they are related (Isaacson et al., 1971, p. 3).
- Third, the techniques used to induce strokes in animal studies do not model the events which cause infarcts in humans. Most experimental techniques produce an acute infarct, whereas clinical insults often develop more gradually. Patients suffering from ischemic stroke often have a history of multifocal artherosclerosis which has developed over weeks to years (Wiebers, Adams & Whisnant, 1990). Moreover, a variety of risk factors might predispose an individual to stroke, such as genetic factors, chronic hypertension, diabetes, cigarette smoking (Wiebers, Adams & Whisnant, 1990), or hypercholesterolemia (Caplan, 1995). These factors affect the vascular system before stroke and also are important co-determinants of the outcome of post-stroke treatment. To the best of my knowledge, no occlusion technique has been developed which mimics the gradual development of a thrombus over a period of hours to weeks. However, the combination of different experimental approaches to model deficiencies, such as old age and hypertension, might be one way to improve the similarity of the conditions mimicked in the animal model and those seen in humans.
- Fourth, it is not possible to assess the disrupting effects of stroke on higher functions (sensorimotor, cognition). This objection against the use of animal models of stroke to study infarct-induced dysfunctions might be valid with respect to cognitive functioning. However, several groups world-wide are putting considerable effort into the development of appropriate test systems. Typical human abilities such as speech, which is often impaired in patients suffering from stroke (e.g. Adams, Victor & Ropper, 1997), cannot be mimicked in animal models (Dawson, Heyes & Iversen, 1992). Moreover, it is not clear whether those tests that are normally used to assess behavioral dysfunction in rodents are comparable to the tests which are used to assess functional decline in non-human primates and humans (Gallagher, 1993). However, if one takes into account the species-specific behavioral repertoire (e.g. Bolles, 1975; Willner, 1991), valid behavioral tests can be developed for a broad range of cognitive processes. Although some of these tests, at first sight, might lack similarity to tests used for humans, they might very well tap similar functions and processes.

A number of similarities and differences are apparent when comparing the deficits seen after MCA-O in rodents with those typically seen in patients after MCA-O (Adams, Victor & Ropper, 1997, p. 790; cited in Chapter 4.0). In both humans and rodents, occlusion of the MCA gives rise to sensorimotor

deficits contralateral to the infarct. However, the deficits in rats and mice (e.g. Klapdor-Dulfer, 1996) appear to be mild when compared with those seen in patients. Moreover, most sensorimotor deficits in rodents are transient and disappear within a few weeks after stroke. This contrasts with the finding that approximately half of the patients who survive a stroke show persistent neurological impairments (Gorelick, 1995). Cognitive impairments, a major symptom in stroke patients (Adams, Victor & Ropper, 1997), were virtually absent in our experiments (see Chapters 3.2, and 3.3).

According to Hunter, Green, and Cross (1995) it can be doubted whether permanent interruption of the blood flow in the core of the infarct, as produced by permanent focal MCA-O, has relevance for most strokes caused by occlusion of the MCA in patients. In stroke patients, the thrombus usually disintegrates, and reperfusion of the infarcted area occurs. Permanent focal ischemia after MCA-O usually does not occur in the hippocampus. This brain structure, and particularly the CA1 region (Kuroiwa, Bonnekoh & Hossmann, 1991; Block & Schwarz, 1996; Nelson et al., 1997), seems to play a prominent role in spatial orientation (Jarrard, 1993, 1995). Therefore, it is more likely that deficits in (spatial) learning and memory are found in reperfusion models of ischemic stroke, such as carotis occlusion models (Puurunen et al., 1997) or four-vessel occlusion (Block & Schwarz, 1996; Nelson et al., 1997) models, where the hippocampus is damaged, than after permanent occlusion of the MCA.

To summarize, permanent occlusion of the MCA-O is a well established method to cause reproducible brain damage in rodents. MCA-occluded rats are useful to assess the processes leading to stroke-induced damage and to evaluate the effects of putative neuroprotective compounds on the size and volume of the infarct (e.g. Gotti, et al, 1990; Mahadik & Wakade, 1992; Matsumoto et al., 1996; Ren & Finklestein, 1997). However, transient ischemic insults of the MCA-O (e.g. Sakai et al., 1996; Zhao et al., 1996; Belayev et al., 1996) or the damage caused by combinations of different (permanent and/or transient) occlusion techniques (e.g. Netto et al., 1993; see also Chapter 4.0, Table 1) might be a better model to assess the consequences of brain infarcts in humans, i.e. complex, long-lasting and severe behavioral dysfunctions. More studies are needed before clear recommendations can be given about the appropriate stroke model(s) to investigate acute functional deficits and their long-term recovery.

Animal models: a definition

The concept of animal models is controversial (McKinney, 1984), and consequently those working with these models should try to define this concept properly. McKinney (1984) gives the following definition: *"Basically, animal models represent experimental preparations developed in one species for the purpose of studying phenomena occurring in another species."* (p. 77). In the General Introduction (Chapter 1), animal models were defined as living experimental systems used to analyze brain-behavior relations under controlled conditions. In my opinion this definition can be extended to:

> *An animal model in the behavioral neurosciences is a living organism used to study brain-behavior relations under controlled conditions, with the final goal to gain insight into these relations in humans and/or a species other than the one studied, or in the same species under conditions different from those under which the study was performed.*

Kaplan (1973) considered that a model and what it is supposed to model are isomorphs of one another. The degree of correspondence should be defined, and the qualification 'animal model' can

only be given if clear, predefined criteria are met. This does not mean that a model in development must fulfill all criteria immediately; it should be possible to refine the model and to continuously re-evaluate it, using a set of generally accepted criteria. Unfortunately, this set of generally accepted criteria does not yet exist. At this point, experts from many scientific disciplines should take up the challenge and try to set up an evaluation system for animal models. For deficiency models, this evaluation involves input from different, highly specialized disciplines.

From the above definition it follows that a scientist who is interested in, for example, vision in pigeons, is not working with an animal model when he or she assesses the performance of pigeons in visual tasks. However, if he or she is trying to elucidate processes involved in vision in the pigeon to gain insight into vision in humans (or species other than pigeons), then he or she is using an animal model. Thus, the *purpose* of the investigation directly determines whether a model is used or not.

McKinney (1984) discriminates four kinds of animal models, depending on their primary purpose: first, there are models designed to parallel a specific sign or symptom of the human disorder; second, there are models designed to evaluate a specific etiological theory; third, there are models designed to study underlying mechanisms or processes; and fourth, there are models designed to permit the preclinical evaluation of putative therapeutics. According to Wiebers, Adams, and Whisnant (1990), the primary goal of all animal research on behavioral impairments should be to improve health-care beyond the laboratory. In this respect, the ultimate clinical outcome is the only test of relevance to determine the value of the animal models used.

The concept of validity

In the following discussion, the term *animal model* refers to *animal model of (pathological) behavioral deficiencies*. Validity is defined as *"(..) the agreement between a test score or measure and the quality it is believed to measure."* (Kaplan & Saccuzzo, 1997, p. 131). Willner (1986), in an influential paper on the concept of animals models, specifically for depression, stated that animal models should possess face validity, predictive validity, and construct validity

Face validity, or phenomenological validity

According to McKinney and Bunney (1969) an animal model should at least meet the requirement that it *resembles* the condition to be modeled with respect to its etiology, its symptomatology, its underlying processes, and its treatment, i.e. it should possess face validity. Kaplan and Saccuzzo (1997) consider face validity to be a concept of limited value because it simply states that a measure is meaningful by itself. No attempts are made to generalize. In fact, considering that rodents have their own species-specific behavioral repertoire in order to survive in their habitat (see, for example, Bolles' concept of *species specific defense reaction*; Bolles, 1975, p. 190), the resemblance between their behavior and that of humans might be weak or absent, even though similar underlying processes might guide their behaviors.

Predictive validity

A test with high predictive validity makes it possible to venture a sound prognosis (Lienert, 1969) or to forecast, for example, future behavior (Kaplan & Saccuzzo, 1997). Predictions can be based on purely empirical evidence, without any understanding of the behavior involved (Silva, 1993). Understanding should be based on an underlying (psychological) construct. Unfortunately both models of dementias

(Altman, Gershon & Normile, 1991) and models of stroke (Caplan, 1995) appear to lack preclinical to clinical predictability.

Construct validity

"A construct is defined as something constructed by mental synthesis." (Kaplan & Saccuzzo, 1997, p. 143). Construct validity refers to the theoretical clarification of what a test measures (Lienert, 1961). Animal models possess construct validity if their procedures are theoretically sound. The construct validity is not established by determining the relation between a test and an accepted criterion. Instead, it is based on the establishment of relationships which are in turn based on the definition of a *trait*. Implicitly, a construct is defined by a network of associations (Cronbach & Meehl, 1955; Runkel & McGrath, 1972, pp. 162-163).

Ellenbroek and Cools (1990) considered predictive validity, face validity, and construct validity, in that order, as a hierarchy of categories of validity, where construct validity is the highest category. In accordance with Kaplan and Saccuzzo (1997), I suggest a slightly different hierarchical order: face validity, predictive validity, and construct validity, with construct validity again as highest category.

Face validity is at the *naive* level: i.e. the test *looks like* it is valid, because of the perceived resemblance between the model and the situation or process to be modeled. Predictive validity is at the *empirical* level, i.e. data show that the outcome obtained in the model has some predictive value for the situation or process to be modeled. Finally, construct validity is at the *theoretical* level. Constructs such as *learning ability* or *memory* have no counterparts in the realm of observables. Instead, they define a framework of theoretically relevant relations (Silva, 1993; Kaplan & Saccuzzo, 1997).

The first concern of any researcher working with animal models should be to determine, and if necessary and possible, to improve the construct validity of these models. Improvement of construct validity helps to improve predictive validity! Construct validity is the most important aspect of validity as far as animal models are considered.

Determination of the construct validity of deficiency models

Valid animal models of behavioral deficiencies are needed, because they make explicit the assumptions about the underlying (pathological) processes and about the mechanism of action of putative therapeutic compounds, and thus help the experimenter to perform meaningful experiments. In order to determine the validity of deficiency models, the measures, dependent variables, or 'read-outs' provided by the model must be evaluated with respect to two different aspects of the model.

The first aspect concerns the (pathological) changes in the animal which are supposed to underlie the defective behavior. These changes might be experimentally induced, e.g. by lesioning particular brain regions, or might occur naturally, e.g. as a consequence of normal aging (Gamzu, 1985). A main question is whether the damage seen in the animal model mimics the damage seen in the disease to be modeled. For example, questions such as whether the reduction in ChAT activity in the cortex of nbm-lesioned rats mimics the cholinergic hypofunction in Alzheimer patients, or whether aged rats can be considered as a model for Alzheimer's disease, as suggested by Schuurman and colleagues (Schuurman et al., 1986, Schuurman & Traber, 1989b) must be discussed.

In lesion and stroke models, relevant aspects are the site and size of the lesion, effects on specific neuronal circuits, and neurotransmitter systems, etc. which can be measured by using appropriate histological, biochemical or imaging techniques. Although the damage or pathological changes induced should be as similar as possible to those found in humans, pathologists, clinicians, molecular biologists, etc. should define which criteria a model must meet to be considered as a valid representation of dysfunctions or deficits seen in humans (see Table 1). Unfortunately, a problem that severely hampers the definition of criteria for deficiency models is that the etiology of many neurodegenerative diseases, and even the processes underlying normal aging, are poorly understood.

It might sometimes be better not to model the full-blown symptomatology but to isolate certain aspects of it (McKinney, 1984). For example, one might want to gain insight into the role of degeneration of the nbM in demented patients by studying the effects of lesioning of the nbm in rodents. Although nbm lesions do not induce the full pathology of Alzheimer's disease (Decker, 1995), they produce severe cell loss in the main cholinergic projections to the cortex, similar to that found in Alzheimer's patients (Coyle, Price & DeLong, 1983; Jacobs & Butcher, 1986; Davison, 1987).

Table 1. *The independent and the dependent variables in deficiency models. The validity of a model can only be determined in a multidisciplinary approach. No explicit set of rules exists for the neuropathological changes (second column) which are considered to be the cause of the behavioral dysfunctions. By contrast, a highly formalized set of rules exist for the behavioral changes (third column).*

Independent variable (subject)	**Dependent variable**	
	Neuropathological changes	Behavioral changes
Aged animal, lesioned animal, ischemic animal, hypoxic animal, aged and lesioned animal, etc. (see Chapter 1, Table 1)	Damage induced: site, size, effects on specific neuronal circuits, neurotransmitter systems, etc.	Behavioral dysfunction or malfunction: impaired cognitive performance, impaired sensorimotor functions, etc.
	Homology of damaged area(s) or neuropathological changes.	Homology of disrupted processes or impaired functions
	Expertise: pathologists, clinicians, molecular biologists, etc., depending on which aspects of the animal model are considered	Expertise: behavioral scientists such as (comparative and physiological) psychologists, ethologists
	Experts should define as exactly as possible which criteria the model must meet	Concepts of reliability and validity from psychological test theory

The second aspect of the models concerns the behavioral changes they induce. The validity and reliability of this aspect of animal models can be evaluated by making use of procedures and concepts from testing psychology. These concepts are well established and extensively described in a number of textbooks (e.g. Lienert, 1961; Runkel & McGrath, 1972; Kaplan & Saccuzzo, 1997; see Table 1).

Some of the consequences of using models of restricted validity are that results cannot properly be interpreted, that scientific progress is retarded, and that animals are used unnecessarily. In this context, the bias of journals in favor of hypothesis-confirming results might be a reason for the slow progress in the development of new animal models and their validation. Negative results remain

unpublished, and poor concepts, hypotheses, and models survive, notwithstanding a vast amount of contradictory data, merely because these data are not made available to the scientific community. Thus, two problems emerge:

- By not publishing negative results, either because the experimenter does not want to, or because of the bias of scientific journals in favor of hypothesis-supporting results, important information is not available and cannot be considered when evaluating an animal model.
- Consequently, more researchers will try to replicate the results. Their experiments could have been *avoided* and animals could have been spared if an evaluation could have been made based on all relevant information, which, of course, includes information about the weaknesses of models and the failures to replicate findings reported by others.

Unfortunately, an aspect of animal models that is hardly ever addressed concerns their *replicability* (see van der Staay, 1997; Chapter 2.4). D'Mello and Steckler, 1996, p. 351 when listing the features of an ideal animal model of human cognitive function (see below), stated that all test conditions should be replicated, if practicable. I would go a little further:

Results are preliminary as long as they have not been corroborated, and preferably by investigators other than those who originally performed the investigations.

In this context, notions such as that it should be avoided to do the same experiments in different laboratories, because it is a waste of animals, must be re-considered.

Reproducibility versus generalizability of results

One of the main purposes of scientific experimentation is to standardize the experimental protocol in such a way that the outcomes are reproducible and that the variation due to putative sources of error is minimized (Runkel & McGrath, 1972). A problem inherent in this approach is that one cannot be sure that the relations found are also valid under different (testing) conditions, i.e. whether the findings can be generalized (Kaplan & Saccuzzo, 1997, p.142).

This dilemma can be overcome by using different methods and different experimental subjects when investigating a scientific question. One might, for example, conclude that a certain brain-behavior relationship is a general phenomenon, if similar results are obtained for different lines of a particular species, or when the findings are consistent across species (Isaacson, et al, 1971).

The replicability of results can be increased if inbred lines of rodents or the F_1-hybrids of crosses between two inbred lines are used to reduce error variation (see the section on this topic below). Inbred strains are produced by mating full siblings for at least 20 generations. At this point the probability that all individuals within the line are homozygous for the same allele asymptotically approaches 100%, (Plomin, DeFries & McClearn, 1980). Unfortunately, lines of the same inbred strain, maintained by different breeders, have been found to differ considerably with respect to genetic markers (van Zutphen & den Bieman, 1984). Reproducibility would be expected to be highest if the animals used for experiments come from the same stock, but profound cohort differences have been observed even within inbred lines (e.g. Mos & Hollander, 1987, with respect of the longevity of rats). Permanent quality control is needed to ensure genetic stability in rodent strains.

An alternative approach to increase reproducibility and generalizability, but not necessarily intra-group variability, is to use samples from a 'heterogeneous stock' or from a 'mosaic population' (van Zutphen, 1993). A heterogeneous stock consists of the F_1-hybrids from crossings between a selected number of inbred strains. In a 'mosaic population', inbred animals are also included. Because these stocks are based on inbred strains, they can be reconstructed whenever they are needed, as long as the inbred strains used are available (van Zutphen, 1993). Unfortunately, standardized samples from heterogeneous stocks or from mosaic populations are not readily available.

Deficiency models

The animal models which have been proposed for the study of behavioral dysfunctions can be classified into two main groups: those using normal subjects and those using subjects with behavioral deficits (see Chapter 1). The second group can further be subdivided into models which are based on naturally occurring deficits or dysfunctions, and into models in which deficits or dysfunctions are induced experimentally.

It is not my intention to discuss all the different classes of deficiency models in detail. However, the use of normal subjects, genetic strains, including the quantitative genetic approaches based on these strains, selected extremes from a particular animal population, e.g. good vs. poor learners as examples of models which are based on naturally occurring deficits, and transgenic animals, knockouts, as examples of experimenter-induced deficits, will briefly be discussed. The models of the old animal, CNS-specific lesions, exemplified by ibotenic acid lesions of the nbm, and stroke, exemplified by MCA-O induced ischemia, have already been discussed above.

Normal animals

If one uses *normal subjects*, then one implicitly assumes that these normal subjects function suboptimally and that there is room for improvement. Moreover, if a compound is active in normal subjects, then it is to be assumed to be useful for the treatment of patients (Gamzu, 1985). This model is used rather frequently. To give a few examples, Cook et al. (1990), using young intact rats as subjects, assessed the effects of the putative cognition enhancer Linopirdine (DuP 996), an acetylcholine releaser, in learning paradigms. The effects of the putative cognition enhancer metrifonate, a compound that is transformed nonenzymatically to dichlorvos, a cholinesterase (ChE) inhibitor, was investigated in young, unimpaired rats in the standard Morris water escape task by van der Staay, Hinz, and Schmidt (1996a,b). The effects of nicotine on working and RM performance of young rats in a 16-arm radial maze task were evaluated by Levin, Kaplan, and Boardman (1997). In these three studies, the compounds tested improved cognitive performance, and the authors of these papers concluded that the respective compounds might provide a useful treatment for cognitive dysfunctions in humans.

Normal subjects as models for behavioral dysfunctions can only possess predictive validity. This model fulfills neither the criteria for face validity nor those for construct validity. Compounds which are able to improve cognitive functions in normal subjects might be considered as 'cognition enhancers', but not primarily as 'disease modifiers'. However, an alternative view might be that so-called normal subjects

function suboptimally, or that they suffer from undetected deficiencies which can be antagonized by the putative cognition-enhancing compound.

Genetic lines

Appropriate lines of rats and mice can be selected from the enormous genetic pool provided by the various inbred strains (Altman & Katz, 1979; Festing, 1980; Crawley et al., 1997). According to Russell (1972), inbred strains and F_1-hybrids possess a number of advantages that make them valuable for the study of behavioral dysfunctions:

- First, genetically different lines provide controlled differences for experimental designs (Hazzard et al., 1992; Ingram, 1996).
- Secondly, such strains increase the reproducibility and predictability of results as a consequence of the reproducibility of individuals of a specific strain. F_1 hybrids from crosses between inbred strains possess the advantages of the inbred strains used, but are less variable, i.e. show greater biological uniformity, than the parental inbred strains do (Phelan & Austad, 1994). The inbred strains and the first filial (F_1) generation(s) of crosses between inbred strains provide groups which consist of genetically identical individuals. The variation *within* inbreds or F_1s can be considered to mirror error variation, provided the animals with the same genotype are housed in a highly standardized environment.
- Thirdly, well-defined and characterized inbred strains can be selected as raw material for quantitative genetic studies. These quantitative genetic studies range from strain comparison studies (e.g. van der Staay, Kerbusch & Raaijmakers, 1990; van der Staay & Blokland, 1996a) to classical Mendelian cross-breeding (e.g. Kerbusch, van der Staay & Hendriks, 1981) and diallel cross-breeding studies (e.g. Kerbusch, 1974; Crusio, 1993; Crusio & Schmitt, 1998).

  A number of association-based approaches have been developed, such as genetic correlations (e.g. van der Staay, Kerbusch & Raaijmakers, 1990) and techniques to identify *quantitative trait loci* (Lander & Botstein, 1989; Buck et al., 1997). Especially, if the environmental effects are small and the number of homogeneous groups (and subjects within groups) is high, the correlation between measures across genotypes approaches the genetical correlation.

Studies using these approaches can contribute to our understanding because they provide information about the genetic background of the measures under investigation, allow a closer examination of the relation between neurobiological and psychological factors, and provide information about the generalizability of results (Hazzard et al., 1992).

Quantitative genetic approaches to study normal and dysfunctional behavior

At least part of normal as well as abnormal behavior appears to be under genetic control. Some decades ago, predominantly in the seventies, the mere mention of ‘genetic control of behavior’ to social scientists would immediately have provoked a discussion about ‘nature and nurture’, and the role of genetic factors would have been minimized, or explained away, or, in the most extreme case, be neglected all together. We now know that genetic factors play an important role in the control of behavior, although we are far from understanding *how*.

What is the role of (animal) behavior genetics for our understanding of behavior?

Normal behavior and behavior disorders are extremely complex phenotypes. Behavior mirrors the function of a complex system, the whole organism. Because behavior shows an extreme variability among individuals, it is unlikely that a single gene governs behavior (Oliverio, Cabib & Puglisi-Allegra, 1992).

Many illnesses, for example psychiatric disorders, are considered as extremes of a continuum of behavior. Traits, such as cognition, anxiety, depression, drug dependence appear to be normally distributed in the population, and the distinction between normal and 'abnormal' or clinically conspicuous is a question of convention (Flint & Corley, 1996). This raises an aspect that is potentially of utmost importance for unraveling the genetic basis of diseases, namely, the definition of the heritable phenotype. This definition is central to the identification of genes which cause a disease, or which increase the susceptibility to development of a disease (Smoller & Tsuang, 1998). This aspect might even be more important for animal studies designed to unravel the genetic basis of psychiatric disorders than it is for quantitative genetic approaches in human populations.

The phenotype, abnormal behavior or behavioral dysfunctions, as described by, for example, the *Diagnostic and statistical manual of mental disorders IV* (American Psychiatric Association, 1994), must be translated into testable measures in animal experiments. Unfortunately, the description of the psychiatric symptomatology, the psychiatric nosology, is not easily translated into behaviors which can be defined and operationalized in animal models, because it is atheoretical and descriptive, and because it suffers from a certain degree of arbitrariness. Moreover, this nosology is not based on pathogenetic mechanisms (Smoller & Tsuang, 1998), and different diagnostic systems might further complicate the search for genetic factors underlying behavioral dysfunctions (Oliverio, Cabib & Pulgisi-Allegra, 1992).

On the basis of the idea that most mental disorders are an expression of the extreme of the distribution of a limited number of underlying dimensions, one way to solve this dilemma is to break the symptomatology down into elemental phenotypes which can be tested in both human populations and in animal studies.

Quantitative genetic approaches

Strain comparisons: strain comparison studies are the most simple way to study the contribution of genetic factors to a particular phenotype. This method compares the phenotype under study across a number of highly inbred rodent strains. The environment in which the strains are bred and kept must be highly standardized, as must be the testing environment (Andrews, 1996) and the age at which testing is performed (Meier, 1964). Under these conditions, a significant proportion of the variation between strains can be ascribed to genetic variation. Comparison of inbred strains of rodents provides a useful tool to study the genetic basis of behavior (Paylor & Crawley, 1997).

In order to assess behavioral deficits which occur naturally, for example due to aging, or which are induced experimentally, for example by brain lesions, it is necessary to select animals of a particular inbred strain or F_1 hybrid on the basis of their behavior as intact, young animals. The level of performance of the young, intact animals largely determines the range of behavioral deficiencies that can be measured. For example, a complicating factor when choosing a particular inbred strain or F_1 hybrid for aging research is that there are strong strain differences in learning and memory, even in young-adult to adult animals. The poor performance of aged inbred BN rats and FBNF1 hybrid rats in shock-motivated tasks, such as the inhibitory avoidance or the active avoidance tasks, cannot be

considered as an adverse consequence of normal aging or of the occurrence of age-related pathological changes (van der Staay & Blokland, 1996a) because these genotypes already show these deficiencies at a young age. Thus, it does not make sense to use genotypes such as the BN or the FBNF1 hybrid in shock-motivated tasks as animal models for aging in gerontological and gerontopharmacological research.

Selection studies: selection for a particular phenotype in a heterogeneous population is another method for demonstrating the role of genetic factors in the expression of the phenotype. In most cases, selection experiments try to produce two lines: one which shows a strong expression of the phenotype, and one that shows a weak expression of the phenotype. Both lines should deviate from a control line, in which no selection occurs. Replicates of the selection line are needed to distinguish the response to selection from that due to changes caused by, for example, inbreeding, or to random genetic drift (Holmes & Hastings, 1995).

Classical Mendelian cross-breeding studies: the classical Mendelian cross-breeding design starts with the selection of two highly inbred rodent strains, which should be as different as possible with respect to the phenotype being studied. From these two parental lines, F_1 and F_2 generations are derived. In addition, in most studies the back-crosses from the F_1 generation to the two inbred lines are also produced (see Figure 1). This design allows a very detailed genetic analysis and the detection of very small gene effects. However, the generalizability of this approach is restricted because all genetic material is derived from only two inbred strains (e.g. Kerbusch, Hendriks & van der Staay, 1989).

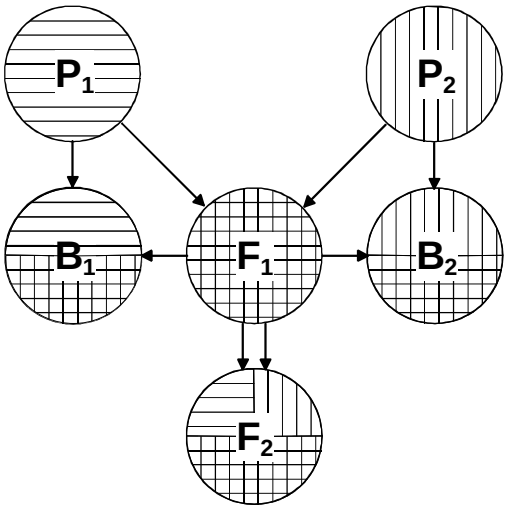

Figure 1. *Schematic representation of a classical cross-breeding design. The P_1 and P_2 should be as different as possible with respect to the phenotype under inverstigation.*
P_1: first parental (inbred) strain
P_2: second parental (inbred) strain
F_1: first filial generation (cross between P_1 and P_2)
F_2: second filial generation (F_1s intercrossed)
B_1: first back-cross (cross between P_1 and F_1)
B_2: second back-cross (cross between P_2 and F_1)

Diallel crosses: in the diallel cross-breeding design, a number of highly inbred rodent strains are crossed in all possible combinations (see Figure 2). The results obtained from the diallel cross-breeding approach are much more generalizable than those obtained from classial Mendelian cross-breeding approaches, but the information obtained about the genetic architecture underlying the phenotype studied is less detailed (Crusio & Schmitt, 1998).

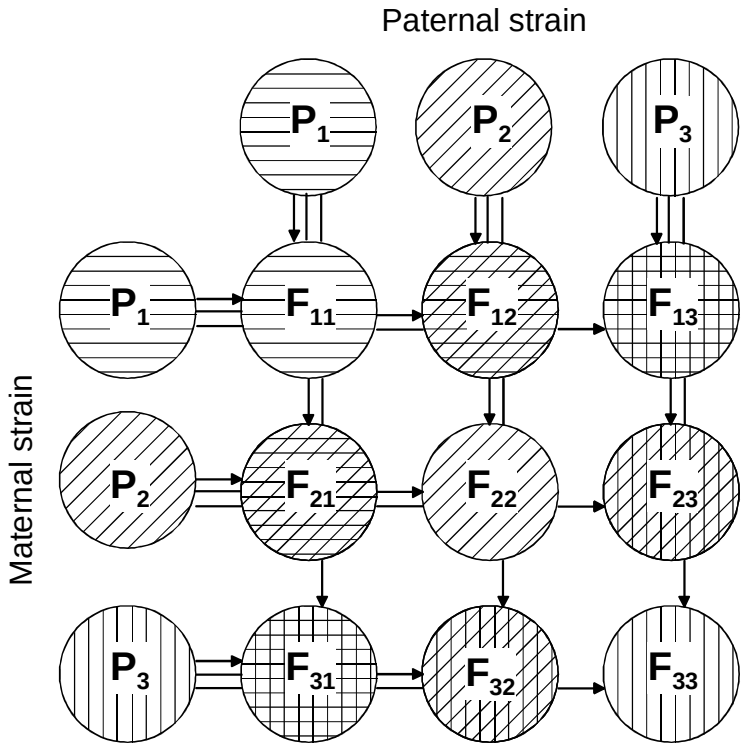

Figure 2. *Schematic representation of a diallel cross-breeding design.*

- P_1: *first parental (inbred) strain*
- P_2: *second parental (imbred) strain*
- P_3: *third parental (inbred) strain*
- F_{12}: *first filial generation of cross between P_1 and P_2; note, that the mother is from P_1 and the father is from P_2*
- F_{21}: *reciprocal of F_{12}; now the mother is from P_2 and the father is from P_1*

Recombinant inbred strains: recombinant inbred strains are derived by crossing two highly inbred, but genetically unrelated strains, and by crossing the F_1 generation to obtain the segregating F_2 generation (see Figure 3). From this generation onward, a series of inbred strains is obtained by at least 20 generations of brother-sister mating. The chance recombination of the genes becomes fixed in the resulting battery of inbred strains, creating a replicable recombinant population (Oliverio et al., 1992).

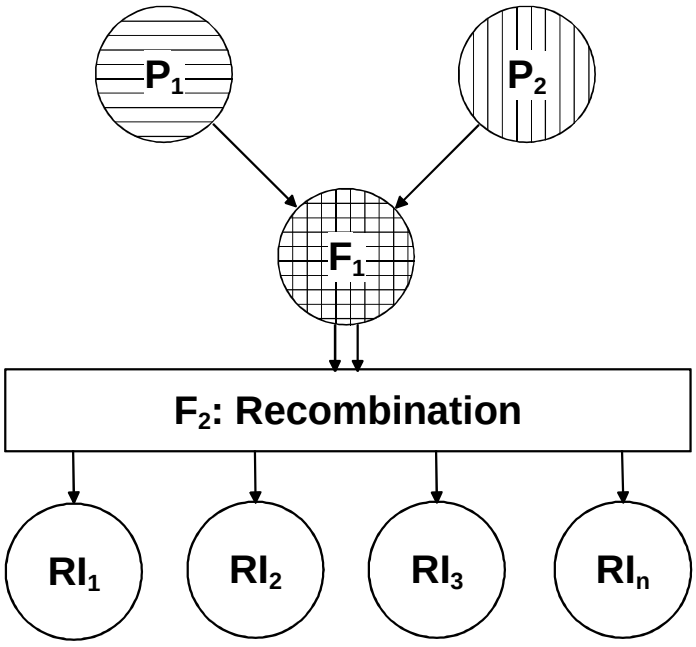

Figure 3. *Schematic representation of a cross-breeding schedule to establish recombinant inbred strains.*

- P_1: *first parental (inbred) strain*
- P_2: *second parental (inbred) strain*
- F_1: *first filial generation of cross between P_1 and P_2*
- RI_1: *first recombinant inbred strain*
- RI_n: *n^{th} recombinant inbred strain*

Quantitative trait loci (QTL): the identification of quantitative trait loci is used as a first step to identify candidate genes for quantitative traits (Rikke, & Johnson, 1998). QTLs can be estimated from recombinant inbred strains or from the second filial generation of two inbred strains which represent extremes of a continuum of the trait under investigation. The conserved chromosomal regions between mice and humans have been mapped and a genetic marker map containing about 6100 dinucleotide markers has been constructed which densely covers the entire genome (Gershenfeld & Paul, 1998). The map of conserved chromosomal regions makes it possible to crudely delimit the chromosomal regions where trait-relevant genes might be located in humans. Unfortunately, because singles genes normally appear to have little effect on the trait under study, it is not very likely that small studies of human pedigrees or affected siblings will confirm QTLs identified in rodent studies (Gershenfeld & Paul, 1997).

Extremes from a population

Several researchers have reported that, in a sample of aged animals tested in learning and memory tasks, some individuals appear to be unimpaired, performing as well as their young counterparts, whereas other individuals show severely impaired performance (e.g. Gallagher, Burwell & Burchinal, 1993; Caprioli, Markowska & Olton, 1995; Quirion et al., 1995; Baxter & Gallagher, 1996; Rasmussen et al., 1996). Caprioli and colleagues (1995) argue that testing the effects of putative therapeutic interventions to improve behavioral deficits in unimpaired individuals decreases the sensitivity of the test system, because unimpaired individuals may not respond to the treatment (but see the paragraph on normal animals, p. 227). They will, however, increase the variability of the sample. Therefore, elimination of unimpaired animals can reduce the variability in the sample, increase the sensitivity of the test, and prevent the waste of resources.

However, Ingram (1996) is concerned about this splitting up of a sample of aged animals into impaired and unimpaired groups. First, the measures used to characterize the behavioral deficits induced by aging may not be strong enough predictors of the effects of aging. Instead of subdividing a sample of aged rats into subsamples of impaired and unimpaired individuals, it may make more sense to improve the age-sensitivity of the behavioral measures used. Second, subdivisions may be based on non-cognitive factors (e.g. differences in the visual ability of animals performing a visually guided discrimination) or pathological differences between individuals, rather than on differences in cognitive abilities. Third, the identification of an impaired subgroup in one cognition test does not necessarily mean that this group will suffer from impairments in another test.

Transgenic and knockout animals

Quantitative genetic studies have already provided convincing evidence for the importance of both environmental and genetic factors in the regulation of behavior (Oliverio, Cabib & Puglisi-Allegra, 1992). But nowadays, to put it provocatively, quantitative genetic approaches in the behavioral sciences, such as strain comparisons or cross-breeding studies, are ‘out of fashion’, they are ‘dinosaurs’ which have not survived because they are extremely time-consuming and expensive, and because new approaches have completely replaced them. These new approaches emerged from molecular biology. The development of highly sophisticated molecular biological approaches will make it possible to answer the question *how* genetic factors control behavior. These techniques are being

rapidly developed and have led to models which allow the investigation of the genetic control of behavior more exactly than was earlier possible with quantitative genetic methods.

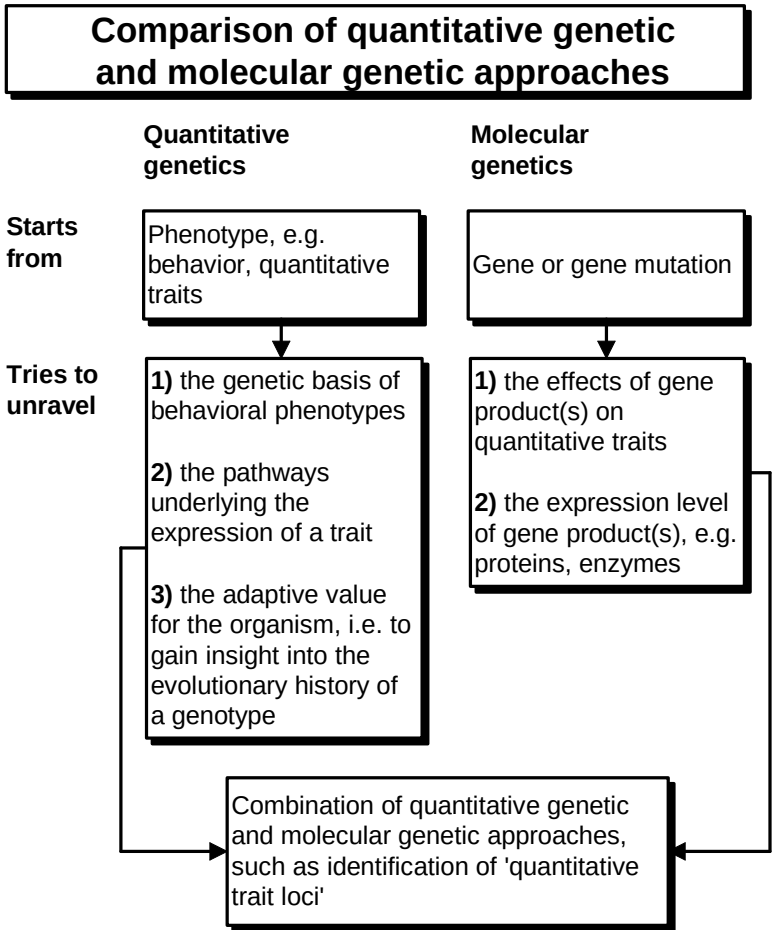

Figure 4. *Comparison of quantitative genetic and molecular genetic approaches to unravel the genetic basis of normal and dysfunctional behavior.*

It can be questioned whether quantitative genetic approaches still have a place in the behavioral sciences. To answer this question it is necessary to compare these two approaches (see also Figure 4).

- Quantitative genetics starts from the phenotype to deduce the underlying genetic factors. The basis of quantitative genetics is that Mendelian rules of discrete inheritance can be extended to study the inheritance of complex characteristics that show a continuous phenotypic distribution (McClearn & DeFries, 1973). The continuous distribution of a trait is the result of the sum of the small genetic influence of different genes, each of which is transmitted in a Mendelian fashion, i.e. quantitative traits are polygenically controlled. Quantitative genetic approaches can be used to estimate the number and interaction of genes involved in a quantitative trait, such as behavior, but they cannot provide any information about the underlying genes (Oliverio, Cabib & Puglisi-Allegra, 1992). Quantitative genetic approaches also provide insight into the evolutionary history of a phenotype, and into brain-behavior relationships (Crusio & Schmitt, 1998).

- Molecular genetic approaches can be considered as 'reverse genetics', in that they start from a detailed knowledge of the gene under investigation, which allows the experimental manipulation of its expression and function. This in turn provides a powerful tool to investigate the physiological relevance of the gene product, usually a protein (Picciotto & Wickman, 1998). Thus, molecular biology starts from the genotype, a single gene, and deduces the gene product and, ultimately, the effects on the phenotype, i.e. behavior.

Molecular biological approaches: transgenics, knockouts, and antisense oligonucleotides

In the last few years, the antisense oligonucleotide technique and a number of gene-targeting, or knockout, models and transgenic models have been developed which are expected to provide new insight into brain-behavior relations and into the neuropathological processes underlying neurodegenerative disorders such as Alzheimer's disease (Higgins & Cordell, 1995; Duff, 1997; Price et al., 1998). These are at present nearly exclusively murine models (Brandon, Idzerda & McKnight, 1995a,b,c; Jucker & Ingram, 1997). Many recent reviews have addressed the experimental techniques used to produce transgenic and knockout mice and antisense oligonucleotides (e.g. Campbell & Gold, 1996; Picciotto & Wickman, 1998).

In short, *transgenic* animals express a mutant gene or overexpress a wild-type gene that has been introduced into the genome of a rodent strain. Trangenic animals might model gain-of-function disorders (Picciotto & Wickman, 1998). For example, transgenic mice that overexpress the amyloid-precursor protein detected in individuals with familial Alzheimer disease mutations are used as a model of Alzheimer's disease (Sommer, 1998; Riekkinen, Schmidt & van der Staay, 1998).

In *knockout* mice, an endogenous gene is inactivated, or a mutant protein is made non-functional (Campbell & Gold, 1996). Knockout animals might model loss-of-function disorders (Picciotto & Wickman, 1998). However, many genes play a significant role during ontogeny, and thus compensatory mechanisms may completely overshadow the effects of the experimentally induced mutations. Recently, inducible knockouts have been developed to overcome the problems associated with functional compensation occurring as a consequence of the absence or inactivation of an endogenous gene during ontogeny (Gold, 1996; Gingrich & Roder, 1998; Picciotto & Wickman, 1998).

Antisense oligonucleotides (knockdowns) are produced by injecting a short sequence of synthetic DNA into the target area, for example, a specific brain region. The synthetic DNA sequence possesses a nucleotide arrangement that is complementary to a specific sequence of mRNA. The injected antisense oligomer blocks the translation of specific genes and thus inhibits synthesis of the relevant protein. Antisense-induced inhibition is transient and reversible, thus allowing assessment of the function of a gene in a *within subjects* design (Gold, 1996).

Investigations of the role of long-term potentiation (LTP) in spatial learning in the Morris task might serve as an example of the approach based on genetically engineered animals. The relationship between this electrophysiolgocial phenomon and spatial orientation in the Morris task has been studied by using mice that lack the gene for protein kinase A (Huang et al., 1995), protein kinase C (PKC_{γ}; Abeliovich et al., 1993), non-receptor tyrosine kinase C gene, *fyn* (Grant et al., 1992), or *Thy-1* (Nosten-Bertrand et al., 1996). These mutations result in disruption or elimination of hippocampal LTP. Spatial learning deficits (Abeliovich et al., 1993; Grant et al., 1992) or no effects on spatial learning (Huang et al., 1995; Nosten-Bertrand et al., 1996) have been reported. These results are inconclusive with respect to the role of LTP in Morris water escape learning, perhaps because LTP is usually elicited in hippocampal slice preparations. Correlation of these *in vitro* findings with spatial learning

performance may lead to results which are difficult to interpret (Errington et al., 1997), especially when LTP and learning are not assessed in the same individuals. However, Silva and colleagues (1998), in their review of the literature on genetically engineered mutations which affect LTP, concluded that LTP appears to be necessary, but not sufficient, for place learning.

There are many pitfalls in molecular genetic studies. For example, the genetic background into which transgenes are introduced is a determinant of the sensitivity of the approach. The importance of the genetic background for the interpretation of behavioral phenotypes expressed by genetically engineered animals has recently been recognized in a number of publications (e.g. Crawley et al., 1997; Logue et al., 1997; Owen et al., 1997; Gerlai, 1998; Lipp & Wolfer, 1998).

Another problem of gene-targeting or knockout animals is that the gene is inactivated not only in the structure(s) believed to be essential for spatial orientation behavior, such as the hippocampus (Jarrard, 1993, 1995), but in every cell that expresses the gene. As a consequence, the deficits may not be restricted to learning ability, but may affect other non-mnemonic processes, with the result that there is an apparent deficit in spatial learning and memory (Deutsch, 1993). New, refined techniques are under development which allow the inactivation of specific genes in small, well-defined brain regions (Roush, 1997; Picciotto & Wickman, 1998).

Although the new techniques undoubtedly will broaden the range of deficit models and, consequently, the understanding of brain-behavior relationships under normal and pathological conditions, Balaban and co-workers caution against premature and unjustified conclusions provoked by the identification of candidate genes in the regulation of behavior: *"(...) speaking about genes and (...) any behavioral phenomenon(...) requires an understanding of the specificity of the biological pathways which link gene differences to behavioral ones, and a knowledge of how consistently and robust certain genetic differences lead to certain behavioral differences in the face of variation in developmental histories, social and environmental circumstances, and different genetic backgrounds"* (Balaban, Alper & Kasamon, 1996, p. 4). Many of these factors are non-hereditary, and these non-genetic sources of variation in behavioral phenotypes are not less important than genetic factors (Oliverio, Cabib & Puglisi-Allegra, 1992).

Animals with CNS-specific lesions, or with cerebral ischemic damage

Animals with lesions of the nbm and with infarcts, induced by occlusion of the MCA, have been discussed above. The points raised in the evaluation of these models might also be relevant for other lesion and infarct models. It is very important to identify possible confounders of the measures under consideration in order to distinguish between specific and nonspecific effects of the experimentally induced brain damage on behavior.

Animals with behavioral deficits induced electrically, pharmacologically, or by hypoxia or anoxia

Many of the issues that have been discussed with respect to the nbm-lesioned animal and the animal with experimentally induced brain infarct are also true for models in which deficits are induced electrically (e.g. by electroconvulsive shock), pharmacologically (e.g. by administration of scopolamine), or by hypoxia and anoxia. These models will not be discussed further here and the reader is referred to recent reviews by, for example, Decker (1995) and Andrews (1996).

Combination of different models

It might be useful to combine different models. For example, the combination of cholinergic lesions and aging might produce a model that possesses more resemblance to the disease state of patients suffering from Alzheimer's disease than a model in which these lesions are induced in young subjects. However, our findings, reported in Chapter 3.3, do not support this notion. A similar combined approach might be appropriate for stroke. Most stroke patients are elderly. Indeed, Sutherland, Dix, and Auer (1996) presented results which support the notion that ischemic vulnerability is age-dependent; however, Yager, Shuaib & Thornhill (1996) only partially confirmed this increased vulnerability of old animals in a rat model of global transient ischemia. They found that susceptibility to hypoxia-induced brain injury did not increase linearly with age. Instead, the very young and the old rats were less susceptible to ischemia-induced damage than were the middle-aged subjects. Thus, a complex pattern of interactions might exist between the effects of chronological age and the neuropathological changes and functional consequences caused by experimentally induced brain damage. More investigations are needed before it can be decided whether the combination of different experimental approaches increases the validity of animal models of behavioral dysfunctions and their underlying pathology.

Non-human primates versus rodents

A topic that has not yet been addressed concerns the species that should be used in animal studies on naturally occurring or experimentally induced cognitive deficits. Intuitively, one would consider that the most human-like species is by definition the most appropriate model to assess the effects of putative cognition enhancing compounds. However, as D'Mello and Steckler (1996, p. 349) correctly stated, *"A close relationship to humans does not guarantee a similar pharmacodynamic response"*. Comparing cognitive functions of humans and non-human primates, Roberts (1996) concluded that even though the two species were tested in the same behavioral paradigm, they might have relied on different cognitive processes for the solution and the performance of the task, i.e. perceived similatities between species in the execution of a behavioral task might be due to mere analogy, not homology (Robbins, 1998). However, non-human primates might provide the models of choice if the higher level of organization and the higher complexity of the cortex are crucial components of the model (Kordower & Gash, 1996; Robbins, 1998), an aspect that cannot satisfactorily be mimicked in rodents. According to Roberts two requirements must be fulfilled before one can speak of behavioral homology:

- First, humans and monkeys should perform similarly in a given test.
- Second, the behavior shown should be supported by homologous brain structures.

Macphail (1996), who looked at the cognitive functions of mammals from an evolutionary perspective, concluded that there is a substantial degree of communality in the mechanisms of learning among vertebrates, which makes non-human and non-primate species suited for use in, for example, behavioral pharmacology. Moreover, there are severe constraints to the use of non-human primates as model species in neurosciences.

- First, and most important, ethical considerations preclude, or severely restrict, the use of non-human primates. This is a position which I share.

- Second, the availability of non-human primates is a limiting factor (Allain et al., 1997). This factor does not play a role if rodent models are used. In most cases, therefore, rodent-based approaches will be the models of choice.

However, a comparative approach, in which two or more species are considered, is to be favored, and the decision to accept or reject a specified model to screen putative cognition enhancers, or the decision to develop a substance clinically as a therapeutic should be based on the data from *more than one species*, not necessarily including non-human primates.

Are animal models necessary?

The discussion about the necessity and benefits of using animals in scientific research continues with the animal rights movement (e.g. Singer, 1985) playing a catalyzing role. However, the discussion is basically emotional, and the standards of exchange are low (Rowan, 1997). Even in the scientific community, positions about the usefulness of animal research range from *"(...) animal models can mislead researchers, or even contribute to illness and deaths by failing to predict the toxic effects of drugs".* (Barnard & Kaufman, 1997, p. 64) to *"Experiments using animals have played a crucial role in the development of modern medical treatments, and they will continue to be necessary as researchers seek to alleviate existing ailments and respond to the emergence of new disease."* (Botting & Morrison, 1997, p. 67).

Most researchers doing animal research are aware of the problems and pitfalls of animal models, and alternative approaches are continuously being evaluated. Nevertheless, many scientists conclude that animal models are still necessary (e.g. Kordower & Gash, 1986; Allain et al., 1997). For example, with respect to gerontological research, Ingram (1985, p. 328) concluded: *"Despite the apparent demographic demand for greater research of geriatric memory impairment using animal models, one can argue that the use of aged animals provides a potentially contaminated resource. Control over noncognitive performance factors presents a very demanding challenge. Indeed, researchers might contend that the use of young animals with specific brain lesions is perhaps a better approach. Certainly, this strategy should be fostered; however, even if it proves effective in identifying specific mechanisms and suggesting certain remedies, ultimately these remedies and therefore a test of the hypothesized mechanism will have to be attempted in aged animals. Therefore, we are stuck with aged animals as ultimate models."*

In 1988 Kordower and Gash pointed out that only a certain number of symptoms of a disease were mimicked by the animal models then available, which limited the predictive validity of the models. Unfortunately, this is still true approximately one decade later and highlights the need to *improve* the validity of animal models.

Reasoning similar to that of Ingram (1985) with respect to geriatric memory impairments is also valid for the neurological defects induced by stroke. As Tamura, Kawai, and Takagi, (1997, p. 283) stated: *"Recent advantages in medical technology are astonishing:it is possible to obtain some metabolic information in human stroke using PET or MRI. In a sense, we are in a new era of clinical ischemic study. However, we cannot obtain complete information without sampling brain tissue itself."* Although highly sophisticated *in vitro* approaches have been established, which allow the investigation of necrotic or apoptotic processes (Thompson, 1995; Bär, 1996; Choi, 1996), and the effects of

pharmacological interventions to attenuate or even prevent these processes of cell death, animals suffering from experimentally induced infarcts are still the ultimate model. Cells cannot model the interactions in complex systems, such as the living brain. This is true not only for neuropathological processes at the cell and system level, but also for the behavioral consequences of these neuropathological changes.

It is, of course, the responsibility of all researchers doing animal research to critically question the cost-benefits of their research. In this context, the principle of the three 'Rs' proposed by Russel and Burch (1959) serves as a good guideline. The three 'Rs' are *replacement*, *reduction*, and *refinement*.

- Animal research should be *replaced* by *in vitro* methods whenever possible. However, as Garattini (1997) pointed out, many effects measurable *in vivo* do not have adequate counterparts *in vitro*. It is, for example, not possible to study *in vitro* the effects of a therapeutic intervention on learning and memory processes, or on sensorimotor dysfunctions. Moreover, *in vivo* studies are necessary to assess the validity of *in vitro* approaches.
- The number of animals can be *reduced* by using appropriate experimental designs, sensitive and reliable methods, and state of the art statistical analytical methods. Moreover, an experiment should be adequately described, i.e. provide complete information about, for example, the animals used (sex, age, strain, supplier) and the rearing and testing conditions (description of the housing and testing environment, timing of testing, equipment used, etc.) (Russell, 1997). This information is indispensable for identifying intervening variables which might affect the outcome of a study, and which must be controlled experimentally to obtain valid and interpretable results.
- *Refinement* of the methods should reduce the suffering of animals. Every student of animal behavior should realize that optimal welfare conditions significantly reduce variation in behavioral studies. A secondary effect of refining experimental methods is an increased validity of the model used, a reduced variability in the data obtained, and consequently, a reduction of the number of animals needed to answer a particular scientific question.

The three 'Rs' provide a guideline for humane animal research. They guide the investigator to answer questions such as: what information could be obtained from the study? Is this information really necessary and essential? Can the information be obtained by means other than animal studies? How can the yield of a study, in terms of scientific information, be maximized and the use of animals, both with respect to number involved and the degree of discomfort and suffering, be minimized?

Biomedical research seems to be in a better position than fundamental scientific research with respect to the cost-benefit analysis of animal research. However, this is only superficially so because biomedical research aiming at the development of new therapeutic approaches relies on the knowledge provided by basic scientific research. For example, in order to identify and understand *pathological conditions*, one needs a sound knowledge of the *normal conditions* (Jucker & Ingram, 1997).

In my opinion, animal models of behavioral dysfunctions will no longer be necessary if the following conditions are fulfilled:

- Animal models will become obsolete if the human nervous system can be investigated by means of direct, non-invasive techniques. However, it has to be realized that the state of the patient assessed by (non-invasive) diagnostics is the endpoint of processes which led to dysfunctions. Therefore, the diagnostic data must be complemented by information which covers the processes preceding and

eventually causing the deficits. This can only be achieved if all potentially relevant information about the patient is systematically registered and stored.

- There is no further need of animal models if there is general consent that these models are not acceptable, even if this has direct consequences for our understanding of the processes underlying behavioral deficiencies, of their prophylaxis and treatment.

How to proceed

Animal models of behavioral dysfunctions remain the main tool to characterize compounds which can be used as therapeutics in the treatment of behavioral deficits caused by disease (e.g. Sarter, Hagan & Dudchenko, 1992a,b; Allain et al., 1997). Therefore, in this part of the discussion, animal models of behavioral dysfunction and the test paradigms thought to be appropriate to analyze the effects of putative cognition enhancers will be discussed. This is of course a personal view.

To continue with a statement made in the General Introduction (Chapter 1), animal models of behavioral dysfunction serve two main goals to enhance our understanding of the underlying substrates and mechanisms, i.e. brain-behavior relations, and to assess the effects of putative therapeutics to alleviate these dysfunctions. With respect to both goals, performance in learning and memory tests and functional ability in sensorimotor tests are usually the 'read-outs' or dependent variables.

Taking the assessment of cognitive function or dysfunction as an example, a major challenge is to develop valid test systems which enable the measurement of cognitive deficits, identification of classes of compounds which might act as putative cognition enhancers, and determination of their potency and efficacy.

"Potency indicates the dose or concentration which is usually needed to produce a certain effect. More potent agents produce specific effects at lower doses or concentrations." (Mutschler et al., 1995, p. 3). The potency of a drug can be determined by using a single test or test system. However, a substance does not qualify as a therapeutic because it is potent in a particular test. The test in which its potency is demonstrated has to be relevant to the indication, i.e. it should possess at the very least face validity and predictive validity.

"Efficacy is a term to describe the sum of all beneficial activities for prevention, alleviation, cure and diagnosis of diseases." (Mutschler et al., 1995, p. 3). This means that efficacy cannot be determined by using a single test or test system. Instead, a battery of tests has to be used, which in the case of cognition enhancers should ideally cover acquisition processes, consolidation processes, and memory processes in deficiency models which mimic the behavioral dysfunctions seen in humans as closely as possible, such as age-associated memory impairments (AAMI; McEntee & Crook, 1991) or the severe cognitive deficits observed in patients suffering from dementias (American Psychiatric Association: DSM IV, 1994). The more efficient compound is the one that modulates a broader range of cognitive processes. This means that a series of behavioral test paradigms has to be used to cover this range of cognitive functions.

As Maître and Pepeu (1989) state, the aim is to find cognition enhancers as remedies for the impairment caused by aging or disease processes and to find means of retarding the deterioration of

cognitive faculties. The term 'cognition enhancer' refers to agents that affect cognitive functions and display a characteristic pharmacological profile (Gamzu et al., 1989).

- A cognition enhancer is a compound that improves cognitive functions, such as learning, consolidation, and/or retrieval, (Gamzu et al., 1989; Schindler, 1989; Wenk & Olton, 1989), i.e. the acquisition, storage, and recall of information (Maître & Pepeu, 1989), that increases information processing capacity (Sarter et al., 1996), that augments brain resources (Wenk & Olton, 1989) and that restores brain function (McEntee & Crook, 1991). The improvement should be most readily observed under conditions of disturbed neurometabolism (Poschel, 1988; Schindler, 1989).
- A cognition enhancer has added value if it protects the CNS against brain insults (Gamzu et al. 1989; Schindler, 1989), does not have other 'classical' psychopharmacological activity (Gamzu et al. 1989), is characterized by minimal or no side effects (Poschel, 1988), and has very low toxicity (Schindler, 1989; Gamzu et al., 1989)
- A cognition enhancer must readily cross the blood-brain barrier (Poschel, 1988)

Most cognitive testing in deficiency models is done by using test systems which are cheap, fast, and simple. Spatial discrimination tasks in mazes, most commonly the Morris water maze, are among the most frequently used tests to detect potential cognition enhancers and to assess their potency (Merlini & Pinza, 1989; Andrews, 1996). A major point of discussion is whether these tests and other tests are sensitive, effective, and valid. The Morris water escape task has been found to be sensitive to the effects of putative cognition enhancing compounds (e.g. Brandeis et al., 1991; Pitsikas, Brambilla & Borsini, 1993; Blokland, Hinz & Schmidt, 1995; van der Staay, Hinz & Schmidt, 1996a,b; van Rijzingen et al., 1996; Fong, Neff & Hadjiconstantinou, 1997). However, the predictive validity of the Morris water escape task, which has been used in many of the experiments described in this book, remains to be elucidated, pending the successful introduction of cognition enhancers of high efficacy and potency in the clinic. Points of criticism of the Morris water escape task as behavioral 'read-out' for the identification and evaluation of potential cognition enhancers or disease modifiers are:

- The Morris water escape task is aversively motivated (Hodges, 1996). Owing to its aversive nature, stress responses might interfere with the cognition enhancing potential of test substances, and consequently, might lead to false-negative findings.
- It is not clear which cognitive processes are tapped by this task. In particular, the testing paradigms used might assess behavioral processes different from those compromised in cognitively impaired patients.

Steckler and Muir (1996) suggested that the battery of tests used to assess the effects of putative cognition enhancers should be extended and should be designed to assess those functions in rats that are involved in cognitive performance in humans.

Sarter, Hagan, and Dudchenko (1992b), in their review of the literature on screening of putative cognition enhancers, concluded that *"(...) the reader seeking advice for the establishment of screening procedures which would be cheap, fast and effective, may realize that no such tests are offered"* (p. 469), because *"(...) screening for cognition enhancers (...) is placed in a unique scientific-historical context characterized by the lack of a clinically effective treatment, of high pressure to discover and develop cognition enhancers, and the need for standardized and comprehensive preclinical tests in order to develop drugs for diseases of widely unknown etiology"* (p. 470).

In conclusion, one of the main reasons for the low predictive validity of animal models for putative cognition enhancing compounds for the treatment of cognitively impaired patients is their uncritical and indiscriminate use (Steckler & Muir, 1996), i.e. investigators using these tests are not concerned about their construct validity. Another major reason is, as discussed previously, that the defects induced in animals do not cover the pathology seen in patients, which restricts the relevance of the deficit model used.

D'Mello and Steckler (1996) have tried to formulate the features of what they consider to be the 'ideal' animal model(s) for human cognitive (dys)function. Six of the twenty-three features are summarized below. The summary does not reflect the order in which D'Mello and Steckler presented the features.

- First, two or more species should be considered. As already stated in Chapter 1, the comparative approach aims at studying the effects of experimental manipulations of a brain structure in more than one species (including humans, if possible) in order to try to generalize about brain structures, functions, behavior, and how they are related (Isaacson et al., 1971, p. 3).
- Second, two or more behavioral paradigms must be used which are believed to assess the same process. This provides an estimate of the construct validity of an experimental approach. However, apparently similarly operationalized tests (or measures) might rely on different processes. For example, measures of WM and RM, operationalized identically in a holeboard and a radial alley maze, were found to be poorly correlated (van Luijtelaar, van der Staay & Kerbusch, 1989). Hodges (1996) came to a similar conclusion after comparing Morris water escape tasks (both the standard version and the WM version) with radial arm maze tasks (both the standard WM version and the version in which only a subset of arms contains a food reward).
- Third, tests with positive reinforcement should be used, wherever possible. The main reason for this is that aversively motivated tasks might induce stress. However, deprivation techniques which induce a certain level of hunger or thirst, a prerequisite in tasks motivated by food or water as positive reinforcer, might also induce stress. Severe deprivation should always be avoided.
- Fourth, the performance baseline reached in a test should be neither too high nor too low, because otherwise the effects of drugs, lesions, and their possible interaction might be confounded by floor or ceiling effects.
- Fifth, the dose range used in a study should include doses which have no effect and doses which give rise to side effects. The dose which induces first side effects should be determined before the effects of putative cognition enhancers are tested in learning and memory tasks. Drugs should be first characterized in observational tests such as the functional observation battery (FOB, Moser, 1990) or the (modified) Irwin test (MIT, Irwin, 1968) in order to avoid over-dosing and to learn about the side effect profile and the safety-efficacy ratio of the drug under study. An additional benefit of proper pre-characterization of compounds is that unnecessary animal studies are avoided.

  To assess the role of the non-cognitive effects of naturally occurring or experimentally induced deficits on cognitive processes, and to estimate their contribution to the effects of putative cognition enhancing compounds, it might be necessary to use tests which are believed to measure anxiety, exploration, or locomotor activity (e.g. open field, light-dark preference box, elevated plus maze, etc.). In addition, a variety of neurological tests are available to further characterize non-cognitive deficits and the effects of putative therapeutics on these impairments. A number of these functional tests have been used in the experiments described in Chapters 3.3, 4.1 and 4.2.

- Sixth, tests should be repeated and results should be replicable. Unfortunately, the replicability of test results has received little attention to date. Irreproducible results contribute to the ever increasing number of false-positive hits (Sarter, Hagan & Dudchenko, 1992a,b). Conclusions should not be based on unreplicated findings. Ideally, results should be confirmed by an independent group of researchers.

  The researcher should be aware that important characteristics of the rat or mouse strain(s) used can change considerably over even only a few years (Andrews, 1996; van der Staay, 1997; Chapter 2.4), and that different lines of the same rat or mouse strain can differ (even if genetic markers are considered; e.g. van Zutphen & den Bieman, 1984; Festing, 1993). All these factors might be responsible for differences in the findings of the same laboratory, and even more so for differences in the findings of different laboratories.

Bearing in mind these and additional recommendations of D'Mello and Steckler (1996) and of Sarter, Hagan, and Dudchenko (1992a,b), a broad characterization of putative cognition enhancers is recommended.

Tasks such as the Morris task(s) might provide first hints as to the efficacy and potency of a putative cognition enhancer. Moreover, the Morris water escape task allows a very detailed analysis of the animals' behavior; for example, measures believed to reflect spatial learning and memory, response strategies, and motor performance can be derived. The efficacy of a test compound should be determined in a series of tests that cover different cognitive processes (Maître & Pepeu, 1989). A selection of these tests is discussed below. Shock-motivated tasks, such as the passive and active avoidance test, are included because they have been used extensively to assess the effects of experimentally induced deficits and their alleviation by putative cognition enhancers. The discussion closes with a review of the tests which I consider to be more useful than shock-motivated tasks for evaluating cognitive deficits.

Shock-motivated avoidance tasks

Passive or inhibitory avoidance

The passive or inhibitory avoidance test, which Iversen (1997) considers as a useful early screening test, is still used in the majority of drug screening programs worldwide. However, there is growing agreement that the validity of the passive avoidance test is extremely low. This test yields an unacceptably high number of false-positive hits (Sarter, Hagan & Dudchenko, 1992a,b; Porsolt, McArthur & Lenègre, 1993). This having been said, most cognition enhancers which currently under clinical development affect behavior in the passive avoidance task. This might be the main reason for the incessant use of this test. There is, however, no valid *scientific* reason to rely on this test. There is only one non-scientific reason for using this test: it is fast.

Active avoidance, or shuttle-box learning

Similar objections might hold for active avoidance paradigms. Acquisition and retention of the two-way active avoidance task can be assessed simultaneously in many rats in commercially available, computer-driven rodent shuttleboxes. This task appears to be sensitive to the effects of putative cognition enhancing compounds (e.g. Groó, Pálosi & Szporny, 1989; van der Staay, Hinz & Schmidt, 1996a). Substances which induce hyperactivity are expected to facilitate active avoidance and to impair inhibitory avoidance (Sansone et al., 1991), whereas substances which induce hypoactivity are

expected to facilitate inhibitory avoidance and to impair active avoidance. If these tests are going to be used, it makes sense to test animals in both paradigms (passive and active avoidance) to exclude false-positive and false-negative outcomes.

The restrictions and pitfalls of shock-motivated avoidance tasks should be kept in mind (see, for example, Sarter, Hagan & Dudchenko, 1992b; Porsolt, Roux & Lenègre, 1991; Andrews, 1996). A positive effect in these tasks does not necessarily indicate that the substance tested is a cognition enhancer. Instead, the conclusion should be that the drug affects behavior in one way or another, possibly by modulation of central nervous system processes. The limitations of shock-motivated avoidance tasks are a reason to use more sophisticated tests, preferably tests which are positively motivated (D'Mello & Steckler, 1996). Such tests should tap different processes involved in cognition: attention, acquisition, consolidation, retrieval of information (recall or recognition).

Tests which might be useful to assess behavioral defects and putative therapeutics

A variety of tests that cover relevant aspects of information processing are needed to characterize cognitive impairments in animal models of behavioral deficiency and to evaluate putative cognition enhancing compounds. The tests mentioned below are in my opinion appropriate for achieving these goals. It must be stressed that these tests should continuously be evaluated, refined, or eventually be replaced by alternative procedures, if their validity is considered insufficient.

Spatial orientation tasks

Standard Morris water escape task

The standard Morris task, in which a rat is trained to localize a submerged platform, measures predominantly spatial RM (Mundy, Barone and Tilson, 1990). The relevance of this test has already been discussed above. Spatial orientation learning in this task shows an age-related decline (see Chapters 2.1, 2.2. 2.3) and specific brain lesions, especially of the hippocampus, disrupt performance in this task (e.g. Barnes, 1988b; Stubley-Weatherly, Harding & Wright, 1996). Treatment with a putative cognition enhancer has been found to (partially) antagonize these naturally occurring (e.g. van der Staay, Hinz & Schmidt, 1996a), or experimentally induced deficits (e.g. Lamberty & Gower, 1991a), and even to improve performance in young, intact animals (van der Staay, Hinz & Schmidt, 1996a,b).

Working memory version of the Morris water escape task

Besides RM versions of the Morris water escape task, versions have been developed which allow the assessment of a WM or short-term memory component (Whishaw, 1985, 1987; see also Chapters 2.2, and 4.4) We found that young rats acquire the WM version of the Morris task within the first sessions whereas 24-month-old rats, even after 12 daily training sessions, did not (van der Staay & de Jonge, 1993; Chapter 2.2). We also found that adult mice acquired this task (Klapdor & van der Staay, 1998; Chapter 4.4). No information, however, is currently available about the sensitivity of this version of the Morris task for selective brain lesions or for the effects of putative cognition enhancing compounds. Moreover, it remains to be evaluated whether the standard (RM) and the repeated (WM) version of the Morris water escape task measure independent processes.

The radial maze

The radial maze with eight or more arms radiating from a central platform (Olton and Samuelson, 1976) is a 'free choice' type maze in which the animal is free to visit all places in whichever order it wants to. The most efficient behavior is to remember the list of places already visited during a trial. Rats have been found to use different strategies to negotiate the maze (Hodges, 1996). By adopting a strategy such as *running around in circles*, i.e. to enter adjacent arms in a clockwise or anti-clockwise direction (Yoerg & Kamil, 1982) animals might be able to fully compensate for experimentally induced deficits without using spatial memory. However, although most experimenters consider that adoption of a successful strategy is a confounding variable, it might reflect an 'intelligent' compensation for lost capacities (Davis, 1996). Therefore, investigation of the development of (foraging) strategies in animals with experimentally induced brain damage might provide relevant information about compensatory strategies.

Olton and Samuelson (1976) supposed that in their 'elevated radial maze' the list of visits is held in the WM. Spatial discrimination learning in the WM version of the radial maze is fast and easy. In this version of the task all arms of the maze are baited with food or water reward. Rats acquire the task rapidly and usually reach a nearly error-free performance within as few as 10 trials. Because of the very fast acquisition of the task, it is not easy to detect the effects of putative cognition enhancers. The radial maze, therefore, should only be used in combination with a deficiency model, i.e. with animals which suffer from age-related cognitive deficits or which have experimentally induced brain damage known to disrupt cognitive performance. For example, lesioning of (parts of) the hippocampal formation (review: Jarrard, 1993), the septum (Nilsson & Gage, 1993; Kelsey & Vargas, 1993; Riekkinen, Schmidt & Riekkinen, 1997), or the fimbria (e.g. van der Staay et al., 1989) has been found to induce deficits in spatial learning and memory.

Unlike the WM, which holds information that is relevant only within a specific trial, the RM (Olton and Papas, 1979) holds trial independent information, for example about the locations where the food is hidden. Thus, if food can only be found in a subset of the potential places that can be visited by the foraging rat, two memory components can be distinguished simultaneously: WM and RM. Significantly more trials are needed to distinguish between spatial WM and RM when only one subset of arms is baited (Hodges, 1996). The sensitivity of this version of the radial maze for the effects of putative cognition enhancers still has to be established.

Attention

Recently, workers in the field of animal cognition have focused on attention (e.g. Steckler & Muir, 1996; Turchi & Sarter, 1997). *"Attention refers to those aspects of perception where stimulus elements are actively selected from the environment"* (Steckler & Muir, 1996, p. 301). These authors conclude that attention is an undervalued, but extremely important component when assessing cognitive functions in animals. As a consequence, a number of attention tasks, most of them reaction time tasks, have been developed to remedy this situation.

Simple reaction time task and choice reaction time task

Moore and co-workers (1992) developed a simple reaction time task (SRTT) and a choice reaction time task (CRTT), both of which appear to provide valid measures of vigilance in rats. Both tests are run in an operant conditioning box (Skinner box). Rats are trained to respond to one of two levers after an inter-trial interval (ITI) of 15 ± 8 seconds and a fixed interval of 3 seconds (FI3"). Operating a lever

during these periods resets the ITI. If there is no response during these periods, a light signal is switched on for 50 milliseconds. A lever press response within 3 seconds from this signal is considered as a 'hit', which is reinforced by a food pellet. The response starts the next ITI. In the SRTT, the light signal is positioned equidistant from both response levers, and a correct lever press on either lever produces a reinforcement. In the CRTT, the light signal is randomly presented above the left or right lever, and only pressing the signaled lever within 3 seconds from stimulus presentation produces a response. Lever presses in the 3-second period preceding a light signal are considered as 'false alarm'. It is conceivable that vigilance tasks can be refined further, based on the work by Moore and colleagues (1992).

In a study assessing the effects of experimentally induced subdural hematoma (SDH) in rats, Klapdor and colleagues (1997a) used a modification of the above mentioned CRTT. The animals were first trained to press a lever that was indicated by a visual stimulus (S+), a square consisting of an 8 by 8 matrix of green light emitting diodes (LEDs). They had to respond by pressing the lever underneath the display which had presented the stimulus. As soon as the animals reached a level of 60-70% correct responses, the stimulus duration was varied within trials. The durations were 0.15, 0.3, 0.6, 1.2, or 2.4 seconds. After a correct response, the levers were retracted and a food pellet was delivered into the illuminated feeder. As soon as the pellet was removed from the food tray by the rat, the feeder light was extinguished, and there was a 5-second time-out before the next random trial commenced.

The stimulus duration clearly affected performance before surgery. The longer the duration, the better the performance. SDH impaired performance at stimulus durations of 0.6 and 1.2 seconds, whereas there was only a marginal impairment with the longest (2.4 seconds) stimulus duration. Performance in response to the shorter stimulus durations was not affected, probably because the performance with the shortest stimulus durations was much closer to chance level before to surgery than was the performance with the longer stimulus durations. Although this modification of the task has been found to be sensitive to SDH-induced deficits, it still must be validated pharmacologically.

Five-choice serial reaction time task

The five choice serial reaction time task was developed by Carli and co-workers (1983) to assess visual attention in rats. The task is a modification of the five-choice serial reaction time task developed by Leonard (1959) to measure selective attention in humans. The apparatus consists of a box with a curved rear-wall, into which five equally spaced holes are inserted. The holes are provided with a photobeam detection system which allows the automatic registration of nosepokes. In the front wall, equidistant to the five holes, there is a hinged perspex panel which provides access to a food tray connected to a pellet dispenser. The holes can be illuminated individually by a bulb at the rear of each hole. A rat is trained to poke its nose into the hole that has been illuminated for 0.5 seconds. A correct response is reinforced with a food pellet, delivered into the food tray. An incorrect response or an omission (failure to respond within a preset period, e.g. 5 seconds, after a hole had been illuminated) is followed by a time-out period. Taking a food reward from the tray, or the end of the time-out period starts a new trial.

Performance on this task can be disrupted by scopolamine (Gutnikov, Barnes & Rawlins, 1994; Jones & Higgins, 1995), or by selective brain lesions (Muir, Everitt & Robbins, 1996) and appears to be sensitive to aging (Jones et al., 1995), to strain differences (Didriksen & Christensen, 1993), and to the effects of putative cognition-enhancing compounds (Sirviö et al., 1993). The task might be suited to test the effects of different drugs in the same animal. However, the task appears to be sensitive to the effects of compounds on response strategies and on locomotor activity (Gutnikov, Barnes & Rawlins,

1994). The researcher using this task to assess either attention or WM performance thus must be aware of the possibility that the results obtained reflect effects on non-mnemonic processes.

Matching tasks

In a delayed matching to position (DMTP) task, the subject is randomly offered one of two retractable levers as the sample stimulus in an operant conditioning apparatus (Skinnerbox) (Dunnett, 1985; van Hest, 1989). The lever is retracted immediately after a lever press response, and a delay (ranging from, for example, 0 to 60 seconds) is introduced. A rat must poke its nose into the food tray during the delay. Upon the first poke after the expiration of the programmed delay both levers are inserted as matching stimuli. Nose pokes are required in order to prevent rats from simply waiting in front of the sample stimulus during the delay. Pressing on the previously presented sample lever produces a food reward. In a modification of the DMTP task, the delayed non-matching to position (DNMTP) task, operation of the lever that was *not* presented as the stimulus is reinforced with a food reward (Sahgal, Keith & LLoyd, 1990). The DMTP and DNMTP tasks can be used as screening tools for the assessment of drug effects on short term memory processes. The performance of aged rats and of rats with experimentally induced brain damage is impaired in the DNMTP task (Dunnett et al, 1988; Dunnett, Rogers & Jones, 1989; Roux et al., 1994). However, the sensitivity of this task to the effects of putative cognition enhancers must still be examined.

Timing behavior

Time-discrimination performance in the Skinner box might provide an additional tool to test the cognition enhancing properties of new therapeutics. The timing behavior can be assessed by a peak interval procedure (discussed in detail by Roberts, 1981). Rats are trained to respond in a discrete fixed interval 20 seconds schedule until the maximum response rate occurs at about 20 seconds (‘peak time’: the time during which the rat maximally ‘expects’ food). The start of each trial is signaled by the onset of white noise. The first lever press following the critical 20-second interval is reinforced and the noise is turned off. A 130-second time-out interval is allowed between the termination of the noise signal and its onset at the start of the next trial. Then, no reinforcement is given for half of the trials, chosen at random, and the white noise is turned off after 50 seconds, independently of whether the animal has responded. Responses made during these ‘empty’ trials are fitted to a scalar timing model (Roberts, 1981; Gibbon, Church & Meck, 1984; Meck & Church, 1987).

The maximal response rate can shift to a shorter or longer interval as a consequence of experimental manipulations. The shift can be due to a change in the speed of the internal clock or be due to a change in the remembered time of reinforcement. Meck and co-worker (Meck, 1983; Meck & Church, 1987; Meck, 1996) have elaborated an information-processing model of timing behavior which includes the neuronal connections and the brain structures involved in the processing of timing. This allows them to distinguish between the effects of treatments on the speed of the internal clock and on memory processes. In general, cognition enhancers are expected to shift the peak to the left, whereas substances which impair cognition are expected to shift the peak to the right.

Concluding remarks

It should be clear right now that it will not be possible to use all these tests in a single study. Moreover, additional behavioral tests might be needed to characterize particular deficiency models and to evaluate the effects of putative therapeutics. Many factors affect the development of deficiency models. Although substantial progress has been made with respect to deficiency models, the endpoint, i.e. a set of generally accepted valid models, has not yet been reached. It is the task of comparative and physiological psychologists, together with experts from other disciplines, to further improve the animal models of behavioral deficiency.

What is needed now is the concerted validation of new animal models and less well characterized older models. This would have the advantage that validation is fast, that it is multidisciplinary, and that subsequent research based on poor or inappropriate models can be avoided. Inefficient use of human resources and, equally important, unnecessary use of animals can be avoided if consent is reached about the appropriateness of a model. In my opinion this approach is also a way to reduce animal experimentation, to increase the quality of the models used, and consequently, to increase the knowledge gained from research with animal models. Improving the validity of the models used and eliminating models which are not valid clearly would improve the cost-benefits balance of animal models.

McKinney (1984) drew some key points from his overview of models of depression which can be generalized to models of neuropathological conditions and behavioral dysfunctions (italics are citations, normal printing are my replacements and extensions of the term 'depression' in McKinney's paper):

"1. Animal models are not replicates of human illnesses nor do they represent the illness in miniature.

2. *They should properly be conceived of as experimental systems in which selected and specific questions regarding* neuropathological conditions and behavioral dysfunctions *can be investigated in ways impossible to do in humans.*

3. *There is no simple, comprehensive model for* neuropathological conditions and behavioral dysfunctions *(...). Each approach has advantages and limitations. Therefore the continuing study of a number of such experimental preparations is indicated.*

4. *Multiple variables are involved in the etiology of* neuropathological conditions and behavioral dysfunctions *and special advantage of animal models is the possibility of evaluating the main effect of each, while studying their interactions in controlled, prospective design.*

5. *While one should always be careful with cross species reasoning, nevertheless there are guidelines, and if one exercises proper scientific caution in this regard the continuing development of a comparative approach (...) has great potential."* (McKinney, 1984, p. 94).

Animal models are a controversial topic in science, and even more so in public opinion. They are often limited and crude approximations of the human condition. In such approaches, there is the inherent danger that the conclusions drawn are erroneous. However, *"The dangers are not in working with models, but in working with too few, and those too much alike, and above all, in belittling any efforts to work with anything else."* (Kaplan, 1973, p. 293). Animal models are not used for their own sake, but are a means to an end, the endpoint being advancement of knowledge about function in health and disease. Animal models are one of many alternative approaches to reach this goal. It is the task of the neuroscientist to identify the shortcomings of animal models, to refine the models, and to develop tools which allow a better understanding of the human condition. Developing valid animal models of

behavioral dysfunctions is a difficult endeavor. As Kaplan said, *"It would be a rash to attempt a priori to set limits on the fruitfulness of models in behavioral science."* (1973, p. 292). In the end *"The success of any animal model of human aging* (or of behavioral dysfunction, FJvdS) *depends on the ability to relate the behavioral and neural findings back to the human condition in some meaningful way."* (Barnes, 1988a, p. 563). It is only by identifying the weaknesses and errors of models that improvements can be made. I am convinced that animal models of behavioral dysfunctions will make significant contributions to our understanding of the processes underlying behavioral dysfunctions in humans, provided that the neuroscientists working with animal models are aware of the problems inherent in this approach and contribute to the refinement and validity of the animal models.